Statistik II - Elementare Wahrscheinlichkeitsrechnung

Prof. Dr. Christine Müller Technische Universität Dortmund

Sommersemester 2020

Literatur

- Henze, N. (1997) Stochastik für Einsteiger. Vieweg, Braunschweig.
- Henze, N. (2019) Stochastik: Eine Einführung mit Grundzügen der Maßtheorie. Springer, Berlin.
- Kersting, G. und Wakolbinger, A. (2008). *Elementare Stochastik*. Birkhäuser, Basel, 18.90 EUR.
- Krengel, U. (2005). Einführung in die Wahrscheinlichkeitsrechnung und Statistik. Vieweg, Braunschweig, 24,90 EUR.
- Milbrodt, H. (2010). Wahrscheinlichkeitstheorie Eine Einführung mit Anwendungen und Beispielen aus der Versicherungs- und Finanzmathematik. Verlag Versicherungswirtschaft, Karlsruhe.
- Moeschlin, O. (2003). Experimental Stochastics. Springer, Berlin.

INHALTSVERZEICHNIS

Inhaltsverzeichnis

1	Ein	Einführung		
	1.1	Fragestellungen der Wahrscheinlichkeitsrechnung	4	
	1.2	Simulation des Zufalls	6	
	1.3	Mengentheoretische Grundlagen	9	
2	Laplace-Experimente			
	2.1	Laplace'scher Wahrscheinlichkeitsraum	13	
	2.2	Bestimmung der Kardinalität einer Menge	15	
	2.3	Bildmaße	17	
	2.4	Aus Laplace-Experimenten abgeleitete Wahrscheinlichkeitsmaße	19	
3 Wahrscheinlichkeitsräume und zufällige Bilder		hrscheinlichkeitsräume und zufällige Bilder	21	
	3.1	Allgemeine Wahrscheinlichkeitsräume	21	
	3.2	Zufallsvariable	28	
4	Diskrete und absolut stetige Wahrscheinlichkeitsverteilungen			
	4.1	Diskrete Wahrscheinlichkeitsverteilungen	35	
	4.2	Absolutstetige Wahrscheinlichkeitsverteilungen	40	
	4.3	Verteilung mehrdimensionaler Zufallsvariablen	52	
5	Bed	lingte Wahrscheinlichkeiten und stochastische Unabhängigkeit	5 9	
	5.1	Bedingte Wahrscheinlichkeiten	59	
	5.2	Unabhängigkeit von Ereignissen	65	
	5.3	Unabhängigkeit von Zufallsvektoren	67	
6	Funktionen von Zufallsvektoren			
	6.1	Projektionen von Zufallsvektoren	74	
	6.2	Summe von Zufallsgrößen	77	
	6.3	Weitere Abbildungen von Zufallsvektoren	81	

IN		Christine Müller, Statistik II - Elementare Wahrscheinlichkeitsrechnung, SS 2020 TSVERZEICHNIS	3
7	Stochastische Abhängigkeit		
	7.1	Bedingte Dichten bei diskreten Zufallsvariablen	83
	7.2	Bedingte Dichten bei absolut stetigen Zufallsvariablen	86
8	Ma	ßzahlen für Verteilungen	89
	8.1	Erwartungswert	89
	8.2	Streuungsmaßzahl (Varianz)	96
	8.3	Abhängigkeitsmaßzahlen	102
9	Koı	nvergenz von Zufallsvariablen	109
	9.1	Das schwache Gesetz der großen Zahlen	109
	9.2	Das starke Gesetz der großen Zahlen	116
	9.3	Verteilungskonvergenz	123
	9.4	Zentraler Grenzwertsatz	128

1 Einführung

1.1 Fragestellungen der Wahrscheinlichkeitsrechnung

1.1.1 Beispiel (Ein Problem des Chevalier de Meré, 1607-1685)

Der Chevalier de Meré fragte sich, was wahrscheinlicher bei einem dreifachen Würfelwurf ist: Die Summe der Augen der drei Würfel ergibt eine 11 oder die Summe der Augen der drei Würfel ergibt eine 12. Oder sind beide Ereignisse gleichwahrscheinlich? Zunächst betrachten wir die Möglichkeiten, die Zahlen 11 und 12 als Summe dreier Zahlen aus der Menge $\{1, \ldots, 6\}$ zu schreiben:

11:
$$1+4+6$$
; $1+5+5$; $2+3+6$; $2+4+5$; $3+3+5$; $3+4+4$

$$12: 1+5+6; 2+4+6; 2+5+5; 3+3+6; 3+4+5; 4+4+4.$$

Insgesamt hat man daher jeweils 6 Möglichkeiten, die Zahlen 11 und 12 zu kombinieren, was gleiche Wahrscheinlichkeit beider Ereignisse nahelegt. Ist das richtig?

1.1.2 Beispiel (Geburtstagsprobleme)

- a) Wie groß ist die Wahrscheinlichkeit, dass zwei Personen der Vorlesung am gleichen Tag Geburtstag haben?
- b) Wie groß ist die Wahrscheinlichkeit, dass eine Person der Vorlesung am gleichen Tag Geburtstag hat wie die Dozentin?
- c) Welches von den beiden Ereignissen ist wahrscheinlicher?

1.1.3 Beispiel (Buffon'sches Nadelproblem)

Auf ein liniiertes Papier unendlicher Größe mit Linienabstand 1 wird zufällig eine Nadel der Länge 1 geworfen. Wie groß ist die Wahrscheinlichkeit für das Ereignis, dass die Nadel eine Linie schneidet?

1.1.4 Beispiel

Der ELISA-Test zur Erkennung einer HIV-Infektion (siehe Henze 1997, Stochastik für Einsteiger, S. 108-109) liefert mit Wahrscheinlichkeit 0.997 ein positives Ergebnis, wenn eine Person tatsächlich erkrankt ist. Die Wahrscheinlichkeit, dass eine gesunde Person ein positives Ergebnis erhält, beträgt 0.001. Wie groß ist die Wahrscheinlichkeit, HIV infiziert zu sein, wenn der Test positiv ist und man weiß, dass 0.2% der Bevölkerung die Infektion haben? Wie ändert sich diese Wahrscheinlichkeit, wenn 2% der Bevölkerung infiziert sind?

1.1.5 Beispiel

Beim Sommerfest der TU Dortmund möchte die Fachschaft Statistik ihr Budget aufbessern. Man entschließt sich zu folgender Lotterie: Der Spieler würfelt mit zwei Würfeln. Bei einem Pasch erhält er als Auszahlung die Gesamtaugenzahl (also Zahlen zwischen 2 und 12) in Euro, ansonsten nichts.

1.1 Fragestellungen der Wahrscheinlichkeitsrechnung

- a) Wie hoch muss der Spieleinsatz gewählt werden, damit das Spiel fair ist, d.h. dass die Fachschaft auf lange Sicht weder Gewinn noch Verlust macht?
- b) Die Fachschaft will aber natürlich Gewinn machen. Es wird von 200 Spielen ausgegangen. Die Fachschaft will einen erwarteten Gewinn G von 100 Euro erzielen. Wie hoch muss dazu der Spieleinsatz sein?

1.2 Simulation des Zufalls

Früher haben die Leute viel gewürfelt, um Erfahrungen mit dem Zufall zu sammeln. Heute können wir den Zufall mit Zufallszahlen simulieren.

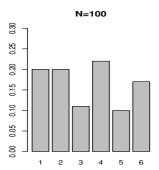
1.2.1 Beispiel (Simulation des Würfelwurfes)

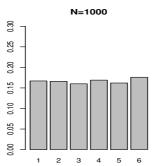
Analysiert man die physikalischen Eigenschaften eines symmetrischen Würfels, so liegt die Vermutung nahe, dass keine Seite (Ziffer) beim Wurf bevorzugt wird. Diese Eigenschaft wird beispielsweise in einer Vielzahl von Spielen genutzt. Alle Seiten (Ziffern) sind somit als gleichwahrscheinlich zu betrachten, und es wird daher etwa erwartet, dass in sechs Würfen eine Sechs fällt. Die gleiche Situation liegt vor, wenn aus einer Urne mit sechs Kugeln mit Zurücklegen gezogen wird und die Kugeln mit den Nummern 1,2,3,4,5,6 nummeriert sind. In R können beide Zufallsexperimente mit der R-Funktion sample simuliert werden.

```
> sample(1:6,size=10,replace=T)
[1] 1 4 2 6 6 1 5 5 1 1
```

Mit den R-Funktionen table und barplot können die absoluten Häufigkeiten der Zahlen 1,2,3,4, 5,6 bei N Würfelwürfen ermittelt und anschließend grafisch in einem Balkendiagramm dargestellt werden. Um drei Grafiken in einer darzustellen, wird hier außerdem die R-Funktion par mit dem Argument mfrow=c(1,3) benutzt, wobei c(1,3) bedeutet, dass die Grafiken in Form einer 1×3 -Matrix angeordnet werden, d.h. eine Zeile mit 3 Spalten.

```
> par(mfrow=c(1,3))
> w<-sample(1:6,size=100,replace=T)
> barplot(table(w)/100,main="N=100",ylim=c(0,0.3))
> w<-sample(1:6,size=1000,replace=T)
> barplot(table(w)/1000,main="N=1000",ylim=c(0,0.3))
> w<-sample(1:6,size=10000,replace=T)
> barplot(table(w)/10000,main="N=10000",ylim=c(0,0.3))
```





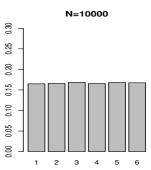


Abbildung 1: Relative Häufigkeiten bei N Würfelwürfen

Insbesondere können wir untersuchen, wie sich der Anteil der Sechsen entwickelt, wenn die Anzahl N der Würfelwürfe immer größer wird. Die Datei Plot_Anteil_der_Sechsen.asc enthält die Funktion Plot.Anteil.der.Sechsen, nämlich:

```
"Plot.Anteil.der.Sechsen" <-
function ()
{
# Plottet den Anteil der Sechsen für Wurfanzahlen von 10 bis 500
plot(seq(0,500,50),seq(0,500,50)*0+1/6,xlab="Wurfanzahl",
    ylab="Anteil der Sechsen", ylim=c(0,1/2),type="l")
for(N in seq(10,500,10)){
   points(N,sum(sample(1:6,size=N,replace=T)==6)/N,pch=16)
}
}</pre>
```

Diese Funktion zeichnet mit plot als erstes eine waagerechte Linie bei 1/6. Dann werden in einer for-Schleife sukzessive mit points die Anteile der Sechsen als Punkte hinzugefügt.

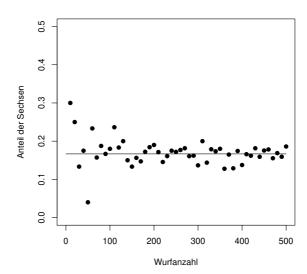


Abbildung 2: Relative Häufigkeiten der Sechsen mit wachsendem N

Die Abbildung 2 zeigt, dass die relative Häufigkeit der Sechs um den Wert 1/6 streut und dass diese Streuung um so geringer wird, desto höher die Anzahl N der Würfelwürfe wird. Es sieht so aus, dass die relative Häufigkeit der Sechs gegen 1/6 konvergiert. Natürlich haben wir dieses Wissen schon in die Konstruktion der Zufallszahlen, die die Würfelergebnisse simulieren, hineingesteckt. Aber mit einem richtigen Würfel würde man das gleiche Verhalten bekommen.

Das Auftreten der Sechs im Würfelwurf mit der Wahrscheinlichkeit 1/6 zu verbinden, war hi-

1.2 Simulation des Zufalls

storisch kein einfacher Schritt. Dieser Wert, der als Grenzwert auftritt, wird ja nie im endlichen Fall beobachtet. Im endlichen Fall treten mit den relativen Häufigkeiten, die man nur beobachten kann, alle möglichen Werte auf und, wenn N nicht durch 6 teilbar ist, tritt 1/6 als relative Häufigkeit gar nicht auf. Noch schwieriger war es, ein mathematisches Konzept zu entwickeln, mit dem dieses **Gesetz der großen Zahlen** auch mathematisch bewiesen werden kann. Dieses Konzept wurde erst vollständig – in axiomatischer Weise – in den zwanziger Jahren des letzten Jahrhundert entwickelt.

In den folgenden Abschnitten wird dieser axiomatischer Ansatz zur Beschreibung des Zufalls vorgestellt und dann wird gezeigt, wie man damit arbeiten kann. Für dieses Konzept sind einige mengentheoretische Begriffe notwendig.

Doch vorher simulieren wir noch das Problem des Chevalier de Meré.

1.3 Mengentheoretische Grundlagen

Bei der Modellierung zufällsabhängiger Vorgänge werden zunächst die möglichen Ergebnisse angegeben, die als Konsequenz dieser Vorgänge möglich sind. Zwei Konsequenzen müssen einander ausschließen.

1.3.1 Definition

Die Menge aller möglichen **Ergebnisse** ω eines Zufallsvorgangs (Zufallsexperiments) wird **Grund-menge** (Grundraum, Ergebnisraum) genannt und mit dem griechischen Buchstaben Ω bezeichnet:

 $\Omega = \{\omega; \ \omega \ \text{ist m\"{o}gliches Ergebnis eines zufallsabh\"{a}ngigen Vorgangs}\}.$

Eine **Menge von Ergebnissen** heißt **Ereignis**. Ereignisse werden mit großen lateinischen Buchstaben A, B, C, \ldots bezeichnet. Ein Ereignis, das genau ein Ergebnis als Element besitzt, heißt **Elementarereignis**.

1.3.2 Beispiel (Würfelwurf)

Wir betrachten das Zufallsexperiment eines einfachen Würfelwurfs. Die möglichen Ergebnisse sind die Ziffern $1, \ldots, 6$, d.h. die Grundmenge ist $\Omega = \{1, 2, 3, 4, 5, 6\}$. Elementarereignisse sind $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ und $\{6\}$. Man beachte, dass Elementarereignisse Mengen sind, während Ergebnisse Elemente sind. Andere Ereignisse sind etwa:

- gerade Ziffer: $A = \{2, 4, 6\},\$
- ungerade Ziffer: $B = \{1, 3, 5\},\$
- Ziffer kleiner als 5: $C = \{1, 2, 3, 4\},\$
- Ziffer ist Summe zweier verschiedener Ziffern: $D = \{3, 4, 5, 6\},\$
- Ziffer ist Primzahl: $E = \{2, 3, 5\},\$
- Ziffer ist kleiner gleich 6: $F = \{1, 2, 3, 4, 5, 6\} = \Omega$.

1.3.3 Beispiel (Ergebnisraum, Grundraum)

Experiment	Grundraum Ω	Element ω
a) Würfelwurf	$\{1,\ldots,6\}$	gewürfelte Augenzahl
b) Münzwurf	{ Kopf (K), Zahl (Z) }	obere Seite der geworfenen Münze
c) Roulette	$\{0, 1, \dots, 36\}$	Ergebnis des Drehrades
d) Warten auf erste "6" beim Würfeln	$\mathbb{N} \cup \{\infty\}$	$\omega \in \mathbb{N}$: erste "6" tritt im ω -ten Wurf auf; $\omega = \infty$: es tritt keine "6" auf
e) Lotto "6 aus 49"	$\begin{cases} \{(\omega_1, \dots, \omega_6); \\ 1 \le \omega_1 < \dots < \omega_6 \le 49 \} \end{cases}$	ω_i ist Nr. der i -t größten gezogenen Kugel
f) Defektstelle einer Leitung (Gas, Telefon) der Länge 1km	[0,1]	Defekt an der Stelle ω
g) Windrichtung an einer Meß- stelle	[0, 360)	ω bezeichnet den Winkel der Windrichtung (z.B. Nord: $ω = 0°$; Ost: $ω = 90°$, etc.)
h) Wartezeit bis zum Ausfall einer Maschine	[0,T)	Zeit bis zum Ausfall (Höchstgrenze T)
i) Temperatur verlauf am Tage x an einer Wetterstation	$\{\omega: [0,24) \to [-273,\infty);$ $\omega \text{ stetige Funktion }\}$	Temperatur verlauf von 0 Uhr bis 24 Uhr in ° Celsius, $\omega(t)$ ist Temperatur zur Zeit t .

Ereignisse sind Teilmengen des Grundraums Ω . Damit können die für Mengen definierten Verknüpfungen auf Ereignisse angewendet werden. Als Abkürzung für die Aussage Das Ergebnis ω ist im Ereignis A enthalten wird kurz $\omega \in A$ geschrieben. Seien A, B, C und $A_i, i \in I$, Ereignisse, wobei I eine Indexmenge ist.

- Schnittereignis zweier Mengen: $A \cap B = \{\omega \in \Omega; \ \omega \in A \ \text{und} \ \omega \in B\}$ Gilt für die Ereignisse A und B: $A \cap B = \emptyset$, so heißen A und B disjunkte Ereignisse.
- Schnittereignis beliebig vieler Mengen: $\bigcap_{i \in I} A_i = \{\omega \in \Omega; \ \omega \in A_i \ \text{für jedes} \ i \in I\}$ Die Ereignisse $A_i, \ i \in I$, heißen **paarweise disjunkt**, falls für jeweils zwei verschiedene Indizes $i, j \in I$ gilt: $A_i \cap A_j = \emptyset$.
- Vereinigungsereignis zweier Mengen: $A \cup B = \{\omega \in \Omega; \ \omega \in A \ \mathbf{oder} \ \omega \in B\}$
- Vereinigungsereignis beliebig vieler Mengen: $\bigcup_{i \in I} A_i = \{\omega \in \Omega; \text{ es gibt ein } i \in I \text{ mit } \omega \in A_i\}$

- Teilereignis: $A \subset B$: Für jedes $\omega \in A$ gilt $\omega \in B$.
- Komplementärereignis: A
 = Ω\A = {ω ∈ Ω; ω ∉ A} = {ω ∈ Ω; ω nicht Element von A};
 A heißt Komplement von A (in Ω).
 Manche Autoren benutzen auch A^c anstelle von Ā.
- Differenzereignis: $B \setminus A = \{ \omega \in \Omega; \ \omega \in B \ \text{und} \ \omega \notin A \} = B \cap \overline{a} = B \setminus (A \cap B);$ $B \setminus A \ \text{heißt} \ \textbf{Komplement} \ \text{von} \ A \ \text{in} \ B.$
- Kommutativgesetze:

$$A \cup B = B \cup A$$
, $A \cap B = B \cap A$.

• Assoziativgesetze:

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C).$$

• Distributivgesetze:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C),$$
 $(A \cap B) \cup C = (A \cup C) \cap (B \cup C).$

• Regeln von de Morgan:

$$\overline{A \cap B} = \overline{A} \cup \overline{B}, \qquad \overline{A \cup B} = \overline{A} \cap \overline{B}.$$

Die Sprechweisen sind in folgender Tabelle zusammengestellt.

Mathematisches Objekt	Interpretation
Ω	Grundraum, Ergebnisraum
$\omega \in \Omega$	(mögliches) Ergebnis
A	Ereignis
\mathscr{A}	Menge aller Ereignisse
Ω	sicheres Ereignis
Ø	unmögliches Ereignis
$\omega \in A$	Ereignis A tritt ein
$\omega \in \overline{A}$	Ereignis A tritt nicht ein
$\omega \in A \cup B$	Ereignis A oder Ereignis B tritt ein
$\omega \in A \cap B$	Ereignis A und Ereignis B treten ein
$A \subset B$	Eintreten von Ereignis A impliziert das Eintreten von Ereignis
	B
$A \cap B = \emptyset$	Ereignisse A und B schließen einander aus
$\omega \in \bigcup A_i$	mindestens ein Ereignis A_i , $i \in I$, tritt ein
$\omega \in \bigcap_{i \in I} A_i$	alle Ereignisse A_i , $i \in I$, treten ein

1.3.4 Beispiel

Gegeben sei das Zufallsexperiment eines Würfelwurfs mit zwei symmetrischen Würfeln. Dann ist der Grundraum Ω gegeben durch

$$\Omega = \{(i,j); i,j=1,\ldots,6\} = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(5,6),(6,1),\ldots,(6,6)\}.$$

Wir betrachten nun die Ereignisse A, B, C und D mit A = erste Ziffer gerade, B = Summe beider Ziffern ist gerade, C = erste Ziffer ist eine Sechs und D = beide Ziffern sind gleich. Dann gilt:

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

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

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

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

Dann gilt etwa:

$$A \cap B = \{(2,2), (2,4), (2,6), (4,2), (4,4), (4,6), (6,2), (6,4), (6,6)\}$$

$$A \cup B = \Omega \setminus \{(1,2), (1,4), ((1,6), (3,2), (3,4), (3,6), (5,2), (5,4), (5,6)\}$$

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

$$C \setminus A = \emptyset$$

$$D \cap B = D$$

$$C \cap D = \{(6,6)\}$$

2 Laplace-Experimente

2.1 Laplace'scher Wahrscheinlichkeitsraum

2.1.1 Beispiel

Wie groß ist die Wahrscheinlichkeit, bei einem Würfelwurf eine gerade Zahl zu würfeln?

 $\Omega = \{1, 2, 3, 4, 5, 6\}$ Menge der möglichen Ergebnisse

 $A = \{2, 4, 6\}$ interessierendes Ereignis, dass eine gerade Zahl gewürfelt wird.

 $P(A) = \frac{\sharp A}{\sharp \Omega} = \frac{3}{6} = \frac{1}{2}$ ist die Wahrscheinlichkeit eine gerade Zahl zu würfeln.

Dabei bedeutet $\sharp A$ die Anzahl der Elemente der Menge A.

2.1.2 Definition (Laplace'scher Wahrscheinlichkeitsraum)

Ist Ω eine endliche Menge, und $P: \mathscr{P}(\Omega) \ni A \to P(A) := \frac{\sharp A}{\sharp \Omega} \in [0,1]$, dann heißt $(\Omega, \mathscr{P}(\Omega), P)$ Laplace'scher Wahrscheinlichkeitsraum.

2.1.3 Bemerkung

- (i) Ω ist die Ergebnismenge. $\omega \in \Omega$ heißt Ergebnis.
- (ii) $\mathscr{P}(\Omega) := \{A; A \subset \Omega\}$, die Potenzmenge von Ω , ist das System der interessierenden Ereignisse.

 $A \in \mathscr{P}(\Omega)$ heißt Ereignis. $\{\omega\} \in \mathscr{P}(\Omega)$ heißt Elementar-Ereignis.

- (iii) Die Abbildung P ist das Laplace'sche Wahrscheinlichkeitsmaß.
- (iv) Der Laplace'sche Wahrscheinlichkeitsraum ist eindeutig durch die Ergebnismenge Ω bestimmt. Speziell besitzt jedes Elementarereignis $\{\omega\}$ die Wahrscheinlichkeit

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

2.1.4 Lemma (Fundamentale Eigenschaften des Laplace'schen Wahrscheinlichkeitsraums) $(\Omega, \mathscr{P}(\Omega), P)$ Laplace'scher Wahrscheinlichkeitsraum

 \Rightarrow

- (i) $P(\Omega) = \frac{\sharp \Omega}{\sharp \Omega} = 1$ ("Normierung")
- $$\begin{split} \text{(ii)} \quad &A,B \in \mathscr{P}(\Omega) \text{ disjunkt, d.h. } A \cap B = \emptyset, \\ \Rightarrow &P(A \uplus B) = \frac{\sharp (A \uplus B)}{\sharp \, \Omega} = \frac{\sharp \, A + \sharp \, B}{\sharp \, \Omega} = P(A) + P(B) \qquad \text{("Additivit")} \end{split}$$

Dabei wurde $A \uplus B := A \cup B$ als Notation für die Vereinigung disjunkter Mengen benutzt.

2.1.5 Lemma (Weitere Eigenschaften)

- (i) $P(\emptyset) = 0$
- (ii) $P(\overline{A}) = 1 P(A)$
- (iii) $A \subseteq B \Rightarrow P(A) \leq P(B)$ und $P(B \setminus A) = P(B) P(A)$
- (iv) $P(A \cup B) = P(A) + P(B) P(A \cap B)$

Beweis

(i)
$$P(\Omega) = P(\Omega \uplus \emptyset) = P(\Omega) + P(\emptyset)$$

(ii)
$$1 = P(\Omega) = P(A \uplus \overline{A}) = P(A) + P(\overline{A})$$

(iii)
$$A \subseteq B \Rightarrow B = A \uplus (B \setminus A) \Rightarrow P(B) = P(A) + P(B \setminus A) \ge P(A) \Rightarrow P(B \setminus A) = P(B) - P(A)$$

(iv)
$$P(A \cup B) = P(A \uplus (B \setminus (A \cap B))) = P(A) + P(B \setminus (A \cap B))$$

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

#

2.1.6 Beispiel (Das Problem des Chevalier de Meré, Fortsetzung von Beispiel 1.1.1)
Analog zu Beispiel 2.1.1 wird der dreifache Würfelwurf durch den Grundraum

$$\Omega = \{(i, j, k); i, j, k = 1, \dots, 6\}$$

modelliert. P bezeichne die Laplace-Verteilung auf Ω , d.h., es wird angenommen, jedes Ergebnis (i, j, k) besitzt die gleiche Wahrscheinlichkeit. Wir betrachten die Ereignisse A = Augensumme 11 und B = Augensumme 12. Zur Berechnung der Wahrscheinlichkeit schreiben wir diese Ereignisse als Teilmengen von Ω und erhalten:

$$\begin{split} A &= \big\{ (i,j,k) \in \Omega; \ i+j+k=11 \big\} \\ &= \big\{ (1,4,6), (1,6,4), (4,1,6), (4,6,1), (6,1,4), (6,4,1), (1,5,5), (5,1,5), (5,5,1), (2,3,6), (2,6,3), (3,2,6), (3,6,2), (6,2,3), (6,3,2), \\ &\qquad \qquad (2,4,5), (2,5,4), (4,2,5), (4,5,2), (5,2,4), (5,4,2), (3,3,5), (3,5,3), (5,3,3), (3,4,4), (4,3,4), (4,4,3) \big\}, \\ B &= \big\{ (i,j,k) \in \Omega; \ i+j+k=12 \big\} \\ &= \big\{ (1,5,6), (1,6,5), (5,1,6), (5,6,1), (6,1,5), (6,5,1), (2,4,6), (2,6,4), (4,2,6), (4,6,2), (6,2,4), (6,4,2), (2,5,5), (5,2,5), (5,5,2), \\ &\qquad \qquad (3,3,6), (3,6,3), (6,3,3), (3,4,5), (3,5,4), (4,3,5), (4,5,3), (5,3,4), (5,4,3), (4,4,4) \big\}. \end{split}$$

Daher gilt $\sharp A=27$ und $\sharp B=25$, so dass wegen $\sharp \Omega=6^3=216$: $P(A)=\frac{27}{216}=\frac{1}{8}=0.125$ und $P(B)=\frac{25}{216}=0.116$ folgt. Das Ereignis A hat somit eine höhere Eintrittswahrscheinlichkeit. De Meré hatte diesen Unterschied bemerkt, konnte ihn aber nicht nachweisen.

2.2 Bestimmung der Kardinalität einer Menge

2.2.1 Satz (Grundabzählprinzip)

Ist A eine Menge von n-Tupeln (a_1, \ldots, a_n) mit:

- (i) Es gibt N_1 Möglichkeiten, a_1 einen Wert zuzuordnen.
- (ii) Nach Festlegung von a_1, \ldots, a_K gibt es N_{K+1} Möglichkeiten a_{K+1} einen Wert zuzuordnen, $1 \le K < n$.

Dann gilt $\sharp A = N_1 \cdot N_2 \cdot \ldots \cdot N_n$.

Beweis

Durch vollständige Induktion.

#

2.2.2 Urnenmodelle

Eine Urne enthält N Kugeln mit den Nummern 1, 2, ..., N. Es wird n-mal aus der Urne gezogen. Wie viel mögliche Ergebnisse gibt es?

(i) Ziehen mit Zurücklegen mit Berücksichtigung der Reihenfolge

$$\Omega = \{(\omega_1, \dots, \omega_n); \ \omega_k \in \{1, \dots, N\}, \ k = 1, \dots, n\}$$

$$\Rightarrow \sharp \Omega = \underbrace{N \cdot N \cdot N \cdot \dots \cdot N}_{n-\text{mal}} = N^n.$$

(ii) Ziehen ohne Zurücklegen mit Berücksichtigung der Reihenfolge

$$\Omega = \{(\omega_1, \dots, \omega_n); \, \omega_k \in \{1, \dots, N\}, \, k = 1, \dots, n, \, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

$$\Rightarrow \sharp \, \Omega = N \cdot (N-1) \cdot (N-2) \cdot \dots \cdot (N-n+1) = \frac{N!}{(N-n)!}$$

(iii) Ziehen ohne Zurücklegen ohne Berücksichtigung der Reihenfolge

$$\Omega = \{(\omega_1, \dots, \omega_n); \, \omega_k \in \{1, \dots, N\}, \, k = 1, \dots, n, \, \omega_1 < \omega_2 < \dots < \omega_n\}$$

Ein Tupel $(\omega_1, \ldots, \omega_n)$ mit paarweise verschiedenen Komponenten kann auf n! Weisen umsortiert (permutiert) werden.

$$\Rightarrow \sharp \Omega = \frac{N!}{(N-n)!n!} =: \binom{N}{n}$$

 $\binom{N}{n}$ ist damit die Anzahl der Möglichkeiten eine n-elementige Teilmenge von einer N-elementigen Menge zu bilden.

(iv) Ziehen mit Zurücklegen ohne Berücksichtigung der Reihenfolge

$$\Omega = \{(\omega_1, \dots, \omega_n); \ \omega_k \in \{1, \dots, N\}, \ k = 1, \dots, n, \ \omega_1 \le \omega_2 \le \dots \le \omega_n\}$$

Betrachten wir die Abbildung

$$S: \Omega \ni (\omega_1, \dots \omega_n) \to (\omega_1 + 1, \omega_2 + 2, \dots, \omega_n + n) \in \Omega'$$

mit

$$\Omega' = \{(\omega_1', \dots, \omega_n'); \ \omega_k' \in \{2, \dots, N+n\}, \ k = 1, \dots, n, \ \omega_1' < \omega_2' < \dots < \omega_n'\},\$$

dann ist S eine Bijektion, also

$$\sharp \Omega = \sharp \Omega' = \binom{N+n-1}{n}$$

nach (iii).

Zusammenfassung der Urnenmodelle (mit R-Befehlen)

	Mit Berücksichtigung	Ohne Berücksichtigung
	der Reihenfolge	der Reihenfolge
Mit Zurücklegen	N^n	$\binom{N+n-1}{n}$
	N^n	choose(N+n-1,n)
Ohne Zurücklegen	$\frac{N!}{(N-n)!}$	$\binom{N}{n}$
	factorial(N)/factorial(N-n)	choose(N,n)

2.2.3 Beispiel (Geburtstagsproblem)

Wie groß ist die Wahrscheinlichkeit p, dass mindestens zwei von den n=25 Schülern einer Klasse am gleichen Tag Geburtstag haben?

2.3 Bildmaße

2.3.1 Beispiel

Bei einem Würfel mit je einer "1" und "2" und je zweimal "3" und "4" ist $\Omega' = \{1, 2, 3, 4\}$ die Ergebnismenge, doch es liegt kein Laplace'scher Wahrscheinlichkeitsraum vor, da die Ergebnisse nicht gleichwahrscheinlich sind. Einen Laplace-Raum erhält man nur, wenn man bei "3" und "4" berücksichtigt, auf welcher Seite des Würfels sie sich befinden, d.h. wenn man z. B. $\Omega = \{1, 2, 3a, 3b, 4a, 4b\}$ betrachtet. Durch die Abbildung

$$X:\Omega\longrightarrow\Omega'$$

mit $X(1)=1, \ X(2)=2, \ X(3a)=3, \ X(3b)=3, \ X(4a)=4, \ X(4b)=4$ werden die Wahrscheinlichkeiten des Raumes $(\Omega, \mathscr{P}(\Omega))$ auf den Raum $(\Omega', \ \mathscr{P}(\Omega'))$ übertragen. Ist P ein Wahrscheinlichkeitsmaß auf $(\Omega, \ \mathscr{P}(\Omega))$, so sagt man, dass $X:\Omega\to\Omega'$ ein Wahrscheinlichkeitsmaß P^X auf $(\Omega', \mathscr{P}(\Omega'))$ induziert. P^X heißt auch Bildmaß von P unter X und X heißt Zufallsvariable.

2.3.2 Beispiel (Dreifacher Würfelwurf)

Von Interesse ist die Summe der Augen beim dreifachen Würfelwurf.

2.3.3 **Definition** (Bildmaß)

Sei P ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{P}(\Omega)), X : \Omega \longrightarrow \Omega', \mathcal{B} \subseteq \mathcal{P}(\Omega'),$ so heißt

$$\begin{split} P^X: \mathscr{B} \ni B &\to P^X(B) := P(X^{-1}(B)) := P(X \in B) \\ &:= P(\{\omega \in \Omega; X(\omega) \in B\}) \in [0,1] \end{split}$$

das Bildmaß von P unter X und X heißt Zufallsvariable.

2.3.4 Lemma (Fundamentale Eigenschaften von P^X)

(i)
$$P^X(\Omega') = P(X \in \Omega') = P(\Omega) = 1$$
 "Normierung"

(ii)
$$B_1, B_2 \in \mathcal{B}, B_1 \cap B_2 = \emptyset$$

 $\Rightarrow P^X(B_1 \uplus B_2) = P(X \in (B_1 \uplus B_2))$
 $= P((X \in B_1) \uplus (X \in B_2)) = P(X \in B_1) + P(X \in B_2)$
 $= P^X(B_1) + P^X(B_2)$ "Additivität"

18

2.3.5 Lemma (Weitere Eigenschaften von P^X)

(i)
$$P^X(\emptyset) = 0$$

(ii)
$$P^X(\overline{B}) = 1 - P^X(B)$$

(iii)
$$B_1 \subseteq B_2 \Rightarrow P^X(B_1) \leq P^X(B_2)$$

(iv)
$$P^X(B_1 \cup B_2) = P^X(B_1) + P^X(B_2) - P^X(B_1 \cap B_2)$$

Beweis

Siehe Lemma 2.1.5. #

2.3.6 Beispiel

Indikatorfunktionen

Für $A \in \mathscr{P}(\Omega)$ definieren wir die Indikatorfunktion

$$1_A: \Omega \ni \omega \to 1_A(\omega) = \left\{ \begin{array}{ll} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{array} \right\} \in \{0, 1\}$$

 1_A ist Zufallsvariable und

$$P^{1_A}(\{1\}) = P(A)$$

$$P^{1_A}(\{0\}) = P(\overline{A}) = 1 - P(A).$$

 1_A zeigt an, ob das Ereignis A eingetreten ist.

2.4 Aus Laplace-Experimenten abgeleitete Wahrscheinlichkeitsmaße

2.4.1 Hypergeometrische Verteilung

In einer Urne befinden sich Kugeln mit den Nummern $1, \ldots, M$, wobei die Kugeln mit den Nummern $1, \ldots, R$ rot sind und die Kugeln mit den Nummern $R+1, \ldots, M$ schwarz sind. D.h. es gibt R rote und S=M-R schwarze Kugeln in der Urne. Es wird N-mal ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, dass bei den N Ziehungen r rote Kugeln gezogen werden?

$$\Omega = \{(\omega_1, \dots, \omega_N); \ \omega_n \in \{1, \dots, M\}, \ n = 1, \dots, N, \ \omega_1 < \omega_2 < \dots < \omega_N\}$$

$$X((\omega_1, \dots, \omega_N)) := \sum_{n=1}^N 1_{\{1, \dots, R\}}(\omega_n) = \text{Anzahl der roten Kugeln in } (\omega_1, \dots, \omega_N)$$

$$\Rightarrow X : \Omega \to \Omega' = \{0, 1, \dots, N\}$$

Gesucht ist

$$P^X(\{r\})$$
 mit $r \in \{0, 1, \dots, N\}$.

$$\Rightarrow P^X(\{r\}) = \frac{\binom{R}{r}\binom{M-R}{N-r}}{\binom{M}{N}} = \frac{\binom{R}{r}\binom{S}{N-r}}{\binom{R+S}{N}} = P_{H(N,R,M-R)}(\{r\}) = P_{H(N,R,S)}(\{r\}).$$

Dabei vereinbaren wir für den Binomialkoeffizienten:

$$\binom{m}{n} = 0$$
, falls $n < 0$ oder $n > m$.

Die Wahrscheinlichkeiten

$$P_{H(N,R,S)}(\{r\}) = \frac{\binom{R}{r}\binom{S}{N-r}}{\binom{R+S}{N}}, r = 0, \dots, N,$$

ergeben die Hypergeometrische Verteilung.

2.4.2 Binomialverteilung

Sei die Urne wie in Abschnitt 2.4.1, nur dass jetzt mit Zurücklegen gezogen wird, d.h.

$$\Omega = \{(\omega_1, \dots, \omega_N); \ \omega_n \in \{1, \dots, M\}, \ n = 1, \dots, N\}$$

ist. Die Abbildung X sei wie in Abschnitt 2.4.1 definiert und gesucht ist $P^X(\{r\})$, die Wahrscheinlichkeit, das r rote Kugeln gezogen werden, $r=0,1,\ldots,n$.

Die Wahrscheinlichkeiten

$$P_{B(N,p)}(\{r\}) = \binom{N}{r} p^r (1-p)^{N-r}, r = 0, 1, \dots, N,$$

ergeben die Binominalverteilung.

2.4.3 Beispiel

Es wird solange gewürfelt, bis die erste "6" erscheint. Wie groß ist die Wahrscheinlichkeit, dass dies im $1, 2, \ldots, n$ -ten Wurf geschieht?

2.4 Aus Laplace-Experimenten abgeleitete Wahrscheinlichkeitsmaße

$$\begin{split} &\Omega_N := \{(\omega_1, \dots, \omega_N); \omega_n \in \{1, \dots, 6\}, n = 1, \dots, N\} \\ &\Rightarrow \sharp \, \Omega_N = 6^N \\ &X_N : \Omega_N \ni \omega = (\omega_1, \dots, \omega_N) \to X_N(\omega) := \inf\{k; \omega_k = 6\} \in \{1, 2, \dots\} \cup \{\infty\} = \Omega' \\ &\text{Für } n \in \{1, \dots, N\} \text{ sei} \\ &A_{N,n} := \{X_N = n\} = \{(\omega_1, \dots, \omega_N) \in \Omega_N; \, \omega_k \neq 6 \text{ für } k = 1, \dots, n - 1, \, \omega_n = 6\}. \\ &\Rightarrow \sharp \, A_{N,n} = 5^{n-1} \cdot 1 \cdot 6^{N-n} \\ &\Rightarrow P^{X_N}(\{n\}) = P(A_{N,n}) = \frac{\sharp A_{N,n}}{\sharp \, \Omega_N} = \frac{5^{n-1}6^{N-n}}{6^N} = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}. \end{split}$$

Da diese Wahrscheinlichkeit nicht von N abhängt, liegt es nahe, eine Verteilung P^X für die Anzahl des Würfels bis zur ersten "6" wie folgt zu definieren:

$$P^X(\{n\}) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}$$

für alle $n \in \mathbb{N}$. Dabei müsste ein Wahrscheinlichkeitsraum mit einem Ω existieren, so dass

$$X:\Omega\longrightarrow\mathbb{N}$$

das Bildmaß P^X ergibt. Da $\mathbb N$ unendlich ist, wäre dann auch Ω unendlich.

3 Wahrscheinlichkeitsräume und zufällige Bilder

3.1 Allgemeine Wahrscheinlichkeitsräume

Interessierende Objekte des Wahrscheinlichkeitsraums sind:

- (i) Ω : Menge der möglichen **Ergebnisse** (des Zufallsexperimentes)
- (ii) $\mathscr{A} \subseteq \mathscr{P}(\Omega)$: System von interessierenden **Ereignissen**
- (iii) $P: \mathcal{A} \ni A \to P(A) \in [0,1]$: Wahrscheinlichkeitsmaß

3.1.1 Definition (σ -Algebra))

Ist $\Omega \neq \emptyset$ eine Menge, dann heißt ein Mengensystem $\mathscr{A} \subset \mathscr{P}(\Omega)$ eine σ -Algebra über Ω

 \Leftrightarrow

- (i) $\Omega \in \mathscr{A}$
- (ii) $A \in \mathscr{A} \Rightarrow \overline{A} \in \mathscr{A}$
- (iii) $(A_n \in \mathscr{A})_{n \in \mathbb{N}} \Rightarrow \bigcup_{n \geq 1} A_n \in \mathscr{A}.$

3.1.2 Lemma (Erste Eigenschaft)

Eine σ -Algebra \mathscr{A} ist abgeschlossen bzgl. abzählbarer Mengenoperationen, z. B.

$$(A_n \in \mathscr{A})_{n \in \mathbb{N}} \stackrel{(ii)}{\Rightarrow} (\overline{A}_n \in \mathscr{A})_{n \in \mathbb{N}}$$

$$\stackrel{(iii)}{\Rightarrow} \bigcup_{n \geq 1} \overline{A}_n \in \mathscr{A}$$

$$\stackrel{(ii)}{\Rightarrow} \bigcap_{n \geq 1} A_n = \overline{\bigcup_{n \geq 1} \overline{A}_n} \in \mathscr{A}.$$

3.1.3 Satz (Erzeugte σ -Algebra)

Ist $\mathscr{E} \subseteq \mathscr{P}(\Omega)$, dann gibt es eine kleinste von \mathscr{E} erzeugte σ -Algebra $\sigma(\mathscr{E})$, d.h.

- (i) $\sigma(\mathcal{E})$ ist σ -Algebra,
- (ii) $\mathscr{E} \subseteq \sigma(\mathscr{E})$
- (iii) $\mathscr{E} \subseteq \mathscr{A}$ und \mathscr{A} σ -Algebra $\Rightarrow \sigma(\mathscr{E}) \subseteq \mathscr{A}$.

Beweis

Betrachte $\sigma(\mathscr{E}) := \bigcap_{\mathscr{A} \in \mathbb{A}(\mathscr{E})} \mathscr{A}$, wobei $\mathbb{A}(\mathscr{E}) = \{\mathscr{A}; \ \mathscr{A} \ \sigma\text{-Algebra und} \ \mathscr{E} \subset \mathscr{A} \ \}.$

Dann gilt $\mathbb{A}(\mathcal{E}) \neq \emptyset$, da $\mathscr{P}(\Omega)$ σ -Algebra ist und $\mathcal{E} \subset \mathscr{P}(\Omega)$. Damit ist $\sigma(\mathcal{E})$ wohldefiniert.

#

Es bleibt zu zeigen, dass $\sigma(\mathcal{E})$ eine σ -Algebra gemäß (2.1) ist:

(i)
$$\bigwedge_{\mathscr{A}\in\mathbb{A}(\mathscr{E})}\Omega\in\mathscr{A}\Rightarrow\Omega\in\sigma(\mathscr{E})$$

(ii)
$$A \in \sigma(\mathscr{E}) \Rightarrow \bigwedge_{\mathscr{A} \in \mathbb{A}(\mathscr{E})} A \in \mathscr{A} \Rightarrow \bigwedge_{\mathscr{A} \in \mathbb{A}(\mathscr{E})} \overline{A} \in \mathscr{A} \Rightarrow \overline{A} \in \sigma(\mathscr{E})$$

(iii)
$$(A_n \in \sigma(\mathscr{E}))_{n \in \mathbb{N}} \Rightarrow \bigwedge_{\mathscr{A} \in \mathbb{A}(\mathscr{E})} (A_n \in \mathscr{A})_{n \in \mathbb{N}} \Rightarrow \bigwedge_{\mathscr{A} \in \mathbb{A}(\mathscr{E})} \bigcup_{n \geq 1} A_n \in \mathscr{A} \Rightarrow \bigcup_{n \geq 1} A_n \in \sigma(\mathscr{E}).$$

3.1.4 Beispiel

- (i) Ω abzählbar (endlich oder unendlich) $\mathscr{E} = \{\{\omega\}; \ \omega \in \Omega\} \text{ das System der Elementarereignisse}$ $\Rightarrow \sigma(\mathscr{E}) = \mathscr{P}(\Omega).$
- (ii) Ω überabzählbar $\mathscr{E} = \{\{\omega\}; \ \omega \in \Omega\}$ $\Rightarrow \sigma(\mathscr{E}) = \{A; \ A \in \mathscr{P}(\Omega), \ A \text{ abzählbar oder } \overline{A} \text{ abzählbar } \}.$

3.1.5 Definition (Borel- σ -Algebra)

Sei $\Omega = \mathbb{R}$ (bzw. $\Omega = \mathbb{R}^p$) und $\mathscr{E} = \{[a,b]; a,b \in \mathbb{R}, a \leq b\}$ (bzw. $\mathscr{E}_p = \{[a,b]; a = (a_1,\ldots,a_p)^\top, b = (b_1,\ldots,b_p)^\top \in \mathbb{R}^p, a \leq b\}$, wobei $a \leq b$ gilt, falls $a_i \leq b_i$ für $i = 1,\ldots,p$ gilt). Dann heißt $\mathscr{B}_1 := \sigma(\mathscr{E})$ (bzw. $\mathscr{B}_p := \sigma(\mathscr{E}_p)$) die **Borel**- σ -Algebra über \mathbb{R} (bzw. \mathbb{R}^p).

3.1.6 Bemerkung

 \mathcal{B}_p lässt sich auch durch das System

$$\mathscr{E}'_p = \{(-\infty, b]; b \in \mathbb{R}^p\}$$

erzeugen, d. h. $\mathscr{B}_p = \sigma(\mathscr{E}_p')$, denn: Für jedes $(-\infty,b] \in \mathscr{B}_p$ gilt:

$$\mathscr{E}'_p \ni (-\infty, b] = \bigcup_{n>1} [-n \ 1_p \ , \ b] \in \sigma(\mathscr{E}_p)$$

mit $1_p = (1, \dots, 1)^\top \in \mathbb{R}^p$ und der Konvention $[a, b] := \emptyset$, falls nicht $a \leq b$ gilt. Somit gilt $\sigma(\mathscr{E}'_p) \subseteq \mathscr{B}_p$. Andererseits gilt für jedes $[a, b] \in \mathscr{E}_p$:

$$[a,b] = \bigcap_{n>1} (a^{(n)},b] \text{ mit } a^{(n)} = a - \frac{1}{n} 1_p \text{ und}$$

$$(a^{(n)},\,b]=(-\infty,b]\backslash \bigcup_{\pi=1}^p(-\infty,a^{(n)}(\pi)]\in\sigma(\mathscr{E}'_p),$$

wobei
$$a^{(n)}(\pi) = (a_1^{(n)}(\pi), \dots, a_p^{(n)}(\pi))^T$$
 mit $a_{\varrho}^{(n)}(\pi) = \begin{cases} b_{\varrho} & \text{für } \varrho \neq \pi \\ a_{\pi}^{(n)} & \text{für } \varrho = \pi \end{cases}$.

Folglich $[a,b] \in \sigma(\mathscr{E}'_p)$ und damit $\mathscr{B}_p \subseteq \sigma(\mathscr{E}'_p)$.

Weitere Erzeuger von \mathscr{B}_p sind z.B. (dabei gelte $a < b \Leftrightarrow a_i < b_i$ für $i = 1, \ldots, p$):

(i)
$$\mathscr{E}_p^1 = \{(a,b); \ a,b \in \mathbb{R}^p, \ a < b\} = \{\{x \in \mathbb{R}^p; \ a < x < b\}, \ a,b \in \mathbb{R}^p \ a < b\},$$

(ii)
$$\mathscr{E}_p^2 = \{(a, b]; \ a, b \in \mathbb{R}^p, \ a \le b\} = \{\{x \in \mathbb{R}^p; \ a < x \le b\}, \ a, b \in \mathbb{R}^p \ a \le b\},\ a \in \mathbb{R}^p \ a \le b\}$$

(iii)
$$\mathscr{E}_p^3 = \{(-\infty, b); b \in \mathbb{R}^p\},\$$

(iv)
$$\mathscr{E}_p^4 = \{O; O \subset \mathbb{R}^p \text{ ist offene Menge}\}.$$

Eine Menge $O \subset \mathbb{R}^p$ ist dabei offen, wenn für jedes $x \in O$ ein offenes p-dimensionales Intervall (a_x, b_x) existiert mit $x \in (a_x, b_x) \subset O$. Damit gilt $\mathscr{E}_p^1 \subset \mathscr{E}_p^4$ und somit $\sigma(\mathscr{E}_p^1) \subset \sigma(\mathscr{E}_p^4)$. Umgekehrt gilt auch $\mathscr{E}_p^4 \subset \sigma(\mathscr{E}_p^1)$, denn jedes $O \in \mathscr{E}_p^4$ ist eine abzählbare Vereinigung von offenen Intervallen (a_n, b_n) , wobei a_n und b_n aus rationalen Komponenten bestehen, d.h. $a_n, b_n \in \mathbb{Q}^p$. Ist nämlich $x \in (a_x, b_x) \subset O$ so gibt es $q(x), a_{q(x)}, b_{q(x)} \in \mathbb{Q}^p$ mit $x \in (a_{q(x)}, b_{q(x)}) \subset O$. Da die rationalen Zahlen abzählbar sind, ist somit O eine abzählbare Vereinigung von Elementen aus \mathscr{E}_p^1 .

3.1.7 Bemerkung

In der Maß- und Integrationstheorie wird mit Hilfe des Auswahlaxiomes gezeigt, dass die Borel- σ -Algebra \mathscr{B}_1 (bzw. \mathscr{B}_p) eine echte Teilmenge von $\mathscr{P}(\mathbb{R})$ (bzw. $\mathscr{P}(\mathbb{R}^p)$) ist.

3.1.8 Definition (Wahrscheinlichkeitsmaß)

Ist $\Omega \neq \emptyset$, \mathscr{A} eine σ -Algebra über Ω , dann heißt $P: \mathscr{A} \ni A \to P(A) \in [0,1]$ ein Wahrscheinlichkeitsmaß auf \mathscr{A}

 \Leftrightarrow

(i)
$$P(\Omega) = 1$$
 "Normierung",

(ii)
$$(A_n \in \mathscr{A})_{n \in \mathbb{N}}$$
 (paarweise) disjunkt, d.h. $A_i \cap A_j = \emptyset$ für $i \neq j$

$$\Rightarrow P(\biguplus_{n \geq 1} A_n) = \sum_{n \geq 1} P(A_n) \qquad \text{``} \sigma\text{-}Additivit".$$

3.1.9 Lemma

- (i) $P(\emptyset) = 0$,
- (ii) A_1, \dots, A_N disjunkt $\Rightarrow P\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N P(A_n),$ speziell

$$A_1, A_2 \text{ disjunkt} \Rightarrow P(A_1 \uplus A_2) = P(A_1) + P(A_2),$$

- (iii) $P(\overline{A}) = 1 P(A)$,
- (iv) $A_1 \subseteq A_2 \Rightarrow P(A_1) \leq P(A_2)$,
- (v) $P(A_1 \cup A_2) = P(A_1) + P(A_2) P(A_1 \cap A_2)$

Beweis

(i)
$$A_1 := \Omega, A_n := \emptyset \text{ für } n \ge 2 :$$

$$P(\Omega) = P\left(\biguplus_{n \ge 1} A_n\right) = P(A_1) + \sum_{n \ge 2} P(A_n)$$

$$= P(\Omega) + \sum_{n \ge 2} P(\emptyset)$$

$$\Rightarrow P(\emptyset) = 0.$$

(ii)
$$A_n = \emptyset$$
 für $n > N$:

$$P\left(\biguplus_{n=1}^N An\right) = P\left(\biguplus_{n \ge 1} A_n\right) = \sum_{n \ge 1} P(A_n)$$

$$= \sum_{n=1}^N P(A_n) + \sum_{n \ge N+1} P(\emptyset) = \sum_{n=1}^N P(A_n).$$

(iii)-(v) folgen aus der Normierung und (ii) wie im Beweis von Lemma 2.1.5.

3.1.10 Satz (Poincaré-Sylvester-Formel)

(Formel des Ein- und Ausschließens) (Verallgemeinerung von (3.1.9)(v))

(i)
$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{m=1}^{N} (-1)^{m+1} \sum_{1 \le n_1 \le \dots \le n_m \le N} P(A_{n_1} \cap \dots \cap A_{n_m}).$$

(ii) Speziell für N=3:

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$
$$-P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3)$$
$$+P(A_1 \cap A_2 \cap A_3).$$

(iii) Gilt
$$P(A_{n_1} \cap \ldots \cap A_{n_m}) = p_m$$
 für alle $1 \le n_1 < \ldots < n_m \le N, \ m = 1, \ldots, N,$

dann vereinfacht sich die Formel zu:

$$P\left(\bigcup_{n=1}^{N} A_n\right) = \sum_{m=1}^{N} (-1)^{m+1} \binom{N}{m} p_m.$$

Beweis

Für (i) vollständige Induktion: Die Formel gilt für N=2 nach (3.1.9)(v).

Induktionsschritt $N \to N+1$:

$$P\left(\bigcup_{n=1}^{N+1} A_{n}\right) \stackrel{(3.1.9)(v)}{=} P\left(\bigcup_{n=1}^{N} A_{n}\right) + P(A_{N+1}) - P\left(\left(\bigcup_{n=1}^{N} A_{n}\right) \cap A_{N+1}\right)$$

$$= P\left(\bigcup_{n=1}^{N} A_{n}\right) + P(A_{N+1}) - P\left(\bigcup_{n=1}^{N} (A_{n} \cap A_{N+1})\right)$$

$$\operatorname{Ind.Vor.} \sum_{m=1}^{N} (-1)^{m+1} \sum_{1 \leq n_{1} < \dots < n_{m} \leq N} P(A_{n_{1}} \cap \dots \cap A_{n_{m}}) + P(A_{N+1})$$

$$- \sum_{m=1}^{N} (-1)^{m+1} \sum_{1 \leq n_{1} < \dots < n_{m} \leq N} P(A_{n_{1}} \cap \dots \cap A_{n_{m}} \cap A_{N+1})$$

$$= \sum_{n=1}^{N+1} P(A_{n}) + \sum_{m=2}^{N} (-1)^{m+1} \sum_{1 \leq n_{1} < \dots < n_{m} \leq N} P(A_{n_{1}} \cap \dots \cap A_{n_{m}})$$

$$- \sum_{m=2}^{N} (-1)^{m} \sum_{1 \leq n_{1} < \dots < n_{m-1} < n_{m} = N+1} P(A_{n_{1}} \cap \dots \cap A_{n_{m}})$$

$$- (-1)^{(N+1)} P(A_{1} \cap \dots \cap A_{N+1})$$

$$= \sum_{n=1}^{N+1} P(A_n) + \sum_{m=2}^{N} (-1)^{m+1} \sum_{1 \le n_1 < \dots < n_m \le N} P(A_{n_1} \cap \dots \cap A_{n_m})$$

$$+ \sum_{m=2}^{N} (-1)^{m+1} \sum_{1 \le n_1 < \dots < n_{m-1} < n_m = N+1} P(A_{n_1} \cap \dots \cap A_{n_m})$$

$$+ (-1)^{(N+1)+1} P(A_1 \cap \dots \cap A_{N+1})$$

$$= \sum_{m=1}^{N+1} (-1)^{m+1} \sum_{1 \le n_1 < \dots < n_m \le N+1} P(A_{n_1} \cap \dots \cap A_{n_m}).$$

(iii) Die vereinfachte Formel folgt aus der Überlegung, dass die Anzahl der Summanden in die Anzahl $\binom{N}{m}$ der möglichen Auswahl von m Elementen aus einer N-elementigen Menge entspricht.

3.1.11 Beispiel (Julklapp)

In einer Schulklasse mit N Kindern soll zur geplanten Weihnachtsfeier jedes Kind jeweils ein anderes beschenken. Um die Zuteilung zufällig zu machen, werden N mit den verschiedenen Namen versehene gleichartige Zettel in einen Hut geworfen und jedes Kind zieht zufällig einen Zettel ohne Zurücklegen. Wie groß ist die Wahrscheinlichkeit, dass das Verfahren funktioniert, d. h. dass kein Kind seinen eigenen Zettel zieht?

3.1.12 Definition

$$\begin{split} \limsup_{n \to \infty} A_n &= \{\omega; \ \omega \in A_n \ \text{für unendlich viele } n \} \\ &= \bigcap_{n \ge 1} \bigcup_{k \ge n} A_k \\ \liminf_{n \to \infty} A_n &= \{\omega; \ \omega \in A_n \ \text{für alle bis auf endlich viele } n \} \\ &= \bigcup_{n \ge 1} \bigcap_{k \ge n} A_k \end{split}$$

 $\lim_{n \to \infty} A_n \text{ existient, falls } \limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n,$

und dann gilt

$$\lim_{n \to \infty} A_n := \limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n.$$

3.1.13 Lemma

(i) Gilt
$$A_1 \subset A_2 \subset \dots$$
 so existiert $\lim_{n \to \infty} A_n$ und es gilt $\lim_{n \to \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.
(ii) Gilt $A_1 \supset A_2 \supset \dots$ so existiert $\lim_{n \to \infty} A_n$ und es gilt $\lim_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

3.1.14 Satz (σ -Stetigkeit)

Ein Wahrscheinlichkeitsmaß ist σ -stetig, d.h. $P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n)$, falls $\lim_{n\to\infty} A_n$ existiert. Insbesondere gilt:

(i)
$$A_1 \subset A_2 \subset \ldots \Longrightarrow \lim_{n \to \infty} P(A_n) = P(\lim_{n \to \infty} A_n) = P(\bigcup_{n=1}^{\infty} A_n).$$

(ii)
$$A_1 \supset A_2 \supset \ldots \Longrightarrow \lim_{n \to \infty} P(A_n) = P(\lim_{n \to \infty} A_n) = P(\bigcap_{n=1}^{\infty} A_n).$$

Beweis

Wir zeigen zuerst (i): Aus $A_1 \subset A_2 \subset \dots$ folgt für alle $n \in \mathbb{N}$

$$\bigcup_{k\geq n}A_k=\bigcup_{k\geq 1}A_k\quad \text{ und }\quad \bigcap_{k\geq n}A_k=A_n.$$

Damit gilt

$$\limsup_{n \to \infty} A_n = \bigcap_{n \ge 1} \bigcup_{k \ge n} A_k = \bigcap_{n \ge 1} \bigcup_{k \ge 1} A_k = \bigcup_{k \ge 1} A_k = \bigcup_{n \ge 1} A_n = \bigcup_{n \ge 1} \bigcap_{k \ge n} A_k = \liminf_{n \to \infty} A_n$$

und somit $\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$. Setze nun $B_k = A_k \setminus A_{k-1}$ mit $A_0 = \emptyset$. Dann gilt $A_n = \bigcup_{k=1}^n B_k$ und $\lim_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} B_k$. Aus der σ -Additivität von P folgt

$$P(\lim_{n \to \infty} A_n) = P\left(\biguplus_{k=1}^{\infty} B_k\right) = \sum_{k=1}^{\infty} P(B_k)$$
$$= \lim_{n \to \infty} \sum_{k=1}^{n} P(B_k) = \lim_{n \to \infty} P\left(\biguplus_{k=1}^{n} B_k\right) = \lim_{n \to \infty} P(A_n).$$

(ii) Die Aussage über die Existenz des Limes folgt analog zu (i). Die Vertauschung von Limes und Wahrscheinlichkeitsmaß kann aus (i) gefolgert werden, wenn $B_k = A_1 \setminus A_k$ betrachtet wird (Übung!).

Die σ -Stetigkeit folgt aus (i) und (ii), indem

$$I_m := \bigcup_{n=1}^m \bigcap_{k \ge n} A_k = \bigcap_{k \ge m} A_k \subset A_m \subset \bigcup_{k \ge m} A_k = \bigcap_{n=1}^m \bigcup_{k \ge n} A_k =: S_m$$

betrachtet wird (Übung!).

3.1.15 Definition (Messbarer Raum, Wahrscheinlichkeitsraum)

- (i) Ist $\Omega \neq \emptyset$, \mathscr{A} eine σ -Algebra über Ω , so heißt (Ω, \mathscr{A}) ein messbarer Raum.
- (ii) Ist (Ω, \mathcal{A}) ein messbarer Raum und P ein Wahrscheinlichkeitsmaß auf \mathcal{A} , dann heißt (Ω, \mathcal{A}, P) Wahrscheinlichkeitsraum.

3.1.16 Beispiele

- (i) Jeder Laplace'scher Wahrscheinlichkeitsraum ist ein Wahrscheinlichkeitsraum im Sinne von Definition (3.1.15).
- (ii) Geometrische Verteilung:

$$\Omega = \mathbb{N}, \ \mathscr{A} = \mathscr{P}(\Omega) = \sigma(\{\{\omega\}; \ \omega \in \Omega\})$$
und $P : \ \mathscr{A} \ni A \to P(A) = \sum_{\omega \in A} P(\{\omega\}) \in [0, 1]$
mit $P(\{\omega\}) := p(1 - p)^{\omega - 1}$

und $p \in (0,1)$ als Erfolgswahrscheinlichkeit (z. B. $P=\frac{1}{6}$ beim Würfeln bis zur ersten "6"). Bei dieser Definition von P muss nur überprüft werden, ob $P(\Omega)=1$ gilt. Mit der Formel für die geometrische Reihe gilt aber

$$P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\}) = \sum_{\omega \in \mathbb{N}} p(1-p)^{\omega-1}$$

= $p \sum_{n=0}^{\infty} (1-p)^n = p \frac{1}{1-(1-p)} = 1.$

3.1.17 Bemerkung

Wie im Beispiel (3.1.16)(ii) angedeutet, lässt sich ein Wahrscheinlichkeitsmaß durch die Werte auf einem Erzeuger definieren, sofern dieser "anständig" ist. Im Fall einer abzählbaren Ergebnismenge und des Erzeugers der Elementarereignisse lässt sich das explizit nachweisen, sofern $\sum_{w \in \Omega} P(\{\omega\}) = 1 \text{ gilt:}$

$$P\left(\biguplus_{n\geq 1} A_n\right) = \sum_{\omega\in\biguplus_{n\geq 1} A_n} P(\{\omega\})$$

$$= \sum_{n\geq 1} \sum_{\omega \in A_n} P(\{\omega\}) = \sum_{n\geq 1} P(A_n).$$

Im Allgemeinen (etwa für $\Omega = \mathbb{R}$ und $\mathscr{A} = \mathscr{B}$) benötigt man den Maßerweiterungssatz von Carathéodory aus der Maß- und Integraltheorie. Dabei wird auch die "Anständigkeit" eines Erzeugers präzisiert wird. Danach sind sowohl \mathscr{E} (bzw. \mathscr{E}_p) als auch $\mathscr{E}' := \mathscr{E}'_1$ (bzw. \mathscr{E}'_p) aus (3.1.6) "anständig". Das bedeutet, dass ein Wahrscheinlichkeitsmaß auf der Borel- σ -Algebra bereits durch die Werte für $(-\infty, b], b \in \mathbb{R}^p$, festgelegt ist.

3.2 Zufallsvariable

3.2.1 Definition (Messbare Abbildung, Zufallsvariable) Seien (Ω, \mathcal{A}) und (Ω', \mathcal{B}) messbare Räume.

- (i) Eine Abbildung $X: \Omega \ni \omega \to X(\omega) \in \Omega'$ heißt \mathscr{A} - \mathscr{B} -messbar $\Leftrightarrow \bigwedge_{B \in \mathscr{B}} X^{-1}(B) := \{\omega \in \Omega; X(\omega) \in B\} \in \mathscr{A}.$
- (ii) Eine Abbildung $X: \Omega \ni \omega \to X(\omega) \in \Omega'$ heißt **Zufallsvariable (zufälliges Element)** auf (Ω, \mathscr{A}) mit Werten in (Ω', \mathscr{B}) , falls (Ω, \mathscr{A}, P) ein Wahrscheinlichkeitsraum ist und $X \mathscr{A}$ - \mathscr{B} -messbar ist.
- (iii) In diesem Fall ist

$$P^X: \mathcal{B} \ni B \to P^X(B) := P(X \in B) := P(X^{-1}(B)) \in [0, 1]$$

das durch X von \mathscr{A} nach \mathscr{B} transportierte (induzierte) Wahrscheinlichkeitsmaß, welches die Verteilung von X ergibt.

3.2.2 Beispiel (Zweifacher Wurf mit Würfel mit 1,2,3 auf jeweils 2 Seiten)

Hier haben wir

$$\Omega = \{(i,j); i,j \in \{1,2,3\}\} = \{(1,1),(1,2),(1,3),(2,1),(2,2),(2,3),(3,1),(3,2),(3,3)\},$$

$$X: \Omega \ni (i,j) \longrightarrow |i-j| \in \{0,1,2\} = \Omega'$$

und sei

$$\mathscr{B} := \{\emptyset, \Omega, \{0\}, \{1, 2\}\}$$

eine σ -Algebra über Ω' . Für welche der folgenden σ -Algebren ist X \mathscr{A} - \mathscr{B} -messbar?

(a)
$$\mathscr{A} = \{\Omega, \emptyset, A, \overline{A}\} \text{ mit } A = \{(1, 1), (2, 2), (3, 3)\},\$$

(b)
$$\mathscr{A} = \{\Omega, \emptyset, A, \overline{A}\} \text{ mit } A = \{(1, 1), (1, 2), (1, 3)\}.$$

3.2.3 Bemerkung

Wegen $P^X(\Omega') = P(\Omega) = 1$ und

$$P^{X}\left(\biguplus_{n\geq 1}B_{n}\right) = P\left(\biguplus_{n\geq 1}X^{-1}(B_{n})\right) = \sum_{n\geq 1}P^{X}(B_{n})$$

ist P^X tatsächlich ein Wahrscheinlichkeitsmaß auf \mathcal{B} und damit $(\Omega', \mathcal{B}, P^X)$ ein Wahrscheinlichkeitsraum (der Bildraum). In vielen Problemen werden wir nur Aussagen und Modellbildungen für diesen Bildraum betrachten.

3.2.4 Satz (Charakterisierung der Messbarkeit)

Seien (Ω, \mathscr{A}) und (Ω', \mathscr{B}) messbare Räume und \mathscr{E} ein Erzeuger von \mathscr{B} . Dann ist X \mathscr{A} - \mathscr{B} -messbar

 \Leftrightarrow

$$\bigwedge_{E \in \mathscr{E}} X^{-1}(E) \in \mathscr{A}.$$

Beweis

 \Leftarrow : Das System $\mathscr C$ aller Mengen $C \in \mathscr P(\Omega')$ mit $X^{-1}(C) \in \mathscr A$ ist eine σ-Algebra über Ω' . Folglich gilt $\mathscr B \subset \mathscr C$ genau dann, wenn $\mathscr E \subset \mathscr C$ ist. Nach Voraussetzung gilt $\mathscr E \subset \mathscr C$, was $\mathscr B \subset \mathscr C$ impliziert. $\mathscr B \subset \mathscr C$ ist aber gleichbedeutend mit der Messbarkeit von X.

3.2.5 Definition (Zufallsgröße, Zufallsvektor)

Sei \mathscr{B}_1 bzw. \mathscr{B}_p die Borel- σ -Algebra über \mathbb{R} bzw. \mathbb{R}^p . Ist X eine Zufallsvariable auf (Ω, \mathscr{A}, P)

- (i) mit Werten in $(\mathbb{R}, \mathcal{B}_1)$, dann heißt X **Zufallsgröße** (reellwertige Zufallsvariable);
- (ii) mit Werten in $(\mathbb{R}^p, \mathcal{B}_p)$, dann heißt X **Zufallsvektor** (mehrdimensionale Zufallsvariable).

3.2.6 Bemerkung (Existenz einer Zufallsvariablen)

Zu einer vorgegebenen Verteilung P^* auf $(\mathbb{R}^p, \mathcal{B}_p)$ existiert stets eine Zufallsvariable X auf

einem geeigneten Wahrscheinlichkeitsraum mit $P^X = P^*$, z. B. $(\Omega, \mathscr{A}, P) = (\mathbb{R}^p, \mathscr{B}_p, P^*)$ und X Identität.

3.2.7 Satz (Kriterium für reelwertige messbare Abbildungen)

Sei \mathscr{B}_p die Borel- σ -Algebra auf \mathbb{R}^p und \mathscr{A} eine σ -Algebra auf Ω . Dann ist $X:\Omega\ni\omega\longrightarrow X(\omega)\in\mathbb{R}^p$ \mathscr{A} - \mathscr{B}_p -messbar

$$\Leftrightarrow \bigwedge_{b \in \mathbb{R}^p} X^{-1}((-\infty, b]) := \{\omega \in \Omega; \ X(\omega) \le b\} \in \mathscr{A},$$

$$\Leftrightarrow \bigwedge_{b \in \mathbb{R}^p} X^{-1}((-\infty, b)) := \{\omega \in \Omega; \ X(\omega) < b\} \in \mathscr{A},$$

$$\Leftrightarrow \bigwedge_{b \in \mathbb{R}^p} X^{-1}([b, \infty)) := \{\omega \in \Omega; \ X(\omega) \ge b\} \in \mathscr{A},$$

$$\Leftrightarrow \bigwedge_{b \in \mathbb{R}^p} X^{-1}([b, \infty)) := \{\omega \in \Omega; \ X(\omega) \ge b\} \in \mathscr{A}.$$

Beweis

Der Beweis folgt mit Satz 3.2.4 aus der Tatsache, dass $\{(-\infty, b]; b \in \mathbb{R}^p\}, \{(-\infty, b); b \in \mathbb{R}^p\}, \{(b, \infty); b \in \mathbb{R}^p\}, \{(b, \infty); b \in \mathbb{R}^p\}$ Erzeuger von \mathscr{B}_p sind.

3.2.8 Folgerung

$$X: \Omega \ni \omega \longrightarrow X(\omega) := \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_p(\omega) \end{pmatrix} \in \mathbb{R}^p \text{ ist } \mathscr{A}\text{-}\mathscr{B}_p\text{-messbar}$$

 \Leftrightarrow

$$\bigwedge_{\pi=1,\dots,p} X_{\pi}: \Omega \ni \omega \to X_{\pi}(\omega) \in \mathbb{R} \text{ ist } \mathscr{A}\text{-}\mathscr{B}_1\text{-messbar}.$$

(D.h. X ist Zufallsvektor $\Leftrightarrow X$ ist Vektor von Zufallsgrößen).

Beweis

Mit Hilfe des Kriteriums 3.2.7:

$$X_{\pi}^{-1}((-\infty, b_{\pi}]) = \{\omega; X_{\pi}(\omega) \leq b_{\pi}\}$$

$$= \bigcup_{n \geq 1} \{\omega; X_{\pi}(\omega) \leq b_{\pi}, X_{\varrho}(\omega) \leq n, \varrho = 1, \dots, p, \varrho \neq \pi\}$$

$$= \bigcup_{n \geq 1} \{\omega; X(\omega) \leq b^{(n)}\} \in \mathscr{A},$$

wobei $b^{(n)} \in \mathbb{R}^p$ mit $b_{\pi}^{(n)} = b_{\pi}$ und $b_{\varrho}^{(n)} = n$ für $\varrho \neq \pi$.

"⇐"

$$X^{-1}((-\infty, b]) = \{\omega; \ X(\omega) \le b\}$$

$$= \bigcap_{\pi=1}^{p} \{\omega; \ X_{\pi}(\omega) \le b_{\pi}\} \in \mathscr{A} \text{ falls } b = (b_{1}, \dots, b_{p})^{\top}.$$

3.2.9 Satz

Sei \mathscr{B}_1 die Borel- σ -Algebra auf \mathbb{R} und \mathscr{A} eine σ -Algebra auf Ω . Sind $X,Y:\Omega\to\mathbb{R}$ \mathscr{A} - \mathscr{B}_1 -messbar, so gilt

$$\{X < Y\} := \{\omega; \ X(\omega) < Y(\omega)\} \in \mathscr{A}, \qquad \{X \le Y\} := \{\omega; \ X(\omega) \le Y(\omega)\} \in \mathscr{A},$$

$$\{X = Y\} := \{\omega; \ X(\omega) = Y(\omega)\} \in \mathscr{A}, \qquad \{X \ne Y\} := \{\omega; \ X(\omega) \ne Y(\omega)\} \in \mathscr{A}.$$

Beweis

Da die Menge \mathbb{Q} der rationalen Zahlen abzählbar ist und dicht in \mathbb{R} liegt, folgt die Behauptung mit Hilfe des Kriteriums 3.2.7 aus

$$\{X < Y\} = \bigcup_{q \in \mathbb{Q}} \{X < q\} \cap \{q < Y\}, \qquad \{X \le Y\} = \overline{\{X > Y\}},$$

$$\{X = Y\} = \{X \le Y\} \cap \{X \ge Y\}, \qquad \{X \ne Y\} = \overline{\{X = Y\}}.$$

#

3.2.10 Satz

Sei \mathscr{B}_1 die Borel- σ -Algebra auf \mathbb{R} , \mathscr{A} eine σ -Algebra auf Ω und $a, c \in \mathbb{R}$. Sind $X, Y, X_n : \Omega \to \mathbb{R}$, $n \in \mathbb{N}$, \mathscr{A} - \mathscr{B}_1 -messbar, so sind auch

$$aX + c$$
, X^2 , $X + Y$, $X - Y$, $X * Y$, $\sup_{n \in \mathbb{N}} X_n$, $\inf_{n \in \mathbb{N}} X_n$, $\limsup_{n \in \mathbb{N}} X_n$, $\liminf_{n \in \mathbb{N}} X_n$ \mathscr{A} - \mathscr{B}_1 -messbar.

Beweis

Mit dem Kriterium 3.2.7 ist mit X zunächst auch $aX + c \mathscr{A}$ - \mathscr{B}_1 -messbar, denn $\{aX + c \leq b\} = \{X \leq \frac{b-c}{a}\} \in \mathscr{A}$ für alle $b \in \mathbb{R}$, falls a > 0, und $\{aX + c \leq b\} = \{X \geq \frac{b-c}{a}\} \in \mathscr{A}$ für alle $b \in \mathbb{R}$, falls a < 0 (der Fall a = 0 ist trivial).

Ebenso ist wegen $\{X^2 \le b\} = \{X \le \sqrt{b}\} \cap \{X \ge -\sqrt{b}\}$ auch X^2 messbar.

Mit Y ist dann auch $b-Y \mathcal{A}-\mathcal{B}_1$ -messbar. Damit gilt

$$\{X + Y \le b\} = \{X \le b - Y\} \in \mathcal{A},$$

weshalb X+Y nach Satz 3.2.9 $\mathscr{A}\text{-}\mathscr{B}_1$ -messbar ist. Die Messbarkeit von X*Y folgt aus

$$X * Y = \frac{1}{4}(X + Y)^2 - \frac{1}{4}(X - Y)^2.$$

Wegen $\{\sup_{n\in\mathbb{N}} X_n \leq b\} = \bigcap_{n=1}^{\infty} \{X_n \leq b\} \in \mathscr{A}$ für alle $b\in\mathbb{R}$, ist auch $\sup_{n\in\mathbb{N}} X_n$ messbar. Mit $\inf_{n\in\mathbb{N}} X_n = -\sup_{n\in\mathbb{N}} (-X_n)$ folgt die Messbarkeit von $\inf_{n\in\mathbb{N}} X_n$. Der Rest folgt aus

$$\lim\sup_{n\in\mathbb{N}}X_n=\inf_{n\in\mathbb{N}}\sup_{m>n}X_m,\ \lim\inf_{n\in\mathbb{N}}X_n=\sup_{n\in\mathbb{N}}\inf_{m\geq n}X_m.$$

Ħ

3.2.11 Satz

Ist $X: \mathbb{R}^p \to \mathbb{R}^q$ stetig, so ist $X \mathscr{B}_p$ - \mathscr{B}_q -messbar.

Beweis

Sei $\mathscr{E}_p = \{O \subset \mathbb{R}^p; O \text{ ist offen}\}$ und $\mathscr{E}_q = \{O \subset \mathbb{R}^q; O \text{ ist offen}\}$. Eine Charakterisierung der Stetigkeit ist, dass das Urbild jeder offenen Menge wieder offen ist, d.h. es gilt $X^{-1}(O) \in \mathscr{E}_p$ für alle $O \in \mathscr{E}_q$. Da \mathscr{E}_p und \mathscr{E}_q nach Bemerkung 3.1.6 Erzeuger der Borel- σ -Algebren \mathscr{B}_p und \mathscr{B}_q sind, folgt die Behauptung aus Satz 3.2.4.

3.2.12 Definition (Verteilungsfunktion)

Ist X Zufallsvektor bzw. Zufallsgröße, dann heißt

$$F_X : \mathbb{R}^p \ni b \to F_X(b) := P(X < b) := P^X((-\infty, b]) \in [0, 1]$$

die Verteilungsfunktion von X.

3.2.13 Bemerkung

 F_X beschreibt die Verteilung P^X von X vollständig (Maßerweiterungssatz der Maß- und Integrationstheorie) und es gilt (vgl. 3.1.6) für alle $a, b \in \mathbb{R}^p$

$$P^{X}((a,b]) = P^{X}((-\infty,b] \setminus \bigcup_{\pi=1}^{p} (-\infty, a(\pi))]$$
$$= F_{X}(b) - P^{X}\left(\bigcup_{\pi=1}^{p} (-\infty, a(\pi))\right)$$

mit $a_{\pi}(\pi) = a_{\pi}$, $a_{\varrho}(\pi) = b_{\varrho}$ für $\varrho \neq \pi$, und die Wahrscheinlichkeit der Vereinigung lässt sich mit Hilfe der Poincaré-Sylvester-Formel und F_X bestimmen. Speziell gilt für

p = 1:

$$P^X((a,b]) = F_X(b) - F_X(a)$$

p = 2:

$$P^{X}((a,b]) = F_{X}(b) - F_{X}(\begin{pmatrix} a_{1} \\ b_{2} \end{pmatrix}) - F_{X}(\begin{pmatrix} b_{1} \\ a_{2} \end{pmatrix}) + F_{X}(a).$$

3.2.14 Satz (Fundamentale Eigenschaften der Verteilungsfunktion einer Zufallsgröße) Ist $F_X : \mathbb{R} \ni b \to F_X(b) \in [0,1]$ eine Verteilungsfunktion, dann gilt:

- (i) F_X ist monoton wachsend, d.h. $b_1 < b_2 \Rightarrow F_X(b_1) \leq F_X(b_2)$.
- (ii) F_X ist rechtsseitig stetig, d. h. $b_n \downarrow b \Rightarrow \lim_{n \to \infty} F_X(b_n) = F_X(b)$.
- (iii) $F_X(-\infty) := \lim_{b \downarrow -\infty} F_X(b) = 0.$
- (iv) $F_X(\infty) := \lim_{b \uparrow \infty} F_X(b) = 1.$

Beweis

(i)
$$b_1 \le b_2 \Rightarrow (-\infty, b_1] \subset (-\infty, b_2]$$

 $\Rightarrow F_X(b_1) = P^X((-\infty, b_1]) \le P^X((-\infty, b_2]) = F_X(b_2).$

(ii)
$$b_n \downarrow b \Rightarrow (-\infty, b] = \bigcap_{n>1} (-\infty, b_n] = \lim_{n\to\infty} (-\infty, b_n].$$

Mit der σ -Stetigkeit von P folgt

$$F_X(b) = P^X((-\infty, b]) = P^X(\lim_{n \to \infty} (-\infty, b_n]) = \lim_{n \to \infty} P^X((-\infty, b_n]) = \lim_{n \to \infty} F_X(b_n).$$

(iii)
$$b_n \downarrow -\infty \Rightarrow \emptyset = \bigcap_{n \geq 1} (-\infty, b_n] = \lim_{n \to \infty} (-\infty, b_n].$$

Mit der σ -Stetigkeit von P folgt wieder

$$\lim_{n\to\infty} F_X(b_n) = \lim_{n\to\infty} P^X((-\infty,b_n]) = P^X\left(\lim_{n\to\infty} (-\infty,b_n]\right) = P^X(\emptyset) = 0.$$

(iv)
$$b_n \uparrow \infty \Rightarrow \mathbb{R} = \bigcup_{n \geq 1} (-\infty, b_n].$$

Mit der σ -Stetigkeit von P folgt wieder

$$\lim_{n \to \infty} F_X(b_n) = \lim_{n \to \infty} P^X((-\infty, b_n]) = P^X\left(\lim_{n \to \infty} (-\infty, b_n]\right) = P^X(\mathbb{R}) = 1.$$

3.2.15 Bemerkung

Da nach Bemerkung 3.1.6 $\mathscr{E}^2 = \{(a,b]; a,b \in \mathbb{R}, a \leq b\}$ ein Erzeuger der Borel- σ -Algebra \mathscr{B}_1 ist und mit $\mu((a,b]) = F(b) - F(a)$ ein Prämaß auf \mathscr{E}^2 definiert wird, ist nach dem Maßerweiterungssatz der Maßtheorie jede Funktion F mit den Eigenschaften (i) bis (iv) in 3.2.14 tatsächlich eine Verteilungsfunktion, d. h. es gibt zu F ein Wahrscheinlichkeitsmaß P_F auf $(\mathbb{R}, \mathscr{B}_1)$ mit $F(b) = P_F((-\infty, b])$ für alle $b \in \mathbb{R}$ bzw. $P_F((a, b]) = F(b) - F(a)$ für alle $a, b \in \mathbb{R}$ mit $a \leq b$. Nach dem Eindeutigkeitssatz der Maßtheorie gibt es zu F sogar nur ein Wahrscheinlichkeitsmaß P_F , das F als Verteilungsfunktion hat.

3.2.16 Beispiel

Ħ

3.2 Zufallsvariable

(i) X konstant, d. h. $X: \Omega \ni \omega \to X(\omega) = a \in \mathbb{R}$.

Dann gilt:
$$X^{-1}(B) = \left\{ \begin{array}{l} \Omega, & \text{falls } a \in B \\ \emptyset, & \text{falls } a \notin B \end{array} \right\} \in \mathscr{A}$$

für alle $B \in \mathcal{B}_1$, weshalb X Zufallsgröße ist.

$$F_X(b) = 1_{[a,\infty)}(b) = \begin{cases} 0 & \text{für } b < a \\ 1 & \text{für } b \ge a \end{cases}$$

(ii) Indikatorfunktionen

Ist $A \in \mathcal{A}$, dann heißt

$$1_A: \Omega \ni \omega \to 1_A(\omega) := \left\{ \begin{array}{l} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } w \notin A \end{array} \right\} \in \mathbb{R}$$

Indikatorfunktion von A.

 1_A ist Zufallsgröße, denn für alle $B \in \mathcal{B}_1$ gilt:

$$1_A^{-1}(B) = \left\{ \begin{array}{l} \Omega, & \text{falls } 0 \in B, \ 1 \in B \\ A, & \text{falls } 0 \notin B, \ 1 \in B \\ \overline{A}, & \text{falls } 0 \in B, \ 1 \notin B \\ \emptyset, & \text{falls } 0 \notin B, \ 1 \notin B \end{array} \right\} \in \mathscr{A}.$$

$$F_{1_A}(b) = P(\overline{A})1_{[0,1)}(b) + 1_{[1,\infty)}(b)$$

$$= \begin{cases} 0 & \text{für } b < 0 \\ P(\overline{A}) & \text{für } 0 \le b < 1 \\ 1 & \text{für } b \ge 1. \end{cases}$$

4 Diskrete und absolut stetige Wahrscheinlichkeitsverteilungen

4.1 Diskrete Wahrscheinlichkeitsverteilungen

- **4.1.1 Definition** (Diskrete Zufallsgröße/-vektor)
- (i) Eine Zufallsgröße (bzw. Zufallsvektor) X ist **diskret**, falls es ein abzählbares (d.h. endliches oder abzählbares unendliches) Ereignis $B^* \in \mathcal{B}_1$ (bzw. $B^* \in \mathcal{B}_p$) gibt mit

$$P^X(B^*) = P(X \in B^*) = 1.$$

(ii) Ist X diskrete Zufallsgröße (bzw. Zufallsvektor), dann heißt

$$p_X : \mathbb{R} \ni x \to p_X(x) := P^X(\{x\}) := P(X = x) := P(\{\omega \in \Omega; X(\omega) = x\}) \in [0, 1]$$

diskrete Dichte oder Zähldichte (der Verteilung) von X.

(iii) $T_X := \{x \in B^*; p_X(x) > 0\} \subseteq B^* \text{ wird Träger (der Verteilung) von } X \text{ genannt.}$

4.1.2 Bemerkung

Für $x \notin B^*$ gilt $p_X(x) = 0$, da $\{x\} \subseteq \overline{B^*}$ und folglich $p_X(x) = P^X(\{x\}) \le P^X(\overline{B^*}) = 0$.

Somit gilt $p_X(x) \neq 0$ nur für abzählbar viele x.

4.1.3 Folgerung

$$\bigwedge_{B \in \mathcal{B}_p} P^X(B) = \sum_{x \in B} p_X(x) \left(:= \sum_{x \in B \cap B^*} p_X(x) \right).$$

Speziell gilt:

$$\bigwedge_{b \in \mathbb{R}^p} F_X(b) = \sum_{x \le b} p_X(x).$$

[Dabei ist die Summation über eine überabzähbare Indexmenge zu verstehen als Summation über die Terme, die von Null verschieden sind, d. h. in obigen Summen treten nur abzählbar viele Summanden ungleich Null auf.]

Beweis

$$\begin{split} P^X(B) &= P^X(B \cap (B^* \cup \overline{B^*})) \\ &= P^X(B \cap B^*) + P^X(B \cap \overline{B^*}) \\ &= P^X\left(\biguplus_{x \in B \cap B^*} \{x\}\right) + 0 \\ &= \sum_{x \in B \cap B^*} P^X(\{x\}). \end{split}$$

4.1.4 Bemerkung

Für diskrete Zufallsgrößen $X: \Omega \to \mathbb{R}$ hat die Verteilungsfunktion F_X Sprungstellen an den Stellen x, für die $p_X(x) > 0$ gilt und verläuft ansonsten konstant.

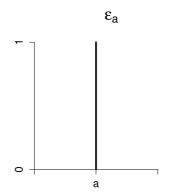
4.1.5 Beispiele (Beispiele für diskrete Wahrscheinlichkeitsverteilungen)

4.1.5.1 Einpunktverteilung ε_a

- X konstant: $X: \Omega \ni \omega \to X(\omega) = a \in \mathbb{R}$
- Träger: $T_X = \{a\}$
- Zähldichte: $p_X(a) = 1$
- Diagramm der Zähldichte: Siehe Abbilung 3 links

4.1.5.2 Diskrete Gleichverteilung $\mathcal{G}(x_1,\ldots,x_n)$

- Verteilungsname in R: sample mit Argumenten x für $x=(x_1,\ldots,x_n)$ und replace = TRUE
- Träger: $T_X = \{x_1, ..., x_n\}$
- Zähldichte: $p_X(x_k) = \frac{1}{n}, \ x_1 < \dots < x_n, \ n \in \mathbb{N}$
- Diagramm der Zähldichte: Siehe Abbilung 3 rechts



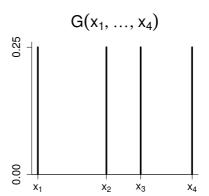
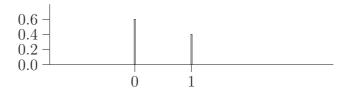


Abbildung 3: Zähldichte der Einpunktverteilung (links) und der diskreten Gleichverteilung (rechts)

4.1.5.3 Zweipunktverteilung / Bernoulli-Verteilung $\mathcal{B}(1,p)$

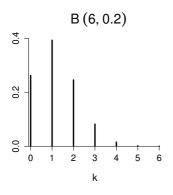
- ullet Verteilungsname in R: binom mit Argumenten size=1 und prob für p
- X Indikatorfunktion: $X = 1_A : \Omega \ni \omega \to 1_A(\omega) = \left\{ \begin{array}{l} 1, \text{ falls } \omega \in A \\ 0, \text{ falls } \omega \notin A \end{array} \right\} \in \mathbb{R}$
- Träger: $T_X = \{0, 1\}$
- Zähldichte: $p_X(1) = P(A) =: p, \ p_X(0) = P(\overline{A}) = 1 p$
- Diagramm der Zähldichte:

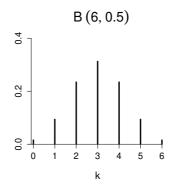




4.1.5.4 Binomial verteilung $\mathcal{B}(n,p)$

- ullet Verteilungsname in R: binom mit Argumenten size für n und prob für p
- Träger: $T_X = \{0, 1, ..., n\}$
- Zähldichte: $p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}; \ n \in \mathbb{N}, \ p \in (0,1)$
- Diagramme von Zähldichten:





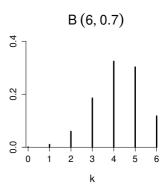


Abbildung 4: Zähldichten verschiedener Binomialverteilungen

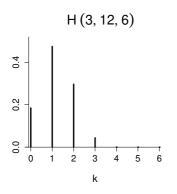
- Bemerkung: Für n = 1 heißt die Binomialverteilung auch Bernoulli-Verteilung.
- Anwendung: Gut-Schlecht-Prüfung (Ziehen mit Zurücklegen bei zwei Ausprägungen)

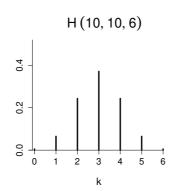
4.1.5.5 Hypergeometrische Verteilung $\mathcal{H}(N; R, S)$

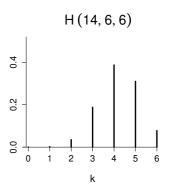
- ullet Verteilungsname in R: hyper mit Argumenten k für N, m für R und n für S
- Träger: $T_X = \{ \max\{0, N S\}, \dots, \min\{R, N\} \}$
- Zähldichte: $p_X(x) = \frac{\binom{R}{x}\binom{S}{N-x}}{\binom{R+S}{N}}, \ N, R, S \in \mathbb{N}, \ N \leq R+S,$ wobei $p_X(x) = 0$, falls x > R oder N-x > S

wo so $P_X(w) = 0$, realis w > 10 odor 10^{-1}

• Diagramme von Zähldichten:



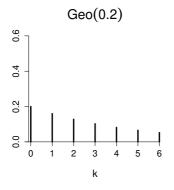


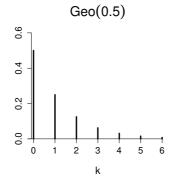


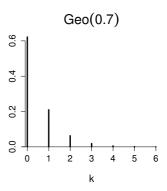
• Anwendung: Gut-Schlecht-Prüfung (Ziehen ohne Zurücklegen bei zwei Ausprägungen)

4.1.5.6 Geometrische Verteilung Geo(p)

- ullet Verteilungsname in R: geom mit Argument prob für p
- Träger: $T_X = \mathbb{N}_0$
- Zähldichte: $p_X(x) = p(1-p)^x, \ p \in (0,1)$
- $\bullet\,$ Diagramme von Zähldichten:



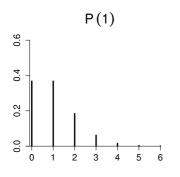


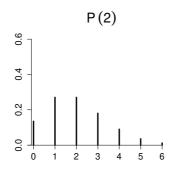


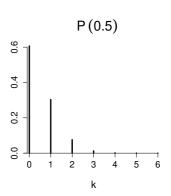
• Anwendung: Wartezeit bei Münzwurfexperimenten

4.1.5.7 Poisson-Verteilung $\mathcal{P}(\lambda)$

- ullet Verteilungsname in R: pois mit Argument lambda für λ
- Träger: $T_X = \mathbb{N}_0$
- Zähldichte: $p_X(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \ \lambda > 0$
- Diagramme von Zähldichten:







• Anwendung: Seltene Ereignisse

4.2 Absolutstetige Wahrscheinlichkeitsverteilungen

4.2.1 Definition (Stetige Zufallsgröße/-vektor)

(i) Eine Zufallsgröße (bzw. ein Zufallsvektor) X heißt absolut stetig

 \Leftrightarrow

$$\bigvee_{f_X:\mathbb{R}\ni x\to f_X(x)\in[0,\infty)} \bigwedge_{a,b\in\mathbb{R}} P^X((a,b]) = \int_a^b f_X(x)dx$$

bzw

DZW.
$$\bigvee_{f_X:\mathbb{R}^p\ni x\to f_X(x)\in[0,\infty)} \bigwedge_{a,b\in\mathbb{R}^p} P^X((a,b]) = \int_a^b f_X(x)dx$$

$$:= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_p}^{b_p} f_X(x_1,\dots,x_p)dx_p\dots dx_2dx_1.$$

(ii) Ist X eine absolut stetige Zufallsgröße (bzw. Zufallsvektor), dann ist $f_X: \mathbb{R}^p \to [0,\infty)$ mit

$$\bigwedge_{a,b\in\mathbb{R}^p} P^X((a,b]) = \int_a^b f_X(x)dx$$

e i n e Wahrscheinlichkeitsdichte (der Verteilung) von X.

(iii) $T_X := \{x; f_X(x) > 0\}$ heißt **Träger** (der Verteilung) von X.

4.2.2 Bemerkung

Im Allgemeinen ist die Wahrscheinlichkeitsdichte nicht eindeutig bestimmt.

Denn: Andert man f_X in einen einzigen Punkt ab, dann ändert sich nichts an den Werten der Integrale.

4.2.3 Folgerung

Ist X absolut stetige Zufallsgröße mit Wahrscheinlichkeitsdichte f_X , dann gilt:

(i)
$$F_X(b) = \int_{-\infty}^b f_X(x) dx$$
.

- (ii) F_X ist stetig.
- (iii) f_X ist stetig in x_0 $\Rightarrow F_X$ differenzierbar in x_0 und $\frac{dF_X}{dx}(x_0) = f_X(x_0)$.
- (iv) $P^X(\lbrace x \rbrace) = 0$ für alle $x \in \mathbb{R}$.

Beweis

(i)

$$F_X(b) = P^X((-\infty, b])$$

$$= P^X \left(\biguplus_{n \ge 2} (-(n+1)|b|, -n|b|] \uplus (-2|b|, b] \right)$$

$$= \sum_{n \ge 2} P^X((-(n+1)|b|, -n|b|]) + P^X((-2|b|, b])$$

$$= \sum_{n \ge 2} \int_{-(n+1)|b|}^{-n|b|} f_X(x) dx + \int_{-2|b|}^{b} f_X(x) dx$$

$$= \int_{-\infty}^{b} f_X(x) dx.$$

- (ii) Nach (i) ist F_X Stammfunktion von f_X und folglich stetig.
- (iii) f_X stetig in x_0

$$\Rightarrow \bigwedge_{\epsilon>0} \bigvee_{\delta>0} \bigwedge_{|x-x_0|<\delta} |f_X(x) - f_X(x_o)| < \epsilon$$

$$\Rightarrow \bigwedge_{\epsilon>0} \bigvee_{\delta>0} \bigwedge_{|y-x_0|<\delta} \left| \int_{x_0}^y (f_X(x) - f_X(x_0)) dx \right| \le \int_{x_0}^y |f_X(x) - f_X(x_0)| dx \le \epsilon |y-x_0|$$

$$\Rightarrow \bigwedge_{\epsilon>0} \bigvee_{\delta>0} \bigwedge_{|y-x_0|<\delta} \left| \frac{\int\limits_{x_0}^y f_X(x)dx}{y-x_0} - f_X(x_0) \right| < \epsilon$$

$$\Rightarrow \bigwedge_{\epsilon>0} \bigvee_{\delta>0} \bigwedge_{|y-x_0|<\delta} \left| \frac{F_X(y) - F_X(x_0)}{y - x_0} - f_X(x_0) \right| < \epsilon$$

$$\Rightarrow \frac{dF_X}{dx}(x_0) = \lim_{y \to x_0} \frac{F_X(y) - F_X(x_0)}{y - x_0} = f_X(x_0)$$

$$\begin{split} P^X(\{x\}) &= P^X \left(\lim_{n \to \infty} \left(x - \frac{1}{n}, x \right] \right) \\ &= \lim_{n \to \infty} P^X \left(\left(x - \frac{1}{n}, x \right] \right) = \lim_{n \to \infty} \left(F(x) - F\left(x - \frac{1}{n} \right) \right) = 0. \end{split}$$

4.2.4 Bemerkung

(i) Für absolut stetige Zuvallsvektoren gilt ein mehrdimensionales Analogon z. B. für (iii)

$$f_X$$
 stetig in $x_0 \Rightarrow \frac{\partial^p F_X}{\partial x_1 \partial x_2 ... \partial x_p}(x_0) = f_X(x_0)$.

Dabei ist F_X die mehrdimensionale Verteilungsfunktion (siehe dazu Abschnitt 4.3).

- (ii) (4.2.3)(iii) ist sehr hilfreich für die Bestimmung einer Wahrscheinlichkeitsdichte, wenn die Verteilungsfunktion gegeben ist.
- (iii) Für Wahrscheinlichkeitsdichten gilt

$$\int_{-\infty}^{\infty} f_X(x)dx = 1 \quad \text{bzw.} \quad \int_{\mathbb{R}^P} f_X(x)dx = 1.$$

4.2.5 Satz (Lokations- und Skalenverschiebungen)

(Affine Abbildungen von Zufallsgrößen)

Ist X eine absolut stetige Zufallsgröße mit Wahrscheinlichkeitsdichte f_X und sind μ , $\sigma \in \mathbb{R}$, $\sigma > 0$, dann ist

$$Y: \Omega \ni \omega \to Y(\omega) := \sigma X(\omega) + \mu \in \mathbb{R}$$

wieder eine absolut stetige Zufallsgröße mit Wahrscheinlichkeitsdichte

$$f_Y : \mathbb{R} \ni y \to f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) \in [0,\infty)$$

und Verteilungsfunktion

$$F_Y: \mathbb{R} \ni b \to F_Y(b) = F_X\left(\frac{b-\mu}{\sigma}\right) \in [0,1].$$

Beweis

$$F_Y(b) = P(Y \le b) = P(\sigma X + \mu \le b) = P\left(X \le \frac{b-\mu}{\sigma}\right)$$
$$= F_X\left(\frac{b-\mu}{\sigma}\right) = \int_{-\infty}^{\frac{b-\mu}{\sigma}} f_X(x) dx$$
$$= \int_{-\infty}^{b} f_X\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{\sigma} dy \qquad \left| \begin{array}{c} \text{Substitution} \\ y = \sigma x + \mu \end{array} \right.$$

Damit ist f_Y mit $f_Y(y) = f_X\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{\sigma}$ Wahrscheinlichkeitsdichte von Y.

4.2.6 Bemerkung

Ist eine Wahrscheinlichkeitsdichte f_X vorgegeben, so können über

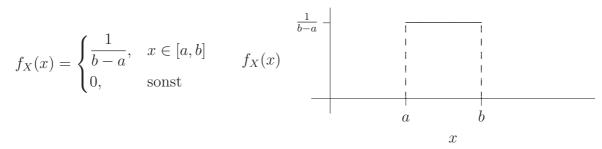
- (i) $\{f_Y; f_Y : \mathbb{R} \ni y \to f_X(y-\mu), \mu \in \mathbb{R}\},\$
- (ii) $\{f_Y; f_Y : \mathbb{R} \ni y \to \frac{1}{\sigma} f_X\left(\frac{y}{\sigma}\right), \ \sigma \in \mathbb{R}, \ \sigma > 0\},$
- (iii) $\{f_Y; f_Y : \mathbb{R} \ni y \to \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right), \ \mu, \ \sigma \in \mathbb{R}, \ \sigma > 0\}.$

ganze Klassen ähnlicher Verteilungen definiert werden. Der Parameter μ beschreibt dabei die Lokation (Lage) und wird deshalb Lokationsparameter genannt. Der Parameter σ beschreibt die Skala ("Streuung") und wird Skalenparameter genannt.

4.2.7 Beispiele (Beispiele für stetige Wahrscheinlichkeitsverteilungen)

4.2.7.1 Rechteckverteilung $\mathcal{R}[a,b]$ (stetige Gleichverteilung $\mathcal{U}(a,b)$) mit Parametern a und b

- Verteilungsname in R: unif für Spezialfall a = 0, b = 1
- Träger: $T_X = [a, b]$ mit $a, b \in \mathbb{R}$, a < b
- Dichte:



• Verteilungsfunktion:

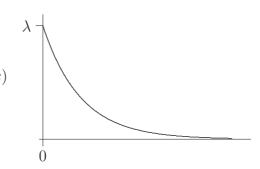
$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x - a}{b - a}, & a \le x < b \\ 1, & b \le x \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x - a}{b - a}, & a \le x < b \\ 1, & b \le x \end{cases}$$

4.2.7.2 Exponential verteilung $\mathcal{E}(\lambda)$ mit Parameter λ

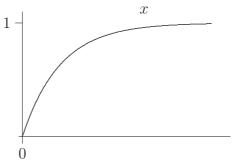
- Verteilungsname in R: exp mit Argument rate für λ
- Träger: $T_X = [0, \infty)$
- Dichte:

Dichte:
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}, \quad \lambda > 0 \qquad f_X(x)$$



• Verteilungsfunktion:

rteilungsfunktion:
$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \ge 0 \end{cases} F_X(x)$$



x

• Anwendung: Lebenszeiten z.B. elektronischer Bauteile

4.2.7.3 Weibull-Verteilung $W(\alpha, \beta)$ mit Parametern α und β

- Verteilungsname in R: weibull mit Argumenten shape für α und scale für β
- Träger: $T_X = [0, \infty)$
- Dichte:

$$f(x) = \begin{cases} 0, & x < 0 \\ \alpha \beta^{-\alpha} x^{\alpha - 1} e^{-(x/\beta)^{\alpha}}, & x \ge 0 \end{cases}, \quad \alpha, \beta > 0$$

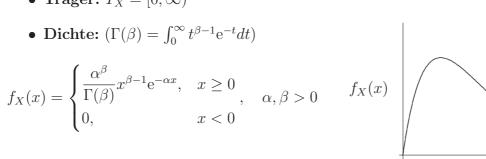
• Verteilungsfunktion:

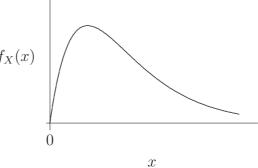
$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - \exp^{-(x/\beta)^{\alpha}}, & x \ge 0 \end{cases}$$

- Anwendung: Lebenszeiten
- Bemerkung: Für $\alpha = 1$, $\beta = 1/\lambda$ ergibt sich die Exponentialverteilung.

4.2.7.4 Gamma-Verteilung $\Gamma(\alpha,\beta)$ mit Parametern α und β

- Verteilungsname in R: gamma mit Argumenten shape für β und rate für α bzw. scale für $\frac{1}{\alpha}$
- Träger: $T_X = [0, \infty)$





- Verteilungsfunktion: Geschlossene Darstellung nur für $\beta \in \mathbb{N}$.
- Bemerkung: Für $\alpha = \lambda$, $\beta = 1$ ergibt sich die Exponentialverteilung.

4.2.7.5 χ^2 -Verteilung χ^2_n mit n Freiheitsgraden

- Verteilungsname in R: chisq mit Argument df für n (df=degree of freedom)
- Träger: $T_X = [0, \infty)$
- Dichte:

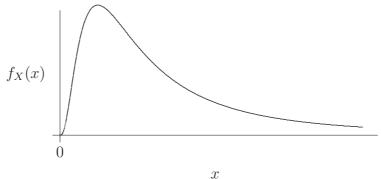
$$f_X(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2 - 1} e^{-x/2}, & x \ge 0\\ 0, & x < 0 \end{cases}, \quad n \in \mathbb{N}$$

- Verteilungsfunktion: Geschlossene Darstellung nur für gerade $n \in \mathbb{N}$.
- Bemerkung: Die χ_n^2 -Verteilung ist eine Gamma-Verteilung mit Parametern $\frac{1}{2}$ und $\frac{n}{2}$.
- Anwendung: Wichtige Verteilung in der Statistik.

4.2.7.6 F-Verteilung $F_{n,m}$ mit n und m Freiheitsgraden

- Verteilungsname in R: f mit Argumenten df1 für n und df2 für m
- Träger: $T_X = [0, \infty)$
- Dichte:

$$f_X(x) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{n/2} \frac{x^{n/2-1}}{(1+n\,x/m)^{(n+m)/2}}, & x \ge 0\\ 0, & x < 0 \end{cases}, \quad n, m \in \mathbb{N}$$

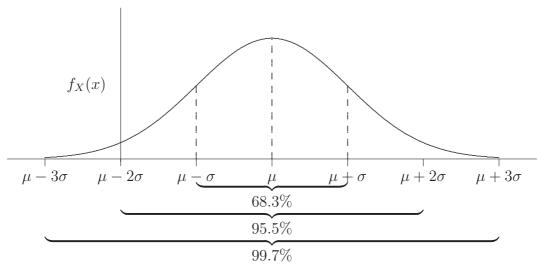


- Verteilungsfunktion: Nicht geschlossen darstellbar.
- Anwendung: Wichtige Verteilung in der Statistik.

4.2.7.7 Normalverteilung $\mathcal{N}(\mu,\sigma^2)$ mit Parametern μ und σ^2

- Verteilungsname in R: norm mit Argumenten mean für μ und sd für σ
- Träger: $T_X = \mathbb{R}$
- Dichte: ('Gaußsche Glockenkurve')

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+.$$



Die Prozentzahlen geben den Wert der Fläche über dem jeweiligen Intervall an. Die Fläche über dem Intervall $[\mu - \sigma, \mu + \sigma]$ beträgt etwa 0.683.

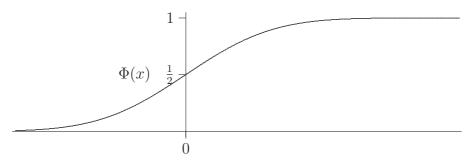
• Verteilungsfunktion: Eine geschlossene Darstellung existiert nicht. Die Verteilungsfunktion der Standardnormalverteilung (i.e. $\mu = 0$, $\sigma^2 = 1$) wird i.a. mit Φ bezeichnet. Die Verteilungsfunktion Φ_{μ,σ^2} einer $\mathcal{N}(\mu,\sigma^2)$ -Verteilung berechnet sich aus Φ gemäß:

$$\Phi_{\mu,\sigma^2}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Da die Dichte der Standardnormalverteilung symmetrisch um 0 ist, gilt für deren Verteilungsfunktion

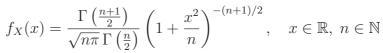
$$\Phi(x) = 1 - \Phi(-x), \quad x \in \mathbb{R}.$$

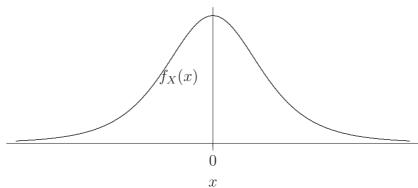
Es reicht daher die Werte von Φ für $x \geq 0$ nur zu kennen.



4.2.7.8 t-Verteilung \mathbf{t}_n mit n Freiheitsgraden

- ullet Verteilungsname in R: t mit Argument df für n
- Träger: $T_X = \mathbb{R}$
- Dichte:





- Verteilungsfunktion: Nur für n = 1 geschlossen darstellbar.
- Anwendung: Wichtige Verteilung in der Statistik.

4.2.7.9 Cauchy-Verteilung $C(\mu, \sigma)$ mit Parametern μ und σ^2

- Verteilungsname in R: cauchy mit Argument location für μ und scale für σ
- Träger: $T_X = \mathbb{R}$
- Dichte:

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

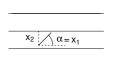
• Verteilungsfunktion:

$$F_X(x) = \frac{1}{\pi} \arctan\left(\frac{x-\mu}{\sigma}\right) + \frac{1}{2}, \quad x \in \mathbb{R}$$

• Bemerkung: Für $\mu = 0$ und $\sigma = 1$ ergibt sich die t-Verteilung t_1 mit einem Freiheitsgrad.

4.2.8 Beispiel (Buffon'sches Nadelproblem)

Auf ein liniiertes Papier unendlicher Größe mit Linienabstand 1 wird zufällig eine Nadel der Länge 1 geworfen. Wie groß ist die Wahrscheinlichkeit für das Ereignis B, dass die Nadel eine Linie schneidet?



Zur Beschreibung des Zufallvorgangs kann man sich auf zwei Komponenten eines Zufallvektors $X = (X_1, X_2)$ beschränken, wobei X_2 den Abstand des unteren Nadelendpunktes zur nächstgelegenen höheren Linie beschreibt, und X_1 ist der Winkel, den die Nadel zu den Linien bildet.

Somit kann X_2 zwischen 0 und 1 variieren und X_1 zwischen 0 und π (wenn wir den Winkel im Bogenmaß messen). Die Zufälligkeit kann nun beschrieben werden durch eine gleichförmige Verteilung von $X = (X_1, X_2)$ auf der Menge

$$M = (0, \pi] \times (0, 1] = \{(x_1, x_2); 0 < x_1 \le \pi, 0 < x_2 \le 1\},\$$

d. h. X ist absolut stetig mit Wahrscheinlichkeitsdichte

$$f_X: \mathbb{R}^2 \ni (x_1, x_2) \to f_X(x_1, x_2) = \frac{1}{\pi} 1_{[0,\pi] \times [0,1]}(x_1, x_2) \in [0, \frac{1}{\pi}].$$

Für das interessierende Ereignis B gilt

$$B = \{(x_1, x_2); \ 0 < x_1 \le \pi, \ 0 < x_2 \le \sin(x_1)\},\$$

und wir erhalten

$$P^{X}(B) = \int_{0}^{\pi} \int_{0}^{1} 1_{B}(x_{1}, x_{2}) \cdot f_{X}(x_{1}, x_{2}) dx_{2} dx_{1} = \frac{1}{\pi} \int_{0}^{\pi} \int_{0}^{\sin(x_{1})} 1 dx_{2} dx_{1}$$
$$= \frac{1}{\pi} \int_{0}^{\pi} \sin(x_{1}) dx_{1} = \frac{1}{\pi} (-\cos(\pi) + \cos(0)) = \frac{2}{\pi} \approx 0,637.$$

4.2.9 Bemerkung

Beim Buffon'schen Nadelproblem haben wir folgende Eigenschaft zur Berechnung der Wahrscheinlichkeit eines beliebigen Ereignisses $B \in \mathcal{B}_p$ ausgenutzt:

$$P^X(B) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} 1_B(x_1, \dots, x_p) f_X(x_1, \dots, x_p) dx_p \dots dx_1.$$

Dazu muss das Integral als Lebesgue-Integral aufgefasst werden. Für das Riemann-Integral kann diese Eigenschaft nur für Intervalle [a, b] und ähnlich einfache Mengen B gezeigt werden.

Ebenso brauchen wir zur Bestimmung von Wahrscheinlichkeiten den Satz von Fubini (siehe Maßtheorie). Dieser besagt insbesondere,

$$\int_{a}^{b} f_{X}(x)dx := \int_{a_{1}}^{b_{1}} \dots \int_{a_{p}}^{b_{p}} f_{X}(x_{1}, \dots, x_{p}) dx_{p} \dots dx_{1}$$

$$= \int_{a_{\pi(1)}}^{b_{\pi(1)}} \dots \int_{a_{\pi(p)}}^{b_{\pi(p)}} f_{X}(x_{1}, \dots, x_{p}) dx_{\pi(p)} \dots dx_{\pi(1)},$$

wobei $\pi: \{1, \ldots, p\} \to \{1, \ldots, p\}$ eine Permutation ist, d. h. der Wert des Integrals lässt sich durch beliebige Reihenfolge der komponentenweisen Integration bestimmen. Dabei ist $[a, b] \in \mathbb{R}^p$, es können aber auch einzelne oder alle Komponenten von a gleich $-\infty$ bzw. von b gleich ∞ sein.

4.2.10 Bemerkung (Zufallszahlen, Dichte, Verteilungsfunktion und Quantile in R)

- rVerteilungsname(n, Argumente) liefert n Zufallszahlen von der Verteilung.
- d*Verteilungsname*(x, *Argumente*) liefert den Wert der Dichtefunktion der Verteilung bei x, d.h. $p_X(x)$ für diskrete Verteilungen und $f_X(x)$ für stetige Verteilungen.
- pVerteilungsname(q, Argumente) liefert den Wert der Verteilungsfunktion der Verteilung bei q, d.h. $F_X(q)$.
- qVerteilungsname(p, Argumente) liefert das p-Quantil der Verteilung, d.h. $F_X^{-1}(p) := \inf\{z \in \mathbb{R}; F_X(z) \geq p\}.$

4.2.11 Bemerkung (Erzeugung von Zufallszahlen mittels runif)

Jede beliebige Verteilung kann aus der Rechteck-Verteilung $\mathbb{R}[0,1]$ auf [0,1] erzeugt werden, die in R durch runif gegeben ist. Für diskrete Verteilungen kann eine Einteilung des Intervalls [0,1] vorgenommen werden, die den einzelnen Wahrscheinlichkeiten entspricht. Sind diese Wahrscheinlichkeiten durch $p_1 = P(\{n_1\}), \ p_2 = P(\{n_2\}), \ p_3 = P(\{n_3\}), \ldots$ gegeben, so wird [0,1] in die Intervalle $[0,p_1], [p_1,p_1+p_2], [p_1+p-2,p_1+p_2+p_3], \ldots$ eingeteilt. Ergibt runif einen Wert in $[\sum_{i=1}^{j-1} p_i, \sum_{i=1}^{j} p_i]$, so wird als Zufallszahl n_j benutzt. Das gleiche Ergebnis erhält man, wenn man die Verteilungsfunktion F der Verteilung benutzt. In diesem Fall wird die Zufallszahl als $F^{-1}(u) := \inf\{z \in \mathbb{R}; F(z) \geq u\}$ gesetzt, wenn runif den Wert u ergibt. Für stetige Verteilungen kann nur die zweite Methode benutzt werden.

4.2.12 Satz (Hauptsatz der Statistik/Stochastik)

Sei F_X die Verteilungsfunktion der Zufallsgröße X und sei deren Inverse $F_X^{-1}:[0,1] \to \mathbb{R}$ durch $F_X^{-1}(u) := \inf\{z \in \mathbb{R}; F_X(z) \geq u\}$ gegeben. Hat U eine Rechteckverteilung auf [0,1], d.h. $U \sim R[0,1]$, und ist $Y = F_X^{-1}(U)$, dann gilt $F_Y = F_X$, d.h. Y hat die gleiche Verteilung wie X.

4.2.13 Lemma

Für alle stetigen und diskreten Zufallsgrößen X gilt immer:

- (i) F_X^{-1} ist monoton wachsend.
- (ii) $F_X^{-1}(F_X(b)) \leq b$ für alle $b \in \mathbb{R}$.
- (iii) $F_X(F_X^{-1}(u)) \ge u$ für alle $u \in [0, 1]$.
- (iv) In (ii) und (iii) gilt "=" genau dann, wenn F_X stetig und invertierbar ist.

Beweis von Satz 4.2.12

Für alle $b \in \mathbb{R}$ beliebig gilt

$$F_Y(b) = P(Y \le b) = P(F_X^{-1}(U) \le b)$$

$$= P(\{\omega \in \Omega; F_X^{-1}(U(\omega)) \le b\})$$

$$\stackrel{(*)}{=} P(\{\omega \in \Omega; U(\omega) \le F_X(b)\}) = F_X(b).$$

Dabei gilt die Gleichheit in (*) wegen Lemma 4.2.13.

Ħ

4.3 Verteilung mehrdimensionaler Zufallsvariablen

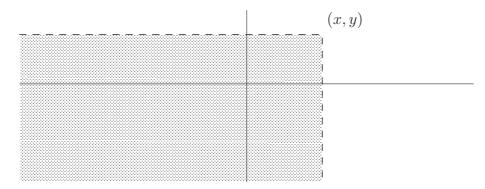
Wir betrachten im folgenden den Fall zweier Zufallsvariablen X und Y. Der Zufallsvektor (X,Y) nehme Werte im \mathbb{R}^2 an, d.h. die Realisationen von (X,Y) sind Paare (x,y) mit $x,y \in \mathbb{R}$. Die Verteilung $P^{(X,Y)}$ wird analog zum eindimensionalen Fall definiert:

$$P^{(X,Y)}(B) := P((X,Y) \in B) = P(\{\omega \in \Omega; \ (X(\omega),Y(\omega)) \in B\}), \quad B \in \mathcal{B}_2, \text{ d.h. } B \subset \mathbb{R}^2.$$

Wie im eindimensionalen Fall kann eine Verteilungsfunktion definiert werden:

$$F_{(X,Y)}(x,y) := P(X \le x, Y \le y) = P(\{\omega \in \Omega; \ X(\omega) \le x, Y(\omega) \le y\}), \quad x, y \in \mathbb{R}.$$

Der Wert der Verteilungsfunktion F an der Stelle (x, y) ist die Wahrscheinlichkeit, dass die Werte des Zufallsvektors (X, Y) in den folgenden schattierten Bereich fallen:



4.3.1 Satz

Für die Verteilungsfunktion $F_{(X,Y)}$ gilt:

1.
$$\lim_{x,y\to-\infty} F_{(X,Y)}(x,y) = 0$$
, $\lim_{x,y\to\infty} F_{(X,Y)}(x,y) = 1$,

2.
$$\lim_{y\to\infty} F_{(X,Y)}(x,y) = F_X(x) \ \forall x\in\mathbb{R}, \ \lim_{x\to\infty} F_{(X,Y)}(x,y) = F_Y(y) \ \forall y\in\mathbb{R},$$

3.
$$F_{(X,Y)}$$
 ist monoton wachsend in jeder Komponente, d.h. $F_{(X,Y)}(x_1, y) \leq F_{(X,Y)}(x_2, y)$ für $x_1 < x_2$ und $F_{(X,Y)}(x, y_1) \leq F_{(X,Y)}(x, y_2)$ für $y_1 < y_2$.

Die Verteilungen F_X und F_Y heißen **Randverteilungen** von $P^{(X,Y)}$ bzw. $F_{(X,Y)}$. Wie im eindimensionalen Fall werden diskrete und stetige Zufallsvariablen unterschieden. Es werden nur solche Fälle betrachtet, in denen X und Y beide diskret bzw. beide stetig sind. Entsprechend zum eindimensionalen Fall können die Verteilungen durch (Zähl-) Dichten beschrieben werden:

• Diskreter Fall: $p_{ij} = P(X = x_i, Y = y_j)$, wobei $x_1 < x_2 < \dots$ und $y_1 < y_2 < \dots$ die Sprungstellen der Verteilungsfunktionen F_X bzw. F_Y sind, so dass

$$F_{(X,Y)}(x,y) = P(X \le x, Y \le y) = \sum_{i,j: x_i \le x, y_j \le y} p_{ij}.$$

Die Zahlen p_{ij} heißen Zähldichte von $P^{(X,Y)}$ bzw. von (X,Y). Die (Rand-) Zähldichten von X bzw. Y können berechnet werden durch:

$$p_i^X = \sum_j p_{ij}, \qquad p_j^Y = \sum_i p_{ij}.$$

• Stetiger Fall:

$$F_{(X,Y)}(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(s,t) dt ds,$$

wobei $f_{(X,Y)}$ eine Riemann-integrierbare Funktion ist mit

$$f_{(X,Y)}(s,t) \ge 0$$
, für alle $s,t \in \mathbb{R}$ und $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{(X,Y)}(s,t) dt ds = 1$.

Für die (Rand-) Dichten von X bzw. Y gelten:

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,t) dt, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(s,y) ds.$$

4.3.2 Beispiel

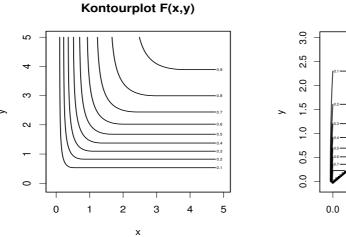
Der Zufallsvektor (X, Y) habe eine diskrete Verteilung mit

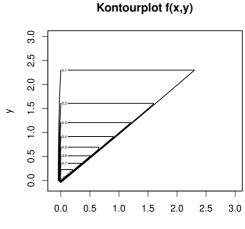
$$p_{ij} = \binom{n}{j} \frac{1}{2^{n+i+1}}, \quad i \in \mathbb{N}_0, j \in \{0, \dots, n\}.$$

4.3.3 Beispiel

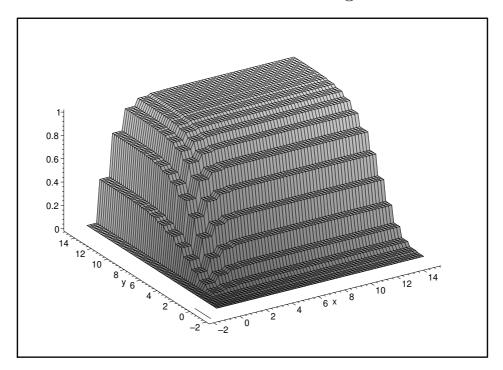
Die Zufallsvariable (X, Y) habe die Verteilungsfunktion:

$$F_{(X,Y)}(x,y) = \begin{cases} 1 - e^{-x} - xe^{-y}, & 0 \le x \le y, \\ 1 - e^{-y} - ye^{-y}, & 0 \le y < x, \\ 0, & \text{sonst.} \end{cases}$$

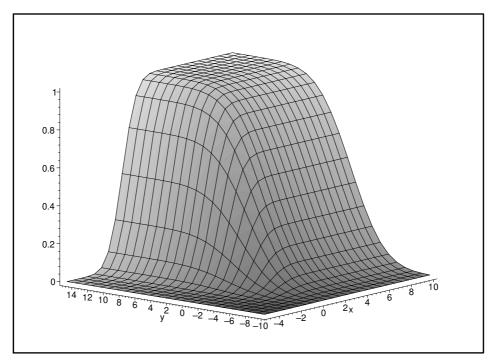




Zweidimensionale diskrete Verteilungsfunktion



Zweidimensionale stetige Verteilungsfunktion



4.3.4 Beispiel (Mehrdimensionale diskrete Verteilungen)

4.3.4.1 Verallgemeinerung hypergeometrische-Verteilung $\mathcal{H}(n; N_1, \dots, N_k)$

• Träger:
$$T_X \subset B^* = \{(x_1, \dots, x_k); x_1, \dots, x_k \in \{0, \dots, n\}, n = x_1 + \dots + x_k\}$$

• Zähldichte:
$$p_X(x) = p_X(x_1, \dots, x_k) = \frac{\binom{N_1}{x_1} \cdot \binom{N_2}{x_2} \cdot \dots \cdot \binom{N_k}{x_k}}{\binom{N}{n}}$$
 mit $N = N_1 + \dots + N_k, n = x_1 + \dots + x_k$

- 1 10/
- Zufallsvektor
- Anwendung: Ziehen ohne Zurücklegen bei mehr als zwei Ausprägungen

4.3.4.2 Multinomial-Verteilung $\mathcal{M}(n; p_1, \dots, p_k)$

- Träger: $T_X = \{(x_1, \dots, x_k); x_1, \dots, x_k \in \{0, \dots, n\}, n = x_1 + \dots + x_k\}$
- Zähldichte: $p_X(x) = p_X(x_1, \dots, x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} \cdot p_2^{x_2} \cdot \dots \cdot p_k^{x_k}$ mit $n = x_1 + \dots + x_k$ und $p_1 + \dots + p_k = 1$
- Zufallsvektor
- Anwendung: Ziehen mit Zurücklegen bei mehr als zwei Ausprägungen

4.3.5 Beispiel (Mehrdimensionale stetige Verteilungen)

4.3.5.1 Zweidimensionale Normalverteilung

Eine zweidimensionale Wahrscheinlichkeitsverteilung von besonderer Bedeutung ist die **zweidimensionale Normalverteilung**. Sie besitzt in ihrer allgemeinsten Form fünf Parameter $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$ und $\rho \in (-1, 1)$. Entsprechend wird sie bezeichnet mit $\mathcal{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Ihre Dichtefunktion ist gegeben durch $(x, y \in \mathbb{R})$:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}.$$

Gilt $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ und $\rho = 0$, so heißt die entsprechende Verteilung zweidimensionale Standardnormalverteilung.

Die zweidimensionale Normalverteilung ist im Standardpaket von R nicht enthalten. Sie kann aber schnell selber geschrieben werden. Die Datei d2norm.asc enthält die folgende Funktion:

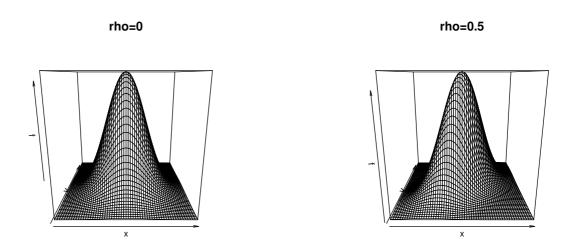


Abbildung 5: Darstellung mit persp

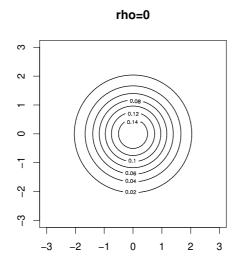
```
"d2norm" <-
function (x,y,m1=0,m2=0,sd1=1,sd2=1,rho=0)
{
# Dichte der zweidimensionalen Normalverteilung
f<-(1/(2*pi*sd1*sd2*sqrt(1-rho^2)))*exp((-1/(2*(1-rho^2)))*((x-m1)^2/sd1^2 -2*rho*(x-m1)*(y-m2)/(sd1*sd2)+(y-m2)^2/sd2^2))
f
}</pre>
```

Diese Funktion hat neben den Argumenten x und y die Parameter μ_1 =m1, μ_2 =m2, σ_1 =s1, σ_2 =s2 und ρ =rho, wobei die Parameter mit Voreinstellungen versehen sind, die die zweidimensionale Standardnormalverteilung ergeben. Die Dichte der zweidimensionalen Normalverteilung kann man sich als einen Berg vorstellen. Zur grafischen Darstellungen eines "Berges" gibt es mehrere Möglichkeiten in R. Mit der R-Funktion persp erhält man eine perspectivische Darstellung, mit contour eine Darstellung mit Höhenlinien und mit image ein Bild, in dem helle Werte hohe Werte anzeigen. Für die Anwendung dieser Funktionen muss ein Intervall von \mathbb{R}^2 gerastert werden. Mit zwei Vektoren \mathbf{x} und \mathbf{y} kann diese Rasterung erzeugt werden. Dann muss eine Matrix mit den Funktionswerten an den Rasterpunkten erstellt werden. Hier wird diese Matrix \mathbf{f} genannt, auch wenn sie in den R-Funktionen \mathbf{z} genannt wird. Die Programm-Datei dnorm_plot.asc enthält die folgende Funktion, die die Dichtefunktion der Normalverteilung über dem Intervall $[-3,3] \times [-3,3]$ darstellt.

```
"d2norm.plot" <-
function (m1=0,m2=0,sd1=1,sd2=1,rho=0.5)
{
```

```
# Plottet die zweidimensionale Normalverteilung
par(pty="s")
x < -seq(-3,3,0.1)
y < -seq(-3,3,0.1)
I<-length(x)</pre>
J<-length(y)
f<-matrix(rep(0,I*J),ncol=J)
for(i in 1:I){
  for(j in 1:J){
     f[i,j]<-d2norm(x[i],y[j],m1=m1,m2=m2,sd1=sd1,sd2=sd2,rho=rho)
  }
persp(x,y,f)
# contour(x,y,f)
# image(x,y,f)
title("rho=0.5")
}
```

par(pty=''s'') erzeugt einen quadratischen Darstellungsbereich für die Grafik. Das ist insbesondere für contour und image wichtig, weil ansonsten das Bild verzerrt ist. Alles, was hinter # steht, wird ignoriert, so dass dahinter Kommentare geschrieben werden können. Hier werden insbesondere die Alternativen zu persp angezeigt. Sollen contour oder image benutzt werden, muss das Kommentarzeichen # entsprechend umgeändert werden. Die Funktion muss auch geändert werden, wenn andere Normalverteilungen über anderen Intervallen dargestellt werden sollen.



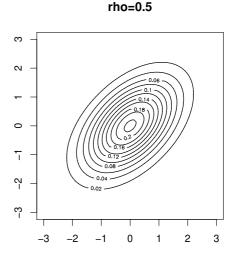
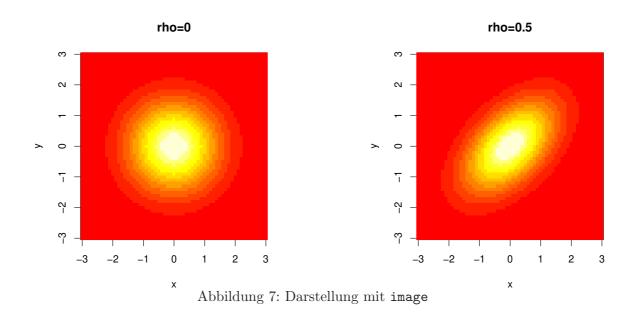


Abbildung 6: Darstellung mit contour



4.3.5.2 Mehrdimensionale Normalverteilung

Die wichtigste mehrdimensionale Verteilung ist die mehrdimensionale Normalverteilung. Diese hat im Fall von s Dimensionen als Parameter einen s-dimensionalen Vektor $\mu \in \mathbb{R}^s$ und eine symmetrische, positiv definite Matrix $\Sigma \in \mathbb{R}^{s \times s}$ ($A \in \mathbb{R}^{s \times s}$ ist positiv definit : $\Leftrightarrow c^{\top}Ac > 0$ für alle $c \in \mathbb{R}^s \setminus \{0\}$). Ist $x \in \mathbb{R}^s$, so ist die Dichte bei $x = (x_1, \dots, x_s)^{\top}$ gegeben durch

$$f(x) = \frac{1}{\sqrt{(2\pi)^s \det(\Sigma)}} \exp\left\{-\frac{1}{2}(x-\mu)^{\top} \Sigma^{-1}(x-\mu)\right\},$$

dabei bezeichnet det die Determinante einer Matrix. Die zweidimensionale Normalverteilung ist ein Spezialfall der mehrdimensionalen Normalverteilung mit

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}.$$

5 Bedingte Wahrscheinlichkeiten und stochastische Unabhängigkeit

5.1 Bedingte Wahrscheinlichkeiten

5.1.1 Beispiel (Bedingte Wahrscheinlichkeiten in Laplace-Experimenten)

$$\frac{\sharp(A\cap B)}{\sharp B} = \frac{\sharp(A\cap B)/\sharp\Omega}{\sharp B/\sharp\Omega} = \frac{P(A\cap B)}{P(B)}$$

5.1.2 Definition (Bedingte Wahrscheinlichkeit)

Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $A, B \in \mathcal{A}, P(B) > 0$, dann heißt

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

die bedingte Wahrscheinlichkeit von A bei gegebenem B.

5.1.3 Beispiele

(i) Zweimaliges Würfeln

Wir sind interessiert an Ereignissen, die die Augensumme betreffen:

 $A_1 = Augensumme gleich 7,$

 $A_2 = Augensumme gleich 8,$

 $A_3 = Augensumme höchstens 5.$

Wir können einen Laplace'schen Wahrscheinlichkeitsraum annehmen mit

$$\Omega = \{(w_1, w_2); w_k \in \{1, 2, \dots, 6\}, k = 1, 2\}.$$

Durch Abzählen erhält man:

$$P(A_1) = \frac{6}{36} = \frac{1}{6} \approx 0.167,$$

$$P(A_2) = \frac{5}{36} \approx 0.139,$$

$$P(A_3) = \frac{1+2+3+4}{36} = \frac{10}{36} = \frac{5}{18} \approx 0.278.$$

Erscheint nun eine 6 beim ersten Wurf, d. h. wir wissen, dass das Ergebnis

$$B = \{(w_1, w_2) \in \Omega; \ w_1 = 6\}$$

eingetreten ist, dann verändern sich die Wahrscheinlichkeiten für uns zu den bedingten Wahrscheinlichkeiten bei gegebenem ${\cal B}$:

5.1 Bedingte Wahrscheinlichkeiten

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{1/36}{6/36} = \frac{1}{6} \approx 0.167,$$

$$P(A_2|B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{1/36}{6/36} = \frac{1}{6} \approx 0.167,$$

$$P(A_3|B) = \frac{P(A_3 \cap B)}{P(B)} = \frac{0}{6/36} = 0.$$

(ii) Geometrische Verteilung

Sei A_n das Ereignis, dass bis zum Zeitpunkt n noch kein Erfolg (keine 6 oder ähnliches) eingetreten ist, d. h.

$$A_n = \{k \in \mathbb{N}; \ k > n\}, \ \text{dann gilt für } N > n$$

$$P(A_{N-n}) = \sum_{k>N-n} p(1-p)^{k-1}$$

$$= p(1-p)^{N-n} \sum_{k=0}^{\infty} (1-p)^k$$

$$= p(1-p)^{N-n} \frac{1}{1-(1-p)} = (1-p)^{N-n}$$

und

$$P(A_N|A_n) = \frac{P(A_N \cap A_n)}{P(A_n)} = \frac{P(A_N)}{P(A_n)}$$

$$= \frac{\sum\limits_{k>N} p(1-p)^{k-1}}{\sum\limits_{k>n} p(1-p)^k}$$

$$= \frac{p(1-p)^N \sum\limits_{k=0}^{\infty} (1-p)^k}{p(1-p)^n \sum\limits_{k=0}^{\infty} (1-p)^k} = (1-p)^{N-n}.$$

Das bedeutet, dass unter Vorkenntnis, dass bis zum Zeitpunkt n noch kein Erfolg eingetreten ist, die Wahrscheinlichkeit, dass in den nächsten N-n Zeitpunkten (kein) Erfolg eintritt, die selbe ist, wie die, dass zu Beginn in den ersten N-n Zeitpunkten (kein) Erfolg eintritt. Man sagt dann auch, dass die geometrische Verteilung eine gedächtnislose Verteilung ist. Aus der bisher verstrichenen Dauer eines durch die geometrische Verteilung beschriebenen Experimentes kann man keine Schlüsse für die Zukunft ziehen.

5.1.4 Satz

Ist (Ω, \mathscr{A}, P) ein Wahrscheinlichkeitsraum, $B \in \mathscr{A}$ mit P(B) > 0, dann ist die bedingte Wahrscheinlichkeit P(A|B) bei gegebenem B aufgefasst als Funktion in A:

$$P(\cdot|B): \mathcal{A} \ni A \to P(A|B) = \frac{P(A \cap B)}{P(B)} \in [0,1]$$

ein Wahrscheinlichkeitsmaß auf \mathscr{A} .

Beweis

$$P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$
 "Normierung".

#

 $(A_n \in \mathscr{A})_{n \geq 1}$ disjunkt $\Rightarrow ((A_n \cap B) \in \mathscr{A})_{n > 1}$ disjunkt, folglich

$$P\left(\biguplus_{n\geq 1} A_n \middle| B\right) = \frac{P\left(\left(\biguplus_{n\geq 1} A_n\right) \cap B\right)}{P(B)} = \frac{P\left(\biguplus_{n\geq 1} (A_n \cap B)\right)}{P(B)} = \frac{\sum\limits_{n\geq 1} P(A_n \cap B)}{P(B)}$$
$$= \sum\limits_{n\geq 1} P(A_n \middle| B).$$

5.1.5 Bemerkung

In Anwendungen und auch in Modellbildung wird meist nicht P(A|B) aus P(B) und $P(A\cap B)$ bestimmt, sondern umgekehrt geht man von der Kenntnis von P(B) und P(A|B) aus um $P(A\cap B)$ zu berechnen.

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

Diese Methode haben wir z. B. implizit bei den Abzählformen für Ziehungen ohne (aber auch mit) Zurücklegen benutzt.

5.1.6 Folgerung

$$A_1, \ldots, A_N \in \mathscr{A}$$
 mit $P(A_1 \cap \ldots \cap A_{N-1}) > 0$. Dann gilt

$$P(A_1 \cap ... \cap A_N) = P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot ... \cdot P(A_N | A_1 \cap ... \cap A_{N-1})$$

$$= P(A_1) \cdot \prod_{n=2}^{N} P\left(A_n | \bigcap_{k=1}^{n-1} A_k\right)$$

Beweis

Durch vollständige Induktion. Für N=2 ist die Behauptung richtig. Sie gelte für $N\geq 2$. Dann folgt

$$P(A_{1} \cap ... \cap A_{N} \cap A_{N+1}) = P(A_{N+1} | A_{1} \cap ... \cap A_{N}) \cdot P(A_{1} \cap ... \cap A_{N})$$

$$\stackrel{\text{(Ind. Vor.)}}{=} P(A_{1}) \prod_{n=2}^{N} P\left(A_{n} | \bigcap_{k=1}^{n-1} A_{k}\right) \cdot P(A_{N+1} | A_{1} \cap ... \cap A_{N})$$

$$= P(A_{1}) \prod_{n=2}^{N+1} P(A_{n} | \bigcap_{k=1}^{n-1} A_{k}).$$

$$\sharp$$

5.1.7 Beispiel

Beim Skat soll die Wahrscheinlichkeit dafür bestimmt werden, dass die drei Spieler je genau ein As erhalten. Aus Symmetriegründen können wir annehmen, dass Spieler 1 die ersten 10 ausgeteilten Karten erhält, Spieler 2 die nächsten 10, dann Spieler 3 zehn und die letzten 2 an den Skat kommen.

5.1.8 Satz (Formel von der totalen Wahrscheinlichkeit)

Ist (Ω, \mathscr{A}, P) ein Wahrscheinlichkeitsraum, $(B_n \in \mathscr{A})_{n \in \mathscr{N}}$, $\mathscr{N} = \{1, \dots, N\}$ oder $\mathscr{N} = \mathbb{N}$, eine disjunkte **Zerlegung von** Ω , d. h. $\Omega = \biguplus_{n \in \mathscr{N}} B_n$, mit $P(B_n) > 0$, $n \in \mathscr{N}$, dann gilt für jedes Ereignis $A \in \mathscr{A}$:

$$P(A) = \sum_{n \in \mathcal{N}} P(A|B_n) \cdot P(B_n)$$

Beweis

$$P(A) = P\left(A \cap \left(\biguplus_{n \in \mathcal{N}} B_n\right)\right) = P\left(\biguplus_{n \in \mathcal{N}} (A \cap B_n)\right)$$
$$= \sum_{n \in \mathcal{N}} P(A \cap B_n) = \sum_{n \in \mathcal{N}} P(A|B_n) \cdot P(B_n).$$

Н

5.1.9 Beispiel

Zwei Firmen V und W stellen ein Maschinenteil her. Sei B das Ereignis, dass das Maschinenteil von Firma V hergestellt wurde, und B^c das Ereignis, dass das Maschinenteil von Firma W hergestellt wurde. A sei das Ereignis, dass das Maschinenteil kaputt ist. Folgende Wahrscheinlichkeiten sind bekannt:

Wahrscheinlichkeit, dass das

Maschinenteil von Firma V hergestellt wurde: P(B) = 0.04

Wahrscheinlichkeit, dass das Maschinenteil kaputt ist,

wenn es von Firma V hergestellt wurde: P(A|B) = 0.7

Wahrscheinlichkeit, dass das Maschinenteil kaputt ist,

wenn es von Firma W hergestellt wurde: $P(A|B^c) = 0.01$

Zunächst folgt aus den Angaben $P(B^c) = 1 - P(B) = 0.96$, so dass

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = 0.7 \cdot 0.04 + 0.01 \cdot 0.96 = 0.0376.$$

Daher beträgt die Wahrscheinlichkeit, dass das Maschinenteil kaputt ist, 3.76%.

5.1.10 Satz (Formel von Bayes)

Ist (Ω, \mathscr{A}, P) ein Wahrscheinlichkeitsraum, $(B_n \in \mathscr{A})_{n \in \mathscr{N}}$ eine disjunkte Zerlegung von Ω mit $P(B_n) > 0$, $n \in \mathscr{N}$, und $A \in \mathscr{A}$ ein Ereignis mit P(A) > 0, dann gilt

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{\sum_{n \in \mathcal{N}} P(A|B_n) \cdot P(B_n)}.$$

Beweis

Mit der Formel von der totalen Wahrscheinlichkeit erhält man

$$P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} = \frac{P(A|B_n) \cdot P(B_n)}{\sum_{n \in \mathcal{N}} P(A|B_n) \cdot P(B_n)}.$$

#

5.1.11 Bemerkung

- Bei der Bayesschen Formel wird von der **Wirkung** A auf die **Ursache** B_i zurückgeschlossen. A kann als Symptom interpretiert werden, das von verschiedenen Krankheiten B_1, \ldots, B_n verursacht wird. Gesucht ist die Wahrscheinlichkeit, dass der Patient bei Beobachtung des Symptoms A an Krankheit B_i leidet. In dieser Situation ist man oft daran interessiert die Krankheit B_i zu finden, die die größte Wahrscheinlichkeit $P(B_i|A)$ besitzt.
- $P(B_1), \ldots, P(B_n)$ heißen **a-priori Wahrscheinlichkeiten**; $P(B_1|A), \ldots, P(B_n|A)$ heißen **a-posteriori Wahrscheinlichkeiten**. Die a-priori Wahrscheinlichkeiten sind als subjektive Vorbewertung zu interpretieren, die durch zusätzliche Information (Eintreten des Ereignisses A) in die a-posteriori Wahrscheinlichkeiten überführt werden. Dieser Zusammenhang wird durch die Bayessche Formel hergestellt.

5.1.12 Beispiel (Fortsetzung von Beispiel 1.1.4)

Die Formel von Bayes kann zur Analyse von medizinischen Labortests verwendet werden. Dabei sind prinzipiell zwei Situationen von Falschergebnissen zu unterscheiden: der Test erkennt eine Krankheit nicht, obwohl die Person erkrankt ist (falsch negativ), bzw. der Test ergibt eine Erkrankung, obwohl die Person tatsächlich gesund ist (falsch positiv). Als Beispiel sei der ELISA-Test zur Erkennung einer HIV-Infektion gewählt (siehe Beispiel 1.1.4).

Der ELISA-Test liefert mit Wahrscheinlichkeit 0.997 ein positives Testergebnis, wenn eine Person tatsächlich erkrankt ist. Die Wahrscheinlichkeit, dass eine gesunde Person ein positives Testergebnis erhält, beträgt 0.001. Die Wahrscheinlichkeit, dass eine Person krank ist, sei $q \in [0, 1]$.

- (i) Wie groß ist die Wahrscheinlichkeit, dass der Test bei einer zufällig ausgewählten Person positiv reagiert?
- (ii) Wie groß ist die Wahrscheinlichkeit, dass die Person tatsächlich erkrankt ist, wenn eine positive Reaktion im Test beobachtet wurde?

5.1.13 Beispiele (Lebenszeit-Verteilungen)

(i) Diskret: Geometrische Verteilung

Im Beispiel (5.1.3)(ii) wurde gezeigt, dass die geometrische Verteilung, die die (diskrete) Wartezeit bis zum ersten Erfolg beschreibt, gedächtnislos ist, d. h. für

$$A_n = \{k \in \mathbb{N}; \ k > n\}, \ n \in \mathbb{N}, \ \text{gilt } P(A_N | A_n) = P(A_{N-n})$$

für N > n. Es liegt nun nahe zu fragen, ob es noch mehr gedächtnislose Verteilungen gibt.

Ist \circledast erfüllt, dann gilt nach Definition der bedingten Wahrscheinlichkeit für N>n

$$P(A_N) = P(A_N \cap A_n) = P(A_n) \cdot P(A_N | A_n)$$
$$= P(A_n) \cdot P(A_{N-n}).$$

Insbesondere gilt also

$$P(A_N) = P(A_1) \cdot P(A_{N-1})$$

und mit Induktion

$$P(A_N) = P(A_1)^N.$$

Folglich für N > 1

$$P(\{N\}) = P(A_{N-1} \setminus A_N)$$

$$= P(A_{N-1}) - P(A_N)$$

$$= P(A_1)^{N-1} - P(A_1)^N$$

$$= (1 - P(A_1))P(A_1)^{N-1}$$

Setzt man noch $p := P(\{1\})$, also $P(A_1) = P(\overline{\{1\}}) = 1 - p$, dann muss gelten:

$$P({N}) = p(1-p)^{N-1}$$

für N > 1 (und sogar N = 1).

Somit ist die geometrische Verteilung die einzige gedächtnislose Verteilung für diskrete Wartezeiten.

(ii) Absolut stetig: Exponential-Verteilung:

Zur Beschreibung von Lebenszeiten ist es naheliegend, sich nicht auf diskrete Zeitpunkte zu beschränken, sondern alle nicht negativen reellen Zahlen als Ergebnisse zuzulassen. Dann bedeutet Gedächtnislosigkeit

$$P(A_T|A_t) = P(A_{T-t})$$

für T>t>0, wobei $A_t:=(t,\infty)=\{s\in\mathbb{R};\ s>t\},\ t>0$, d. h. A_t ist das Ereignis, dass die Lebensdauer größer als t ist.

5.2 Unabhängigkeit von Ereignissen

Ein Ereignis A sollte von einem Ereignis B unabhängig sein, wenn die Vorkenntnis, dass das Ereignis B eingetreten ist, keinen Einfluss auf die Wahrscheinlichkeit für das Auftreten von A besitzt.

In Ausdrücken der bedingten Wahrscheinlichkeit heißt das für B mit P(B) > 0;

$$P(A|B) = P(A).$$

Dies ist äquivalent zu (vgl. (5.1.5))

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

Ist auch P(A) > 0, so ist dies wiederum äquivalent zu

$$P(B|A) = P(B),$$

d. h. B unabhängig von A.

Unabhängigkeit ist somit eine symmetrische Eigenschaft und sollte daher durch eine symmetrische Bedingung charakterisiert werden. Außerdem wäre es günstig, auf die Einschränkungen P(A) > 0, P(B) > 0 verzichten zu können.

Schließlich sollte man die Unabhängigkeit auch verbal von linearer Unabhängigkeit (lineare Algebra) unterscheiden können: Stochastische Unabhängigkeit.

5.2.1 Definition (Stochastische Unabhängigkeit)

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Zwei Ereignisse $A, B \in \mathcal{A}$ heißen (stochastisch) unabhängig

 \Leftrightarrow

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

5.2.2 Lemma (Einfache Eigenschaften)

A,B unabhängig \Leftrightarrow A,\overline{B} unabhängig \Leftrightarrow \overline{A},B unabhängig \Leftrightarrow $\overline{A},\overline{B}$ unabhängig

#

Beweis

A, B unabhängig

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

$$\Rightarrow P(A \cap \overline{B}) = P(A \setminus (A \cap B))$$

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

$$= P(A) (1 - P(B)) = P(A) \cdot P(\overline{B})$$

 $\Rightarrow A, \overline{B}$ unabhängig.

Die restlichen Behauptungen folgen durch geeignete Identifikation.

5.2.3 Definition

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum. Ereignisse $A_1, \ldots, A_N \in \mathcal{A}$ heißen (stochastisch) unabhängig

 \Leftrightarrow

$$\bigwedge_{\substack{2 \le n \le N \\ i_1 < i_2 < \dots < i_n}} P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_n})$$

5.2.4 Beispiel

2-maliges Würfeln: $A_1 = 6$ im ersten Wurf, $A_2 = 6$ im zweiten Wurf, $A_3 = 6$ Augensumme ist 7. Dann sind

$$A_1, A_2$$
 unabhängig,
 A_1, A_3 unabhängig,
 A_2, A_3 unabhängig,

aber A_1, A_2, A_3 sind nicht unabhängig, denn

$$P(A_1 \cap A_2 \cap A_3) = 0 \neq \frac{1}{6^3} = P(A_1) \cdot P(A_2) \cdot P(A_3).$$

5.2.5 Übung

In Analogie zu der Aussage in 5.2.2 zeige man:

 A_1, \ldots, A_N unabhängig

 \Leftrightarrow

$$\bigwedge_{\substack{A_n^* \in \{\emptyset, A_n, \overline{A}_n, \Omega\} \\ n = 1, \dots, N}} P(A_1^* \cap \dots \cap A_N^*) = P(A_1^*) \cdot \dots \cdot P(A_N^*).$$

5.3 Unabhängigkeit von Zufallsvektoren

5.3.1 Vorbemerkung

Häufig treten verschiedene zufällige Erscheinungen auf, die jede für sich durch eine Zufallsgröße beschrieben werden kann. Da diese zufälligen Erscheinungen eventuell miteinander in Verbindung gebracht werden sollen, sollten sie auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definiert sein. Zum Beispiel kann der Produktraum $(\prod_{i=1}^p \Omega_i, \bigotimes_{i=1}^p \mathcal{A}_i)$ als gemeinsamer Wahrscheinlichkeitsraum gewählt werden (siehe Maßtheorie). Die Zufallsgrößen lassen sich zu einem Zufallsvektor zusammenfassen, und die Verteilung des Zufallsvektors bezeichnet man oft als **gemeinsame Verteilung** der Zufallsgrößen.

Zufallsgrößen (aber auch Zufallsvektoren) wird man als (stochastisch) unabhängig ansehen, wenn alle durch sie beschriebenen Ereignisse unabhängig sind, genauer:

5.3.2 Definition

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum.

(i) Zufallsgrößen X_1, \ldots, X_N heißen (stochastisch) unabhängig.

 \Leftrightarrow

$$\bigwedge_{B_n \in \mathcal{B}_1, \ n=1,...,N} P(X_1 \in B_1,...,X_N \in B_N) = P(X_1 \in B_1) \cdot ... \cdot P(X_N \in B_N).$$

(ii) Zufallsvektoren X_1, \ldots, X_N mit Dimensionen p_1, \ldots, p_N heißen (stochastisch) unabhängig

 \Leftrightarrow

$$\bigwedge_{B_n \in \mathscr{B}_{p_n}, n=1,\dots,N} P(X_1 \in B_1,\dots,X_N \in B_N) = P(X_1 \in B_1) \cdot \dots \cdot P(X_N \in B_N).$$

5.3.3 Bemerkung

Da wir zumeist an der Unabhängigkeit von Zufallsgrößen interessiert sind, ist diese Definition explizit angegeben worden, obwohl sie ein Spezialfall von (ii) ist. In Verteilungsschreibweise ist sie noch einprägsamer:

$$\bigwedge_{B_n \in \mathscr{B}_{p_n}, n=1,\dots,N} P^{(X_1,\dots,X_N)}(B_1 \times \dots \times B_N) = P^{X_1}(B_1) \cdot \dots \cdot P^{X_N}(B_N),$$

d. h. die gemeinsame Verteilung P^{X_1,\ldots,X_N} ist das "Produkt" der einzelnen Verteilungen P^{X_1},\ldots,P^{X_N} , nämlich das Produktmaß $\bigotimes_{n=1}^N P^{X_n}=P^{X_1}\otimes\ldots\otimes P^{X_N}$ (siehe Maßtheorie).

5.3.4 Lemma

Sind X_1, \ldots, X_N Indikatorfunktionen, d. h. $X_n = 1_{A_n}, A_n \in \mathcal{A}, n = 1, \ldots, N$, dann gilt: $1_{A_1}, \ldots, 1_{A_N}$ (stochastisch) unabhängig

 \Leftrightarrow

 A_1, \ldots, A_N (stochastisch) unabhängig

#

Beweis

Folgt aus der Übung 5.2.5.

5.3.5 Satz

- (i) X_1, \dots, X_N unabhängig $\Rightarrow (X_1, \dots, X_n), (X_{n+1}, \dots, X_N) \text{ unabhängig}$
- (ii) X_1, \ldots, X_N unabhängig, h_1, \ldots, h_N (messbare) Funktionen auf dem Bildraum von X_1, \ldots, X_N (so dass $Y_n := h_n \circ X_n$, $n = 1, \ldots, N$, Zufallsgrößen bzw. Zuvallsvektoren sind.)
 - $\Rightarrow Y_1, \dots, Y_N \text{ mit } Y_n := h_n \circ X_n, \ n = 1, \dots, N \text{ sind unabhängig.}$

Beweis

(i) Als erstes wird gezeigt, dass X_1, \ldots, X_n (und analog X_{n+1}, \ldots, X_N) unabhängig sind. Das sieht man wie folgt:

$$\bigwedge_{\substack{B_k \in \mathscr{B}_{p_k} \\ k=1,\dots,n}} P^{(X_1,\dots,X_n)}(B_1 \times \dots \times B_n)$$

$$= P^{(X_1,\dots,X_N)}(B_1 \times \dots \times B_n \times \mathbb{R}^{p_{n+1}} \times \dots \times \mathbb{R}^{p_N})$$

$$= P^{X_1}(B_1) \cdot \dots \cdot P^{X_n}(B_n) \cdot P^{X_{n+1}}(\mathbb{R}^{p_{n+1}}) \cdot \dots \cdot P^{X_N}(\mathbb{R}^{p_N})$$

$$= P^{X_1}(B_1) \cdot \dots \cdot P^{X_n}(B_n).$$

Setze nun $Z_1=(X_1,\ldots,X_n),\ Z_2=(X_{n+1},\ldots,X_N)$ und $C_1=B_1\times\ldots\times B_n,\ C_2=B_{n+1}\times\ldots\times B_N$ für beliebige $B_k\in\mathscr{B}_{p_k},\ k=1,\ldots,n,n+1,\ldots,N.$ Dann gilt

$$P^{Z_1}(C_1) \cdot P^{Z_2}(C_2) = P^{(X_1, \dots, X_n)}(B_1 \times \dots \times B_n) \cdot P^{(X_{n+1}, \dots, X_N)}(B_{n+1} \times \dots \times B_N)$$

$$= P^{X_1}(B_1) \cdot \dots \cdot P^{X_n}(B_n) \cdot P^{X_{n+1}}(B_{n+1}) \cdot \dots \cdot P^{X_N}(B_N)$$

$$= P^{(X_1, \dots, X_n, X_{n+1}, \dots, X_N)}(B_1 \times \dots \times B_n \times B_{n+1} \times \dots B_N) = P^{(Z_1, Z_2)}(C_1 \times C_2).$$

Da Mengen der Form $B_1 \times \ldots \times B_n$ die Borel- σ -Algebra auf $\mathbb{R}^{p_1+\ldots+p_n}$ und Mengen der Form $B_{n+1} \times \ldots \times B_N$ die Borel- σ -Algebra auf $\mathbb{R}^{p_{n+1}+\ldots+p_N}$ erzeugen, folgt die stochastische Unabhängigkeit von $Z_1 = (X_1, \ldots, X_n)$ und $Z_2 = (X_{n+1}, \ldots, X_N)$.

Ħ

$$\bigwedge_{C_1,\dots,C_N} P^{(Y_1,\dots,Y_N)}(C_1 \times \dots \times C_N) = P^{(X_1,\dots,X_N)}(h_1^{-1}(C_1) \times \dots \times h_N^{-1}(C_N))$$

$$= P^{X_1}(h_1^{-1}(C_1)) \cdot \dots \cdot P^{X_N}(h_N^{-1}(C_N))$$

$$= P^{Y_1}(C_1) \cdot \dots \cdot P^{Y_N}(C_N).$$

5.3.6 Satz (Charakterisierung der Unabhängigkeit durch die Verteilungsfunktion) $X_1, \ldots X_N$ unabhängige Zufallsvektoren

 \Leftrightarrow

$$\bigwedge_{\substack{x_n \in \mathbb{R}^{p_n} \\ n=1,\dots,N}} F_{(X_1,\dots,X_N)}(x_1,\dots,x_N) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_N}(x_N).$$

Beweis

" \Rightarrow ": Für beliebige $x_1 \in \mathbb{R}^{p_1}, \dots, x_N \in \mathbb{R}^{p_N}$ setze $B_n = (-\infty, x_n], n = 1, \dots, N$. Aus der Unabhängigkeit von X_1, \dots, X_N folgt:

$$F_{(X_1,\dots,X_N)}(x_1,\dots,x_N) = P(X_1 \le x_1,\dots,X_N \le x_N) = P(X_1 \in B_1,\dots,X_N \in B_N)$$
$$= P(X_1 \in B_1) \cdot \dots \cdot P(X_N \in B_N) = F_{X_1}(x_1) \cdot \dots \cdot F_{X_N}(x_N).$$

"\(\sim \)": Nach dem Eindeutigkeitssatz aus der Maßtheorie gibt es auf dem Produktraum $(\prod_{n=1}^{N} \mathbb{R}^{p_n}, \bigotimes_{n=1}^{N} \mathscr{B}_{p_n})$ genau ein Wahrscheinlichkeitsmaß P_0 mit

$$P_0(B_1 \times \ldots \times B_N) = P^{X_1}(B_1) \cdot \ldots \cdot P^{X_N}(B_N)$$
 für alle $B_n \in \mathscr{B}_{p_n}, \ n = 1, \ldots, N.$

Sei $\widetilde{X}_n: \prod_{n=1}^N \mathbb{R}^{p_n} \ni (x_1, \dots, x_N) \to x_n \in \mathbb{R}^{p_n}$ die Projektion auf die *n*'te Komponente von (x_1, \dots, x_N) . Dann gilt

$$P_0^{\widetilde{X}_n}(B_n) = P_0(\widetilde{X}_n^{-1}(B_n)) = P_0(\mathbb{R}^{p_1} \times \ldots \times \mathbb{R}^{p_{n-1}} \times B_n \times \mathbb{R}^{p_{n+1}} \times \ldots \times \mathbb{R}^{p_N}) = P^{X_n}(B_n)$$

für alle $B_n \in \mathscr{B}_{p_n}, \ n = 1, \dots, N$, und $P^{(\widetilde{X}_1, \dots, \widetilde{X}_N)} = P_0$. Also sind $\widetilde{X}_1, \dots, \widetilde{X}_N$ unabhängig. Für diese gilt nach " \Rightarrow " (dem 1. Teil des Beweises):

$$F_{(\widetilde{X}_1,...,\widetilde{X}_N)}(x_1,...,x_N) = F_{\widetilde{X}_1}(x_1) \cdot ... \cdot F_{\widetilde{X}_N}(x_N)$$

= $F_{X_1}(x_1) \cdot ... \cdot F_{X_N}(X_N) = F_{(X_1,...,X_N)}(x_1,...,x_N)$

für alle $x_1 \in \mathbb{R}^{p_1}, \dots, x_N \in \mathbb{R}^{p_N}$. Da die Verteilung $P^{(X_1, \dots, X_N)}$ eindeutig durch die Verteilungsfunktion gegeben ist (Bemerkung 3.2.13), folgt $P^{(X_1, \dots, X_N)} = P_0^{(\tilde{X}_1, \dots, \tilde{X}_N)}$ und somit

Ħ

insbesondere für alle $B_1 \in \mathscr{B}_{p_1}, \ldots, B_N \in \mathscr{B}_{p_N}$

$$P^{(X_1,...,X_N)}(B_1 \times ... \times B_N) = P_0^{(\widetilde{X}_1,...,\widetilde{X}_N)}(B_1 \times ... \times B_N)$$

$$= P_0(\widetilde{X}_1 \in B_1,...,\widetilde{X}_N \in B_N) = P_0(\widetilde{X}_1 \in B_1) \cdot ... \cdot P_0(\widetilde{X}_N \in B_N)$$

$$= P_0^{\widetilde{X}_1}(B_1) \cdot ... \cdot P_0^{\widetilde{X}_N}(B_N) = P^{X_1}(B_1) \cdot ... \cdot P^{X_N}(B_N).$$

5.3.7 Bemerkung

Für N=2 Zufallsgrößen X_1, X_2 sieht man auch ohne dem Eindeutigkeitssatz aus der Maßtheorie, dass für beliebige Intervalle $(a,b]=\left(\left(\begin{pmatrix} a_1\\a_2\end{pmatrix}\right),\left(\begin{pmatrix} b_1\\b_2\end{pmatrix}\right)\right]$ aus $F_{(X_1,X_2)}(x_1,x_2)=F_{X_1}(x_1)\cdot F_{X_2}(x_2)$ für alle $x_1,x_2\in\mathbb{R}$ folgendes folgt:

$$\begin{split} P^{(X_1,X_2)}((a,b]) &= P(X_1 \in (a_1,b_1], \ X_2 \in (a_2,b_2]) \\ &= F_{(X_1,X_2)}(b) - F_{(X_1,X_2)}\left(\binom{a_1}{b_2}\right) - F_{(X_1,X_2)}\left(\binom{b_1}{a_2}\right) + F_{(X_1,X_2)}(a) \\ &= F_{X_1}(b_1) F_{X_2}(b_2) - F_{X_1}(a_1) F_{X_2}(b_2) - F_{X_1}(b_1) F_{X_2}(a_2) + F_{X_1}(a_1) F_{X_2}(a_2) \\ &= (F_{X_1}(b_1) - F_{X_1}(a_1)) \ (F_{X_2}(b_2) - F_{X_2}(a_2)) \\ &= P^{X_1}((a_1,b_1]) \cdot P^{X_2}((a_2,b_2]). \end{split}$$

5.3.8 Satz (Charakterisierung der Unabhängigkeit durch diskrete Dichten)

Sind X_1, \ldots, X_N diskret, dann gilt:

$$X_1, \ldots, X_N$$
 unabhängig

 \Leftrightarrow

$$\bigwedge_{x_1,\dots,x_N} p_{(X_1,\dots,X_N)}(x_1,\dots,x_N) = p_{X_1}(x_1) \cdot \dots \cdot p_{X_N}(x_N).$$

Beweis

Vorüberlegung:

$$X_1, \ldots, X_N$$
 diskret $\Leftrightarrow (X_1, \ldots, X_N)$ diskret,

denn
$$X_1, \ldots, X_N$$
 diskret $\Rightarrow \bigvee_{B_1^*, \ldots, B_{N}^* \text{abz\"{a}hlbar}} P^{X_n}(B_n^*) = 1, \ n = 1, \ldots, N.$

Ist nun $B^* = B_1^* \times \ldots \times B_N^*$, dann ist B^* abzählbar und es gilt

$$P^{(X_1,...,X_N)}(\overline{B^*}) = P(\{X_1 \in \overline{B_1^*}\} \cup ... \cup \{X_N \in \overline{B_N^*}\}) \le \sum_{n=1}^N P^{X_n}(\overline{B_n^*}) = 0,$$

folglich
$$P^{(X_1,\ldots,X_N)}(B^*)=1$$
, also (X_1,\ldots,X_N) ist diskret.

Umgekehrt:
$$(X_1, \ldots, X_N)$$
 diskret \Rightarrow

5.3 Unabhängigkeit von Zufallsvektoren

$$\bigvee_{P_{N-1}, \dots, X_N} P^{(X_1, \dots, X_N)}(B_*) = 1.$$

 \dot{B}_* abzählbar

Sei
$$B_n^* := \{x_n; \bigvee_{x_k, k \in \{1, \dots, N\} \setminus \{n\}} (x_1, \dots, x_n, \dots, x_N) \in B^* \}$$

die Projektion von B^* auf die n-te Komponente.

Dann ist B_n^* abzählbar und

$$P^{X_n}(B_n^*) \ge P^{(X_1,\dots,X_n)}(B^*) = 1,$$

und somit ist X_n diskret für n = 1, ..., N.

Beweis der Charakterisierung

" \Rightarrow ": Klar mit $B_n := \{x_n\}$

" \Leftarrow ": Sei B_n^* abzählbar mit $P^{X_n}(B_n^*)=1$ und B_n beliebig für $n=1,\,\ldots,N$. Dann gilt

$$P^{(X_{1},\dots,X_{N})}(B_{1} \times \dots \times B_{N}) = \sum_{(x_{1},\dots,x_{N})\in(B_{1}\times\dots\times B_{N})\cap(B_{1}^{*}\times\dots\times B_{N}^{*})} p_{(X_{1},\dots,X_{N})}(x_{1},\dots,x_{N})$$

$$= \sum_{(x_{1},\dots,x_{N})\in(B_{1}\cap B_{1}^{*})\times\dots\times(B_{N}\cap B_{N}^{*})} p_{X_{1}}(x_{1})\cdot\dots\cdot p_{X_{N}}(x_{N})$$

$$= \sum_{x_{1}\in B_{1}\cap B_{1}^{*}} \dots \sum_{x_{N}\in B_{N}\cap B_{N}^{*}} p_{X_{1}}(x_{1})\cdot\dots\cdot p_{X_{N}}(x_{N})$$

$$= \left(\sum_{x_{1}\in B_{1}\cap B_{1}^{*}} p_{X_{1}}(x_{1})\right)\cdot\dots\cdot \left(\sum_{x_{N}\in B_{N}\cap B_{N}^{*}} p_{X_{N}}(x_{N})\right)$$

$$= P^{X_{1}}(B_{1})\cdot\dots\cdot P^{X_{N}}(B_{N}).$$

H

5.3.9 Beispiel (Unabhängige, identisch verteilte Indikatorfunktionen (Bernoulli-Experimente))

(Unabhängige Wiederholung von Experimenten mit Ergebnis Erfolg/Mißerfolg, z. b. Würfeln mit Erfolg bei "6" oder zufälliges Auswählen einer Person mit Zurücklegen mit "Erfolg" bei Krankheit).

5.3.10 Satz (Charakterisierung der Unabhängigkeit durch Wahrscheinlichkeitsdichten) X_1, \ldots, X_N seien absolut stetig mit Wahrscheinlichkeitsdichten f_{X_1}, \ldots, f_{X_N} . Dann gilt: X_1, \ldots, X_N unabhängig

 \Leftrightarrow

 (X_1, \ldots, X_N) absolut stetig und $f : \mathbb{R} \ni (x_1, \ldots, x_N) \to f(x_1, \ldots, x_N) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_N}(x_N) \in [0, \infty)$ ist Wahrscheinlichkeitsdichte von (X_1, \ldots, X_N) .

Beweis

Für alle $[a_1, b_1], \ldots, [a_N, b_N]$ gilt

$$P^{X_1}([a_1, b_1]) \cdot \ldots \cdot P^{X_N}([a_N, b_N])$$

$$= \int_{a_1}^{b_1} f_{X_1}(x_1) dx_1 \cdot \ldots \cdot \int_{a_N}^{b_N} f_{X_N}(x_N) dx_N$$

$$= \int_{a_1}^{b_1} \ldots \int_{a_N}^{b_N} f_{X_1}(x_1) \cdot \ldots \cdot f_{X_N}(x_N) dx_N \ldots dx_1$$

Für $[a, b] := [a_1, b_1] \times \ldots \times [a_N, b_N]$ gilt somit:

 X_1, \ldots, X_N unabhängig

 \Leftrightarrow

$$\bigwedge_{a,b} P^{(X_1,\dots,X_N)}([a,b]) = \int_{a_1}^{b_1} \dots \int_{a_N}^{b_N} f_{X_1}(x_1) \cdot \dots \cdot f_{X_N}(x_N) \ dx_N \dots dx_1$$

 \Leftrightarrow

 $f:(x_1,\ldots x_N)\to f_{X_1}(x_1)\cdot\ldots\cdot f_{X_N}(x_N)$ ist Wahrscheinlichkeitsdichte von (X_1,\ldots,X_N) .

ţ

5.3.11 Beispiel (N-dimensionale Normalverteilung (vgl. Beispiel 4.3.5))

Seien X_1, \ldots, X_N unabhängig normalverteilt mit Lokationsparametern μ_1, \ldots, μ_N und Skalenparametern $\sigma_1^2, \ldots, \sigma_N^2$, d. h. $X_n \sim N(\mu_n, \sigma_n^2)$. Dann ist (X_1, \ldots, X_N) absolut stetig und besitzt die Wahrscheinlichkeitsdichte $f_{(X_1, \ldots, X_N)}(x_1, \ldots, x_N) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_N}(x_n)$

$$= \left(\frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}}\right) \cdot \dots \cdot \left(\frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{(x_N - \mu_N)^2}{2\sigma_N^2}}\right)$$

$$= \frac{1}{(2\pi)^{N/2} \sqrt{\sigma_1^2 \cdot \dots \cdot \sigma_N^2}} e^{-1/2 \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \dots + \frac{(x_N - \mu_N)^2}{\sigma_N^2} \right)}. \text{ Setzen wir nun } x = \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}, \ \mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_N \end{pmatrix}$$

und

$$\Sigma = \operatorname{diag} (\sigma_1^2, \dots, \sigma_N^2) := \begin{pmatrix} \sigma_1^2 & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{pmatrix} \in \mathbb{R}^{N \times N}, \operatorname{dann \ l\"{asst \ sich \ die \ Dichte \ darstel-}$$

len als:

$$f_{(X_1,\dots,X_N)}(x) = \frac{1}{(2\pi)^{N/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

(mit $det(\Sigma)$ Determinante von Σ und T als Transpositionszeichen).

Ist Σ irgendeine positiv definite, symmetrische Matrix, d. h. $\bigwedge_{c \in \mathbb{R}^N} c^T \Sigma c > 0$, - also Σ nicht notwendig von Diagonalgestalt -, dann beschreibt \circledast die Dichte einer allgemeinen

N-dimensionalen Normalverteilung mit Lokationsparameter (-vektor) μ und Skalenparameter (-matrix) Σ .

6 Funktionen von Zufallsvektoren

6.1 Projektionen von Zufallsvektoren

Zuerst betrachten wir Funktionen vom Typ $h_n(X_1, \ldots, X_N) = X_n$, d. h. Projektionen.

6.1.1 Definition (Rand- / Marginalverteilung)

Ist $X = (X_1, ..., X_N)$ ein Zufallsvektor, dann heißen die Verteilungen P^{X_n} der Komponenten X_n Rand- oder Marginalverteilungen von X bzgl. der n- ten Komponente, n = 1, ..., N. Entsprechend heißen p_{X_n} bzw. f_{X_n} Rand- oder Marginaldichten.

6.1.2 Folgerung

(i)
$$X = (X_1, ..., X_N)$$
 diskret
$$\Rightarrow X_n \text{ diskret mit diskreter Dichte}$$

$$p_{X_n}(x_n) = \sum_{(x'_1, ..., x'_N): x'_n = x_n} p_{(X_1, ..., X_N)}(x'_1, ..., x'_N)$$

$$\left(= \sum_{x_1} ... \sum_{x_{n-1}} \sum_{x_{n+1}} ... \sum_{x_N} p_{(X_1, ..., X_N)}(x_1, ..., x_N) \right).$$

(ii) $X = (X_1, \dots, X_N)$ absolut stetig

 $\Rightarrow X_n$ absolut stetig und

$$f_{X_n}: x_n \to f_{X_n}(x_n) := \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_N)}(x_1, \dots, x_N) dx_N \dots dx_{n+1} dx_{n-1} \dots dx_1$$

$$\underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_N)}(x_1, \dots, x_N) dx_N \dots dx_{n+1} dx_{n-1} \dots dx_1}_{N-1 \text{ Integrale}}$$

ist Wahrscheinlichkeitsdichte von X_n .

Beweis

(i) Siehe Vorüberlegung im Beweis von Satz 5.3.8 und

$$p_{X_n}(x_n) = P(X_n = x_n)$$

$$= P(X_n = x_n, (X_1, \dots, X_N) \in B^*)$$

$$= P(\{(X_1, \dots, X_N) \in B^*\} \cap \{X_n = x_n\})$$

$$= \sum_{(x'_1, \dots, x'_N) \in B^*, x'_n = x_n} p_{(X_1, \dots, X_N)}(x'_1, \dots, x'_N)$$

$$= \sum_{x_1 \in B_1^*} \dots \sum_{x_{n-1} \in B_{n-1}^*} \sum_{x_{n+1} \in B_{n+1}^*} \dots \sum_{x_N \in B_N^*} p_{(X_1, \dots, X_N)}(x_1, \dots, x_N).$$

#

(ii)

$$P^{X_n}([a_n, b_n]) = P(X_n \in [a_n, b_n])$$

$$= P(X_n \in [a_n, b_n], (X_1, \dots, X_N) \in \mathbb{R}^{\sum_{n=1}^N p_n})$$

$$\stackrel{\oplus}{=} \int_{a_n}^{b_n} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_N)}(x_1, \dots, x_N) dx_N \dots dx_{n+1} dx_{n-1} \dots dx_1 dx_n.$$

Dabei wurde in \oplus der Satz von Fubini ausgenutzt, der besagt, dass die Reihenfolge der Integration vertauscht werden kann.

$$\Rightarrow f: x_n \longrightarrow \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{(X_1, \dots, X_N)}(x_1, \dots, x_N) dx_N \cdot dx_{n+1} dx_{n-1} \cdots dx_1$$

ist Wahrscheinlichkeitsdichte von X_n .

6.1.3 Beispiel (Bivariate (2-dimensionale) Normal-Verteilung)

Sei (X_1, X_2) 2-dimensional normal-verteilt mit Lokationsparameter $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ und

Skalenparameter
$$\sum = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$$
 (siehe 5.3.11) mit $-1 < \rho < 1, \ \sigma_1 > 0, \ \sigma_2 > 0.$

Dann gilt (siehe 5.3.11)

$$(1 - \rho^2)(x - \mu)^T \Sigma^{-1}(x - \mu) = \frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} - 2\frac{\rho}{\sigma_1 \sigma_2}(x_1 - \mu_1)(x_2 - \mu_2)$$

$$= (1 - \rho^2)\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \left(\frac{x_2 - \mu_2}{\sigma_2} - \rho\frac{x_1 - \mu_1}{\sigma_1}\right)^2$$

$$= (1 - \rho^2)\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{1}{\sigma_2^2}\left(x_2 - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)\right)\right)^2.$$

Folglich hat (X_1, X_2) die Dichte

$$f_{(X_1,X_2)}(x_1,x_2) = f_{N(\mu_1,\sigma_1^2)}(x_1) \cdot f_{N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1),(1 - \rho^2)\sigma_2^2)}(x_2),$$

wobei

$$f_{N(\mu_0,\sigma_0^2)}(x_0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{1}{2\sigma_0^2}(x_0-\mu_0)^2}$$

die Dichte der eindimensionalen Normalverteilung mit Lokationsparameter μ_0 und Skalenparameter σ_0^2 ist.

Für festes x_1 ist also $f_{N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2)}$ die Wahrscheinlichkeitsdichte der Normalverteilung mit Lokationsparameter $\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$ und Skalenparameter $(1 - \rho^2)\sigma_2^2$.

Damit gilt

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_2$$

$$= f_{N(\mu_1, \sigma_1^2)}(x_1) \cdot \int_{-\infty}^{\infty} f_{N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2) \sigma_2^2)}(x_2) dx_2$$

$$= f_{N(\mu_1, \sigma_1^2)}(x_1).$$

Analog folgt

$$f_{X_2}(x_2) = f_{N(\mu_2, \sigma_2^2)}(x_2),$$

d. h. die (ein-dimensionalen) Randverteilungen der (zwei-dimensionalen) Normalverteilung sind wieder normal-verteilt mit entsprechenden Lokations- und Skalenparametern.

6.1.4 Lemma (Nützliche Eigenschaft der Normalverteilung) (X_1, X_2) bivariant normalverteilt: X_1, X_2 unabhängig $\Leftrightarrow \rho = 0$.

Beweis

" \Rightarrow ": Siehe Beispiel 5.3.11.

"⇐": Aus Beispiel 6.1.3 folgt

$$f_{(X_1,X_2)}(x_1,x_2) = f_{N(\mu_1,\sigma_1^2)}(x_1) \cdot f_{N(\mu_2,\sigma_2^2)}(x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$
 #

6.2 Summe von Zufallsgrößen

6.2.1 Satz

 (X_1, X_2) diskreter Vektor von Zufallsgrößen mit gemeinsamer diskreter Dichte $p_{(X_1, X_2)}$. Dann gilt für die Zufallsgröße $X_1 + X_2$:

(i)
$$p_{X_1+X_2}(y) = \sum_{x} p_{(X_1,X_2)}(x,y-x)$$

(ii) Sind X_1, X_2 zusätzlich unabhängig, dann gilt

$$p_{X_1+X_2}(y) = \sum_x p_{X_1}(x) p_{X_2}(y-x).$$

Beweis

(i)

$$p_{X_1+X_2}(y) = P(X_1 + X_2 = y)$$

$$= P(\{X_1 + X_2 = y\} \cap \biguplus_{x \in B_1^*} \{X_1 = x\})$$

$$= P(\biguplus_{x \in B_1^*} \{X_1 + X_2 = y, X_1 = x\})$$

$$= \sum_{x \in B_1^*} P(X_1 = x, X_2 = y - x)$$

$$= \sum_{x \in B_1^*} p_{(X_1, X_2)}(x, y - x).$$

- (ii) folgt sofort aus Satz 5.3.8 mit der Charakterisierung der Unabhängigkeit für diskrete Dichten.
- **6.2.2 Bemerkung** (Gemeinsame Dichte von X_1 und $X_1 + X_2$ im diskreten Fall) Bevor wir das Ergebnis 6.2.1 auf den stetigen Fall übertragen können, betrachten wir einen Trick, den wir in 6.2.1 implizit benutzt haben:

Wir bestimmen die gemeinsame Verteilung von X_1 und $X_1 + X_2$ und erhalten die Verteilung von $X_1 + X_2$ als Randverteilung:

$$p_{(X_1,X_1+X_2)}(x,y) = P(X_1 = x, X_1 + X_2 = y)$$

$$= P(X_1 = x, X_2 = y - x) = p_{(X_1,X_2)}(x,y - x)$$

$$\Rightarrow p_{X_1+X_2}(y) = \sum_{x} p_{(X_1,X_1+X_2)}(x,y)$$

$$= \sum_{x} p_{(X_1,X_2)}(x,y - x).$$

6.2.3 Satz

Sei (X_1, X_2) ein absolut stetiger Vektor von Zufallsgrößen mit gemeinsamer Wahrscheinlichkeitsdichte $f_{(X_1,X_2)}$.

(i) Dann ist auch $(X_1, X_1 + X_2)$ absolut stetig mit Wahrscheinlichkeitsdichte

$$f_{(X_1,X_1+X_2)}:(x,y)\to f_{(X_1,X_2)}(x,y-x)$$

(ii) $X_1 + X_2$ ist absolut stetig mit Wahrscheinlichkeitsdichte:

$$f_{X_1+X_2}: y \to f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x,y-x) dx.$$

(iii) Sind X_1, X_2 zusätzlich unabhängig, dann ist die sogenannte **Faltung** der Randdichten

$$f_{X_1+X_2}: y \to f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(y-x) dx$$

eine Wahrscheinlichkeitsdichte von $X_1 + X_2$.

Beweis

(i)

$$P^{(X_1,X_1+X_2)}([a,b]) = P(a_1 \le X_1 \le b_1, a_2 \le X_1 + X_2 \le b_2)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_{[a_1,b_1] \times [a_2,b_2]}(x_1, x_1 + x_2) f_{(X_1,X_2)}(x_1, x_2) dx_2 dx_1$$

$$= \int_{-\infty}^{\infty} 1_{[a_1,b_1]}(x_1) \int_{-\infty}^{\infty} 1_{[a_2,b_2]}(x_1 + x_2) f_{(X_1,X_2)}(x_1, x_2) dx_2 dx_1.$$

Substitution $y = x_1 + x_2$ im inneren Integral liefert

$$P^{(X_1,X_1+X_2)}([a,b])$$

$$= \int_{-\infty}^{\infty} 1_{[a_1,b_1]}(x_1) \int_{-\infty}^{\infty} 1_{[a_2,b_2]}(y) f_{(X_1,X_2)}(x_1,y-x_1) dy dx_1$$

$$= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{(X_1,X_2)}(x_1,y-x_1) dy dx_1.$$

- (ii) Die Randdichte wird entsprechend 6.1.2 gebildet.
- (iii) folgt aus Satz 5.3.10 mit der Charakterisierung der Unabhängigkeit über Wahrscheinlichkeitsdichten.

6.2.4 Beispiele

(i) Binominal-Verteilung

Die Binomial-Verteilung kann als Summe unabhängiger Bernoulli-Experimente hergeleitet werden:

 X_1, \ldots, X_N unabhängig, identisch verteilt mit

$$P(X_n = 1) = p, \ P(X_n = 0) = 1 - p.$$

Dann gilt für $Y_N = \sum_{n=1}^N X_n$:

$$p_{Y_N}(k) = \binom{N}{k} p^k (1-p)^{N-k},$$

d.h. Y_N ist binominal-verteilt mit Parameter N und p (vgl. (4.5.4)).

Interpretation:

$$X_n = 1 \stackrel{\frown}{=} \text{Erfolg}, \ X_n = 0 \stackrel{\frown}{=} \text{Misserfolg}$$

 $Y_N \cong \text{Anzahl der Erfolge in } N \text{ Experimenten.}$

(ii) Normalverteilung:

 (X_1,X_2) bivariant normal-verteilt mit Lokationsparameter $\mu=\begin{pmatrix} \mu_1\\ \mu_2 \end{pmatrix}$ und Skalenparameter

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \text{ (vgl. 6.1.3)}.$$

Dann gilt nach Satz 6.2.3

$$f_{X_1+X_2}(y) = \int_{-\infty}^{\infty} f_{(X_1,X_2)}(x,y-x) \, dx = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \, e^{-\frac{1}{2(1-\rho^2)}g(x,y)} \, dx$$

mit

$$g(x,y) = \frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-x-\mu_2)^2}{\sigma_2^2} - 2\frac{\rho}{\sigma_1\sigma_2}(x-\mu_1)(y-x-\mu_2) = \frac{1}{\sigma_1^2\sigma_2^2}(x^2(\sigma_2^2 + \sigma_1^2 + 2\rho\sigma_1\sigma_2) - 2x(\sigma_2^2\mu_1 + \sigma_1^2(y-\mu_2) + \rho\sigma_1\sigma_2(y-\mu_2) + \rho\sigma_1\sigma_2\mu_1) + \sigma_2^2\mu_1^2 + \sigma_1^2(y-\mu_2)^2 + 2\rho\sigma_1\sigma_2\mu_1(y-\mu_2)).$$

Nach geeigneter quadratischer Ergänzung erhält man:

$$g(x,y) = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{\sigma_1^2\sigma_2^2} (x - c(y,\mu_1,\mu_2))^2 + \frac{(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2} (y - (\mu_1 + \mu_2))^2.$$

Denn mit
$$e := (\sigma_2^2 + \sigma_1^2 + 2\rho\sigma_1\sigma_2)$$
 ergibt sich

$$\begin{split} \frac{\sigma_1^2 \sigma_2^2}{e} &g(x,y) \\ &= x^2 - 2x \, \frac{(\sigma_2 \mu_1 (\sigma_2 + \rho \sigma_1) + \sigma_1 (y - \mu_2) (\sigma_1 + \rho \sigma_2))}{e} \\ &\quad + \frac{(\sigma_2 \mu_1 (\sigma_2 + \rho \sigma_1) + \sigma_1 (y - \mu_2) (\sigma_1 + \rho \sigma_2))^2}{e^2} \\ &\quad - \frac{(\sigma_2 \mu_1 (\sigma_2 + \rho \sigma_1) + \sigma_1 (y - \mu_2) (\sigma_1 + \rho \sigma_2))^2}{e^2} \\ &\quad + \frac{\sigma_2^2 \mu_1^2 + \sigma_1^2 (y - \mu_2)^2 + 2\rho\sigma_1\sigma_2\mu_1 (y - \mu_2)}{e} \\ &\quad = (x - c(y, \mu_1, \mu_2))^2 \\ &\quad + \frac{1}{e^2} [-(\sigma_2\mu_1 (\sigma_2 + \rho \sigma_1) + \sigma_1 (y - \mu_2) (\sigma_1 + \rho \sigma_2))^2 \\ &\quad + (\sigma_2^2 \mu_1^2 + \sigma_1^2 (y - \mu_2)^2 + 2\rho\sigma_1\sigma_2\mu_1 (\mu - \mu_2)) \cdot e] \\ &\quad = (x - c(y, \mu_1, \mu_2))^2 \\ &\quad + \frac{1}{e^2} [-\sigma_2^2 \mu_1^2 (\sigma_2^2 + 2\sigma_2\sigma_1\rho + \rho^2\sigma_1^2) \\ &\quad - \sigma_1^2 (y - \mu_2)^2 (\sigma_1^2 + 2\sigma_1\sigma_2\rho + \rho^2\sigma_2^2) \\ &\quad - 2\sigma_1\sigma_2\mu_1 (y - \mu_2) (\sigma_1\sigma_2 + \rho\sigma_1^2 + \rho\sigma_2^2 + \rho^2\sigma_1\sigma_2) \\ &\quad + (\sigma_2^2 \mu_1^2 + \sigma_1^2 (y - \mu_2)^2 + 2\rho\sigma_1\sigma_2\mu_1 (y - \mu_2)) \cdot e] \\ &\quad = (x - c(y, \mu_1, \mu_2))^2 \\ &\quad + \frac{1}{e^2} \left[(\sigma_2^2 + 2\sigma_1\sigma_1\rho + \sigma_1^2 - e) \left(-\sigma_2^2 \mu_1^2 - \sigma_1^2 (y - \mu_2)^2 - 2\sigma_1\sigma_2\mu_1 (y - \mu_2)\rho \right) \right. \\ &\quad + \sigma_2^2 \mu_1^2 \left(\sigma_1^2 - \rho^2\sigma_1^2\right) \\ &\quad + \sigma_1^2 \left(y - \mu_2\right)^2 (\sigma_2^2 - \rho^2\sigma_2^2\right) \\ &\quad - 2\sigma_1\sigma_2\mu_1 (y - \mu_2) \left(\sigma_1\sigma_2 - \rho^2\sigma_1\sigma_2\right) \right] \\ &\quad = (x - c(y, \mu_1, \mu_2))^2 \\ &\quad + \frac{1}{\sigma_1} [\sigma_1^2 \sigma_2^2 (1 - \rho^2) (\mu_1 - (y - \mu_2))^2]. \end{split}$$

Folglich

$$f_{X_1+X_2}(y) = f_{N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)}(y) \cdot \int_{-\infty}^{\infty} f_{N(c(y,\mu_1,\mu_2),\frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)}}(x)dx$$
$$= f_{N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)}(y),$$

d. h. X_1+X_2 ist normalverteilt mit Lokationsparameter $\mu_1+\mu_2$ und Skalenparameter $\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2$.

Sind insbesondere X_1, X_2 unabhängig $N(0, \sigma^2)$ verteilt, dann hat $X_1 + X_2$ eine $N(0, 2\sigma^2)$ -Verteilung.

6.3 Weitere Abbildungen von Zufallsvektoren

6.3.1 Bemerkung

Ist X ein absolut stetiger Zufallsvektor, so lässt sich die Verteilung von h(X), wobei $h: \mathbb{R}^p \to \mathbb{R}^p$ eine beliebige Funktion ist, die nur bestimmte Regularitätseigenschaften erfüllen muss, mittels der Substitutionsregel für mehrfache Integrale ermitteln. Die Substitutionsregel für mehrfache Integrale, die die Substitutionsregel für einfache Integrale verallgemeinert, lautet wie folgt:

6.3.2 Satz (Substitutionsregel für mehrfache Integrale)

(siehe z. b. Heuser (1981), Lehrbuch der Analysis Teil 2, S. 478)

Die \mathbb{R}^p -wertige Funktion h sei auf der offenen Menge $H \subset \mathbb{R}^p$ injektiv und stetig differenzierbar, und det h'(t) sei auf H entweder ständig positiv oder ständig negativ. Ferner sei T = [a, b] eine Teilmenge von H und f eine auf h(T) stetig reellwertige Funktion. Dann ist f auf h(T) Riemann-integrierbar und es gilt

$$\int_{h(T)} f(x)dx = \int_{T} f(h(t)) |\det h'(t)| dt.$$

Dabei bezeichnet h'(t) die Matrix der 1. Ableitungen der Funktion $h: \mathbb{R}^p \to \mathbb{R}^p$ an der Stelle t und det A die Determinante von A.

6.3.3 Satz (Transformationssatz für Wahrscheinlichkeitsdichten)

Sei X ein absolut stetiger p-dimensionaler Zufallsvektor mit Wahrscheinlichkeitsdichte f_X , $h: \mathbb{R}^p \to \mathbb{R}^p$ eine injektive und stetig differenzierbare Abbildung mit entweder $\det h'(t) > 0$ für alle $t \in \mathbb{R}^p$ oder $\det h'(t) < 0$ für alle $t \in \mathbb{R}^p$. Dann besitzt Y = h(X) die Wahrscheinlichkeitsdichte

$$f_Y : \mathbb{R}^p \ni y \to f_Y(y) = f_X(h^{-1}(y))/|\det h'(h^{-1}(y))|.$$

Beweis

Nach der Substitutionsregel gilt für alle $a, b \in \mathbb{R}^p, a \leq b$,

$$\int_{[a,b]} f_X(h^{-1}(y))/|\det h'(h^{-1}(y))| dy = \int_{h(h^{-1}[a,b])} f_X(h^{-1}(y))/|\det h'(h^{-1}(y))| dy$$

$$= \int_{h^{-1}([a,b])} f_X(h^{-1}(h(t))) \frac{|\det h'(t)|}{|\det h'(h^{-1}(h(t)))|} dt = \int_{h^{-1}([a,b])} f_X(t) dt = P(X \in h^{-1}([a,b]))$$

$$= P(h(X) \in [a,b]) = P(Y \in [a,b]) = P^Y([a,b]).$$

6.3.4 Beispiel (*N*-dimensionale Normalteilung)

Nach Beispiel 5.3.11 hat X eine N-dimensionale Normalverteilung mit Lokationspara-

meter $\mu \in \mathbb{R}^N$ und Skalenparameter $\Sigma \in \mathbb{R}^{N \times N}$ (Σ symmetrisch und positiv definit), falls

$$f_X(x) = \frac{1}{(2\pi)^{N/2} \sqrt{\det \Sigma}} e^{-1/2(x-\mu)^T \sum^{-1} (x-\mu)}$$

gilt. Betrachte nun

$$Y = AX + b$$

mit invertierbarer Matrix $A \in \mathbb{R}^{N \times N}$ und $b \in \mathbb{R}^N$. Dann erfüllt h gegeben durch h(x) = Ax + b die Voraussetzung von Satz 6.3.3 mit h'(x) = A und $h^{-1}(y) = A^{-1}(y - b)$. Nach Satz 6.3.3 hat Y dann die Dichte

$$f_Y(y) = f_X(A^{-1}(y-b))/|\det A|$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{(\det A)^2 \det \Sigma}} e^{-1/2(y-b-A\mu)^T(A^{-1})^T\Sigma^{-1}A^{-1}(y-b-A\mu)}$$

$$= \frac{1}{(2\pi)^{N/2}\sqrt{\det A\Sigma A^T}} e^{-1/2(y-(b+A\mu))^T(A\Sigma A^T)^{-1}(y-(b+A\mu))},$$

d. h. Y besitzt wieder eine N-dimensionale Normalverteilung mit Lokationsparameter $A\mu + b$ und Skalenparameter $A\Sigma A^T$.

7 Stochastische Abhängigkeit

7.1 Bedingte Dichten bei diskreten Zufallsvariablen

7.1.1 Motivation

Für Ereignisse B mit P(B) > 0 war die bedingte Wahrscheinlichkeit eines weiteren Ereignisses A bei gegebenem B (Vorkenntnis, dass B eingetreten ist), definiert als

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

vgl. Definition 5.1.2). Wir sind jetzt interessiert an Wahrscheinlichkeitsaussagen über eine Zufallsgröße X mit der Vorkenntnis, dass eine andere Zufallsgröße Y einen bestimmten Wert y (Realisierung) angenommen hat. Sind X, Y diskret und gilt P(Y = y) > 0, dann gilt:

$$P({X = x} | {Y = y}) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

bzw.

$$P(\{X \le x\} \mid \{Y = y\}) = \frac{P(X \le x, Y = y)}{P(Y = y)}.$$

Für y mit P(Y = y) > 0 ist nach Satz 5.1.4

$$P(\bullet \mid \{Y = y\}) : \mathscr{A} \ni A \to P(A \mid \{Y = y\})$$

ein Wahrscheinlichkeitsmaß auf \mathscr{A} . Folglich ist auch (für y mit P(Y=y)>0)

$$P^{X \mid Y=y}: \mathcal{B} \ni B \to P(X \in B \mid \{Y=y\})$$

ein Wahrscheinlichkeitsmaß auf \mathcal{B} , das von X induzierte (transportierte) Bildmaß des bedingten Wahrscheinlichkeitsmaßes auf \mathcal{A} bei gegebenem $\{Y = y\}$ (vgl. Definition 3.2.1).

7.1.2 Definition

$$(X,Y)$$
 diskret, $\mathscr{Y}^*:=\{y;\ P(Y=y)=p_Y(y)>0\}.$ Für $y\in\mathscr{Y}^*$ heißt

- (i) $p_{X \mid Y=y} : \mathbb{R} \ni x \to p_{X \mid Y=y}(x) := P(\{X=x\} \mid \{Y=y\}) = \frac{p_{(X,Y)}(x,y)}{p_Y(y)} \in [0,1]$ bedingte diskrete Dichte (der Verteilung) von X bei gegebenem Y=y,
- (ii) $P^{X \mid Y=y}: \mathcal{B} \ni B \to P^{X \mid Y=y}(B) = P(\{X \in B\} \mid \{Y=y\}) = \sum_{z \in B} p_{X \mid Y=y}(z) \in [0,1]$ bedingte Wahrscheinlichkeitsverteilung von X bei gegebenem Y=y,

(iii)
$$F_{X \mid Y=y} : \mathbb{R} \ni x \to F_{X \mid Y=y}(x) = P(\{X \le x\} \mid \{Y=y\}) = P^{X \mid Y=y}((-\infty, x])$$

= $\sum_{z \le x} p_{X \mid Y=y}(z) \in [0, 1]$

bedingte Verteilungsfunktion von X bei gegebenem Y = y.

7.1.3 Beispiele

(i) Indikatorenfunktionen

(ii) Binominal-Verteilungen

Es soll die Haltbarkeit von Teetassen untersucht werden. In einer ersten Versuchsreihe werden N Teetassen aus Trinkhöhe fallen gelassen. Es wird

X = Anzahl der Tassen, die**nicht**zerbrochen sind,

beobachtet.

In einer zweiten Versuchsreihe werden unter gleichen Versuchsbedingungen nochmals M Teetassen getestet. Dazu wird

Y = Anzahl der Tassen, die in der 2. Versuchsreihe**nicht**zerbrochen sind

beobachtet. Wie sieht die Verteilung der "überlebenden" Teetassen in der ersten Versuchsreihe aus, wenn bekannt ist, das insgesamt in den beiden Versuchsreihen k von den N+M Teetassen nicht zerbrochen sind?

7.1.4 Satz (Wichtige Eigenschaft)

(i) Für
$$y \in \mathscr{Y}^* = \{y; p_Y(y) > 0\}$$
 gilt $p_{(X,Y)}(x,y) = p_{X|Y=y}(x) \cdot p_Y(y)$.

(ii)
$$P^{(X,Y)}(B_1 \times B_2) = \sum_{y \in B_2 \cap \mathscr{Y}^*} P^{(X,Y)}(B_1 \times \{y\})$$

 $= \sum_{y \in B_2 \cap \mathscr{Y}^*} P(X \in B_1, Y = y) = \sum_{y \in B_2 \cap \mathscr{Y}^*} \sum_{x \in B_1} p_{(X,Y)}(x,y)$
 $= \sum_{y \in B_2 \cap \mathscr{Y}^*} \sum_{x \in B_1} p_{X|Y=y}(x) \cdot p_Y(y) = \sum_{y \in B_2 \cap \mathscr{Y}^*} P^{X|Y=y}(B_1) \cdot p_Y(y).$

Definiert man $p_{X|Y=y}$ bzw. $P^{X|Y=y}$ für $y \in \overline{\mathscr{Y}}^*$ beliebig, so bleibt wegen $p_{(X,Y)}(x,y) \le p_Y(y) = 0$ (i) gültig und in (ii) kann über alle $y \in B_2$ summiert werden.

7.1.5 Definition (Allgemeine Definition bedingter diskreter Dichten) Sei (X, Y) diskret.

(i) $p_{X|Y=y}$ heißt bedingte diskrete Dichte (der Verteilung) von X bei gegebenem Y=y

 \Leftrightarrow

- (α) $p_{X|Y=y}$ ist diskrete Dichte und
- $(\beta) \quad \bigwedge_{x} p_{(X,Y)}(x,y) = p_{X|Y=y}(x)p_{Y}(y),$
- (ii) $P^{X|Y=y}: \mathcal{B} \ni B \to P^{X|Y=y}(B) = \sum_{x \in B} p_{X|Y=y}(x) \in [0,1]$ heißt bedingte Wahrscheinlichkeitsverteilung von X bei gegebenem Y=y.
- (iii) $F_{X|Y=y}: \mathbb{R} \ni x \to F_{X|Y=y}(x) = \sum_{z \leq x} p_{X|Y=y}(z) \in [0,1]$ heißt bedingte Verteilungsfunktion von X bei gegebenem Y=y.

7.1.6 Bemerkung

Wegen Satz 7.1.4 ist die Definition 7.1.5 mit der ursprünglichen Definition 7.1.2 verträglich. Sie ist aber insofern besser handbar, als man aus bedingten Dichten und Randdichten eine gemeinsame Verteilung konstruieren kann und dass die Verallgemeinerung auf den absolut stetigen Fall naheliegend ist.

7.2 Bedingte Dichten bei absolut stetigen Zufallsvariablen

7.2.1 Definition

Sei (X,Y) absolut stetig.

(i) $f_{X|Y=y}$ heißt bedingte Wahrscheinlichkeitsdichte (der Verteilung) von X bei gegebenem Y=y

 \Leftrightarrow

- (α) $f_{X|Y=y}$ ist Wahrscheinlichkeitsdichte und
- $(\beta) \bigwedge_{x} f_{(X,Y)}(x,y) = f_{X|Y=y}(x) f_{Y}(y).$
- (ii) $P^{X|Y=y}$ mit $P^{X|Y=y}([a,b]) = \int_a^b f_{X|Y=y}(x) dx$ heißt **bedingte Wahrscheinlich**keitsverteilung von X bei gegebenem Y=y.
- (iii) $F_{X|Y=y}: \mathbb{R} \ni x \longrightarrow F_{X|Y=y}(x) = \int_{-\infty}^{x} f_{X|Y=y}(z) dz \in [0,1]$ heißt bedingte Verteilungsfunktion von X bei gegebenem Y=y.

7.2.2 Satz (Wichtige Eigenschaften)

- (i) Für y mit $f_Y(y) > 0$ gilt $f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}$.
- (ii) Für y mit $f_Y(y) = 0$ kann $f_{X|Y=y}$ beliebig gewählt werden.

(iii)
$$P^{(X,Y)}([a,b]) = \int_{a_2}^{b_2} P^{X|Y=y}([a_1,b_1]) f_Y(y) dy.$$

Beweis

(i) + (ii): klar.

(iii)
$$P^{(X,Y)}([a,b]) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{(X,Y)}(x,y) dx dy = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{X|Y=y}(x) f_Y(y) dx dy$$

$$= \int_{a_2}^{b_2} \int_{a_1}^{b_1} f_{X|Y=y}(x) \ dx \ f_Y(y) \ dy = \int_{a_2}^{b_2} P^{X|Y=y}([a_1, b_1]) \ f_Y(y) \ dy.$$

7.2.3 Bemerkung (Nachträgliche Motivation)

Das Problem der Definition bedingter Verteilungen bei absolut stetiger Verteilung der bedingten Zufallsvariable Y besteht darin, dass für die Bedingung Y = y immer gilt P(Y = y) = 0. Es liegt aber nahe, die Bedingung Y = y durch eine Näherung "Y nahe bei Y" zu approximieren. Ist $P^Y([y, y + h]) > 0$, so ist die bedingte Wahrscheinlichkeit

$$P(X \in [a, b] | Y \in [y, y + h])$$

definiert.

Wir nehmen den Idealfall an, dass f_Y stetig ist in einer Umgebung von y und $f_Y(y) > 0$ ist. Dann können wir betrachten:

$$\lim_{h \downarrow 0} P(X \in [a, b] \mid Y \in [y, y + h]) = \lim_{h \downarrow 0} \frac{P^{(X, Y)}([a, b] \times [y, y + h])}{P^{Y}([y, y + h])}$$

$$= \lim_{h\downarrow 0} \frac{\int\limits_{y}^{y+h} \int\limits_{a}^{b} f_{(X,Y)}(x,\widetilde{y}) \, dx \, d\widetilde{y}}{\int\limits_{y}^{y+h} f_{Y}(\widetilde{y}) \, d\widetilde{y}} = \frac{\int\limits_{a}^{b} f_{(X,Y)}(x,y) \, dx}{f_{Y}(y)} = \int\limits_{a}^{b} \frac{f_{(X,Y)}(x,y) \, dx}{f_{Y}(y)} \, dx.$$

7.2.4 Beispiele (i) (X,Y) gleichförmig verteilt auf dem Einheitskreis, d. h.

$$f_{(X,Y)}(x,y) = \frac{1}{\pi} 1_{\{(x,y); x^2 + y^2 \le 1\}}(x,y).$$

Daraus folgt für die Randverteilung

$$f_Y(y) = \frac{1}{\pi} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} dx \ 1_{[-1,1]}(y) = \frac{2}{\pi} \sqrt{1-y^2} \ 1_{[-1,1]}(y)$$

und somit für die bedingte Verteilung für |y| < 1:

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{1-y^2}} \mathbb{1}_{[-\sqrt{1-y^2},\sqrt{1-y^2}]}(x)$$

d. h. bei gegebenem Y=y $(y\in (-1,1))$ ist die bedingte Verteilung von X die gleichförmige Verteilung auf $[-\sqrt{1-y^2},\sqrt{1-y^2}]$.

(ii) Bivariante Normalverteilung

Bei der Berechnung der Randverteilung der bivarianten Normalverteilung (vgl. Beispiel 6.1.3) hatten wir gezeigt:

$$f_{(X,Y)}(x,y) = f_{N(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y - \mu_2), (1 - \rho^2)\sigma_1^2)}(x) \cdot f_{N(\mu_2, \sigma_2^2)}(y) ,$$

wobei $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ der Lokationsparameter und $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}$ der Skalenparameter der gemeinsamen Verteilung ist. Folglich ist

$$f_{N(\mu_1+\rho\frac{\sigma_1}{\sigma_2}(y-\mu_2), (1-\rho^2)\sigma_1^2)}$$

eine bedingte Wahrscheinlichkeitsdichte von X bei gegebenem Y=y, d. h. die bedingte Verteilung ist wieder eine Normalverteilung, und zwar mit Lokationsparameter $\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(y-\mu_2)$ linear abhängig von y (sofern $\rho \neq 0$) und von y nicht abhängendem Skalenparameter $(1-\rho^2)\sigma_1^2$.

7.2.5 Bemerkung (Grenzen der Methodik)

Bisher können nicht behandelt werden:

- (i) Mischungen von Verteilungen, wobei die bedingte Verteilung von anderem Typ ist als die Randverteilung. Z. B. gibt es, einen Zusammenhang zwischen Poissonverteilung und Exponentialverteilung.
- (ii) Verteilungen mit diskretem und absolut stetigen Anteil, z. B. $(X_1, \max(X_1, X_2))$ und $P^{\max(X_1, X_2)|X_1=x}$ für absolut stetige X_1, X_2 .

Dafür wird eine allgemeine Theorie basierend auf Maß- und Integrationstheorie benötigt. Siehe Statistik V oder z. B. Bauer (1978), Kapitel X oder Billingssley (1995), Chapter 6.

8 Maßzahlen für Verteilungen

8.1 Erwartungswert

8.1.1 Motivation

Hat man eine endliche Menge von Objekten, an deren Gewicht, Größe oder einer anderen reellwertigen Eigenschaft man interessiert ist, so liegt es nahe, als durchschnittliches Gewicht, Größe, etc. das arithmetische Mittel zu wählen, d. h. das Gesamtgewicht, die Summe aller Größen etc. geteilt durch die Anzahl der Objekte. Diese "durchschnittliche" Zahl zur Beschreibung des Gewichts etc. stellt eine Maßzahl für die Eigenschaft dar.

Wählt man ein Objekt zufällig aus, so kann man vorher – außer in trivialen Fällen – nicht vorhersagen, welches Gewicht, Größe, etc. man beobachten wird. Kennt man aber die Wahrscheinlichkeiten, mit denen die Objekte auftreten, kann man ein gewichtetes arithmetisches Mittel bilden, wobei die Gewichte die Wahrscheinlichkeiten sind. Diese Maßzahl ist insbesondere eine Maßzahl für die "Lage" der Objekte.

8.1.2 Definition (Erwartungswert für diskrete und absolut stetige Zufallsgrößen)

(i) Ist X eine diskrete Zufallsgröße mit Dichte p_X und gilt $\sum_x |x| p_X(x) < \infty$, dann heißt

$$E(X) := \sum_{x} x \ p_X(x)$$

Erwartungswert von X.

(ii) Ist X eine absolut stetige Zufallsgröße mit Dichte f_X und gilt $\int |x| f_X(x) dx < \infty$, dann heißt

$$E(X) := \int x \, f_X(x) \, dx$$

Erwartungswert von X.

(iii) Die Schreibweise

$$E(X) := \int x \ P^X(dx)$$

bedeutet $E(X) := \sum_{x} x \ p_X(x)$ im diskreten Fall und $E(X) := \int x \ f_X(x) \ dx$ im absolut stetigen Fall.

8.1.3 Bemerkung

(i) Der Erwartungswert von X hängt nur von der Verteilung P^X von X ab, d.h. besitzen zwei Zufallsgrößen X und Y die selbe Verteilung $P^X = P^Y$, dann sind auch die Erwartungswerte gleich.

- (ii) Gilt $\sum_{x} |x| p_X(x) = \infty$ bzw. $\int |x| f_X(x) dx = \infty$, so sagt man: "Der Erwartungswert von X existiert nicht" oder "X besitzt keinen (endlichen) Erwartungswert". Es gilt $\sum_{x} |x| p_X(x) < \infty$ genau dann, wenn die Reihe $\sum_{n=1}^{\infty} x_n p_X(x_n)$ absolut summierbar ist.
- (iii) Diskreter Fall: Man stelle sich einen selbst masselosen Stab vor, an dem an den Punkten x Massen $p_X(x)$ sich befinden. E(X) ist dann der Schwerpunkt (das Moment) dieses Stabes.
- (iv) Für absolut stetige Zufallsgrößen X lässt sich $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$ ähnlich wie im diskreten Fall als Schwerpunkt der Fläche unter der Dichte f_X interpretieren.
- (v) Die Schreibweise $E(X) := \int x \ P^X(dx)$ bedeutet, dass der Erwartungswert ein Maßintegral bezüglich des Maßes P^X ist. Siehe dazu die Maßtheorie.
- (vi) Ist Ω abzählbar, so gilt

$$E(X) = \sum_{x} x \ p_X(x) = \sum_{x} x \ P(X = x) = \sum_{x} x \sum_{\omega: X(w) = x} P(\{w\})$$
$$= \sum_{x} \sum_{\omega: X(w) = x} X(w) \ P(\{w\}) = \sum_{\omega \in \Omega} X(w) \ P(\{w\}) \ .$$

In der maßtheoretischen Schreibweise bedeutet dies

$$E(X) := \int x P^X(dx) = \int X(\omega) P(d\omega).$$

Für den absolute stetigen Fall gilt die gleiche Gleichheit, die in der Maßtheorie sehr einfach aus dem sogenannten Transformationssatz der Maßtheorie folgt. Sie ist ein Spezialfall aus dem folgenden Satz.

8.1.4 Satz (Transformationssatz für den Erwartungswert)

Ist X eine Zufallsvariable mit Werten in $(\mathcal{X}, \mathcal{B})$ und $g : \mathcal{X} \ni x \to g(x) \in \mathbb{R}$ eine reellwertige (messbare) Funktion auf \mathcal{X} , so dass $g \circ X$ Zufallsgröße ist. Dann gilt

$$E(g(X)) = \int g(x) P^X(dx)$$
, sofern $\int |g(x)| P^X(dx) < \infty$ existiert.

(i) Ist X ein diskreter Zufallsvektor mit Dichte p_X , so gilt insbesondere

$$E(g(X)) = \sum_{x} g(x) p_X(x)$$
, falls $\sum_{x} |g(x)| p_X(x) < \infty$ existient.

(ii) Ist X ein absolut stetiger Zufallsvektor mit Dichte f_X , so gilt insbesondere

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n,$$

falls
$$\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$$
 existiert.

Beweis

Der allgemeine Fall sowie (ii) folgt aus dem Transformationssatz für Maß-Integrale (siehe Statistik V)). Die Aussage (i) kann aber wieder auch direkt gezeigt werden:

Mit X ist auch $Y := g \circ X$ eine diskrete Zufallsgröße. Nach Definition 8.1.2 gilt

$$E(g(X)) = E(Y) = \sum_{y} y \ p_{Y}(y) = \sum_{y} y \ P(\{\omega; \ g(X(w)) = y\})$$

$$= \sum_{y} \sum_{x: g(x) = y} P(\{\omega; \ X(w) = x\}) = \sum_{y} \sum_{x: g(x) = y} g(x) \ p_{X}(x) = \sum_{x} g(x) \ p_{X}(x),$$

und man sieht, dass

$$\sum_{y} |y| \ p_Y(y) < \infty \Leftrightarrow \sum_{x} |g(x)| \ p_X(x) < \infty$$

gilt, wenn man die obige Rechnung mit |y| bzw. |g(x)| anstelle von y bzw. g(x) durchführt. \sharp

8.1.5 Beispiele

- (i) Indikatorfunktionen (Bernoulli-Variable)
- (ii) Poisson-Verteilung

- (iii) Geometrische Verteilung
- (iv) Exponential-Verteilung

8.1.6 Satz (Eigenschaften des Erwartungswertes)

Seien $X: \Omega \to \mathscr{X}, Y: \Omega \to \mathcal{Y}$ Zufallsvariable, $g: \mathscr{X} \to \mathbb{R}, h: \mathcal{Y} \to \mathbb{R}$ messbare Funktionen, $a, b \in \mathbb{R}$. Unter der Bedingung, dass die betrachteten Erwartungswerte existieren, gilt dann:

- (i) $g(X) \ge 0 \Rightarrow E(g(X)) \ge 0$.
- (ii) $g(X) \ge 0$ und $P(g(X) > 0) > 0 \Rightarrow E(g(X)) > 0$.
- (iii) Linearität:

$$E(a \ q(X) + b \ h(Y)) = a \ E(q(X)) + b \ E(h(Y)).$$

(iv) X, Y unabhängig $\Rightarrow E(g(X) \cdot h(Y)) = E(g(X)) \cdot E(h(Y)).$

Beweis

In der Darstellung des Erwartungswertes als Maß-Integral folgen (i) und (iii) sofort aus den Eigenschaften des Maß-Integrals (siehe Maßtheorie). Ohne Maßtheorie müssen der diskrete und der absolut stetige Fall getrennt bewiesen werden. Das geht folgendermaßen:

(i)
$$E(g(X)) = \left\{ \begin{array}{ll} \sum\limits_{x} g(x) \ p_X(x) & \geq \sum\limits_{x} 0 \cdot p_X(x) \\ \int\limits_{-\infty}^{x} g(x) \ f_X(x) \ dx & \geq \int\limits_{-\infty}^{x} 0 \cdot f_X(x) \ dx \end{array} \right\} = 0.$$

(iii) Ist $\ell:(x,y)\to a\ g(x)+b\ h(y)$, dann gilt nach Satz 8.1.4

$$E(a g(X) + b h(Y)) = E(\ell(X, Y))$$

$$= \begin{cases} \sum_{(x,y)} \ell(x,y) \ p_{(X,Y)}(x,y) \\ \sum_{-\infty}^{\infty} \sum_{-\infty} \ell(x,y) \ f_{(X,Y)}(x,y) \ dydx \end{cases}$$

$$= \begin{cases} \sum_{\substack{x \ y \\ \infty \ \infty}} \sum_{y} (a \ g(x) + b \ h(y)) \ p_{X,Y)}(x,y) \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a \ g(x) + b \ h(y)) \ f_{(X,Y)}(x,y) \ dy dx \end{cases}$$

$$= \begin{cases} \sum_{\substack{x \ y \\ \infty \ \infty}} a \ g(x) \ p_{(X,Y)}(x,y) & + \sum_{\substack{x \ y \\ \infty \ \infty}} b \ h(y) \ p_{X,Y)}(x,y) \\ \int_{-\infty - \infty}^{\infty} \int_{-\infty}^{\infty} a \ g(x) \ f_{(X,Y)}(x,y) \ dy dx & + \int_{-\infty - \infty}^{\infty} \int_{-\infty}^{\infty} b \ h(y) \ f_{(X,Y)}(x,y) \ dy dx \end{cases}$$

$$= \left\{ \begin{array}{ll} a\sum\limits_{x}g(x)\sum\limits_{y}p_{(X,Y)}(x,y) & +b\sum\limits_{y}h(y)\sum\limits_{x}p_{(X,Y)}(x,y) \\ a\int\limits_{-\infty}^{\infty}g(x)\int\limits_{-\infty}^{\infty}f_{(X,Y)}(x,y)\;dxdy & +b\int\limits_{-\infty}^{\infty}h(y)\int\limits_{-\infty}^{\infty}f_{(X,Y)}(x,y)\;dxdy \end{array} \right.$$

$$= \begin{cases} a \sum_{x} g(x) p_X(x) & +b \sum_{y} h(y) p_Y(y) \\ \infty & \infty \\ a \int_{-\infty}^{x} g(x) f_X(x) dx & +b \int_{-\infty}^{x} h(y) f_Y(y) dy \end{cases}$$

$$= a E(g(X)) + b E(h(Y)).$$

Dabei sind die Umsummation (bzw. Vertauschung der Integrale) möglich, da die Summen absolut konvergieren (bzw. da $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx$, $\int_{-\infty}^{\infty} |h(y)| f_y(y) dy$ endlich sind und der Satz von Fubini (siehe Maßtheorie) angewendet werden kann).

(ii) folgt aus (i) durch folgende Betrachtungen: Aus P(g(X) > 0) > 0 folgt die Existenz eines a > 0 mit $P(g(X) \ge a) = p > 0$. Denn würde $P(g(X) \ge a) = 0$ für alle a > 0 gelten, so würde aus der σ -Stetigkeit der Widerspruch $0 = \lim_{a \downarrow 0} P(g(X) \ge a) = P(g(X) > 0) > 0$ folgen. Sei nun $h: \mathcal{X} \ni x \to h(x) = a \, 1_{\{x; \, g(x) \ge a\}}(x) \in \mathbb{R}$. Dann gilt $h(x) \le g(x)$ bzw. $q(x) = g(x) - h(x) \ge 0$ für alle $x \in \mathcal{X}$. Aus (i) und (iii) folgt $E(g(X)) - E(h(X)) = E(q(X)) \ge 0$ und somit mit Beispiel 8.1.5(i)

$$0$$

(iv) X, Y unabhängig

$$\Rightarrow \begin{cases} p_{(X,Y)}(x,y) = p_X(x) \cdot p_Y(y) \\ f_{(X,Y)}(x,y) = f_X(x) \cdot f_Y(y) \end{cases}$$

nach Satz 5.3.8 und Satz 5.3.10.

Für $\ell(x,y) = g(x) \cdot h(y)$ gilt dann nach Satz 8.1.4

$$E(g(X) \cdot h(Y)) = E(\ell(X, Y))$$

$$= \begin{cases} \sum\limits_{(x,y)} \ell(x,y) \ p_{(X,Y)}(x,y) \\ \sum\limits_{-\infty}^{\infty} \sum\limits_{-\infty} \ell(x,y) \ f_{(X,Y)}(x,y) \ dydx \end{cases} = \begin{cases} \sum\limits_{(x,y)} g(x) \ h(y) \ p_X(x) \ p_Y(y) \\ \sum\limits_{\infty}^{\infty} \sum\limits_{\infty} G(x) \ h(y) \ f_X(x) \ f_Y(y) \ dydx \end{cases}$$

$$= \begin{cases} \sum_{x} g(x) \ p_X(x) \sum_{y} h(y) \ p_Y(y) \\ \sum_{x=0}^{\infty} g(x) \ f_X(x) \left(\int_{-\infty}^{\infty} h(y) \ f_Y(y) \ dy \right) dx \end{cases} = \begin{cases} \sum_{x} g(x) \ p_X(x) \ E(h(Y)) \\ \sum_{x=0}^{\infty} g(x) \ E(h(Y)) \ f_X(x) \ dx \end{cases}$$

$$=E(g(X))\cdot E(h(Y)).$$
 #

8.1.7 Satz (Jensensche Ungleichung)

Sei $R = \mathbb{R}$ oder R ein Intervall in \mathbb{R} , $X : \Omega \to R$ eine Zufallsgröße mit Werten in R und $\phi : R \to \mathbb{R}$ eine konvexe Funktion, d.h. es gilt $\alpha\phi(x) + (1-\alpha)\phi(y) \ge \phi(\alpha x + (1-\alpha)y)$ für alle $x, y \in R$ und $\alpha \in [0, 1]$. Dann liegt E(X) in R und es gilt

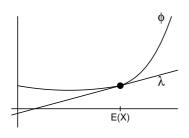
$$\phi(E(X)) \le E(\phi(X)).$$

Beweis

Die Behauptung $E(X) \in R$ folgt aus Satz 8.1.6(i) und (ii). Ist zum Beispiel R = (c, d), so gilt für g(X) = X - c > 0

$$P(g(X) > 0) \ge P(X \in R) = 1$$

und somit mit 8.1.6(ii) und (iii) E(X) - c = E(g(X)) > 0, also E(X) > c. E(X) < d folgt analog.



Wegen der Konvexität von ϕ gibt es zu jedem $x_0 \in R$ eine Gerade $\lambda(x) = a + bx$ mit $\lambda(x_0) = \phi(x_0)$ und $\lambda(x) \leq \phi(x)$ für alle $x \in R$. Sei nun $x_0 = E(X)$. Dann gilt mit Satz 8.1.6(i) und (iii) $E(\phi(X)) \geq E(\lambda(X)) = a + b E(X) = \lambda(E(X)) = \phi(E(X))$.

8.1.8 Beispiele

(i) Binomialverteilung

(ii) Hypergeometrische Verteilung

Ähnlich wie in (i) läßt sich die hypergeometrische Verteilung mit Parametern N und (R, S) als Verteilung der Summe von N identischen Bernoulli-Variablen X_n , n = 1, ..., N, mit Erfolgswahrscheinlichkeit $p = \frac{R}{R+S}$ interpretieren $(X_n = 1 = n - te$ Kugel ist rot). Jedoch sind hierbei die $X_1, ..., X_N$ nicht unabhängig. Eine direkte Summation erfolgt nach folgendem Schema, wobei man noch genau auf die Randbedingungen $x \leq R$ und $N - x \leq S$

zusätzlich zu $0 \le x \le N$ achten muss.

$$E(X) = \sum_{\substack{0 \le x \le N \\ x \le R \\ N - x \le S}} x \frac{\binom{\binom{R}{x}\binom{S}{N-x}}{\binom{R+S}{N}}}{\binom{R+S}{N-x}}$$

$$= \sum_{\substack{1 \le x \le N \\ x \le R \\ N-x \le S}} \frac{R!}{(x-1)!(R-x)!} \binom{S}{N-x} \left/ \frac{(R+S)!}{N!(R+S-N)!} \right.$$

$$= N \frac{R}{R+S} \sum_{\substack{1 \le x \le N \\ x-1 \le R-1 \\ N-x \le S}} \frac{(R-1)!}{(x-1)!(R-x)!} \binom{S}{N-x} \left/ \frac{(R+S-1)!}{(N-1)!(R+S-N)!} \right.$$

$$= N \frac{R}{R+S} \sum_{\substack{1 \le x \le N \\ x-1 \le R-1 \\ N-x \le S}} \frac{\binom{R-1}{x}\binom{S}{N-1-x}}{\binom{R+S-1}{N-1}} = N \frac{R}{R+S},$$

$$\sum_{\substack{1 \le x \le N-1 \\ x \le R-1 \\ N-1-x \le R+S-1-(R-1)}} \frac{\binom{R-1}{x}\binom{S}{N-1-x}}{\binom{R+S-1}{N-1}} = N \frac{R}{R+S},$$

wobei die Summanden im vorletzten Gleichungsglied wieder die Werte der diskreten Dichte $p_{H(N-1; R-1, S)}(x)$ einer hypergeometrischen Verteilung mit Parametern (N-1) und (R-1, S) ist.

8.2 Streuungsmaßzahl (Varianz)

8.2.1 Motivation (Streuungsmaßzahl)

Neben der Frage, wo eine Verteilung konzentriert ist, d. h. wo eine Beobachtung einer entsprechend verteilten Zufallsgröße zu erwarten ist, ist es auch interessant zu wissen, wie weit die Verteilung um diese Lagemaßzahl streut, d. h. wie dicht an diese Lagemaßzahl eine Beobachtung zu erwarten ist. Um diesen Abstand zu messen, muss ein Abstandsbegriff gewählt werden. Aus technischen, wie auch formellen (interpretierbaren Gründen), bietet sich der quadratische Abstand an; und hiervon der Erwartungswert – also der erwartete quadratische Abstand:

$$E((X-a)^2),$$

wobei a eine Maßzahl für die Lage sein soll. Es liegt nahe, für die Lagemaßzahl a^* den Wert zu wählen, so dass der erwartete quadratische Abstand minimiert wird:

$$E((X - a^*)^2) = \min_{a} E((X - a)^2).$$

8.2.2 Lemma

$$E((X - a^*)^2) = \min_{a} E((X - a)^2) \iff a^* = E(X).$$

Beweis

Für beliebiges $a \in \mathbb{R}$ gilt:

$$E((X - a)^{2}) = E((X - E(X) + E(X) - a)^{2})$$

$$= E((X - E(X))^{2} + 2(X - E(X))(E(X) - a) + (E(X) - a)^{2})$$

$$\stackrel{8.1.6(iii)+8.1.5(i)}{=} E((X - E(X))^{2}) + 2(E(X) - a) \underbrace{E((X - E(X)))}_{=0} + (E(X) - a)^{2}$$

$$= E((X - E(X))^{2}) + (E(X) - a)^{2}$$

$$\geq E((X - E(X))^{2})$$

$$\text{mit "="} \Leftrightarrow a = E(X).$$

8.2.3 Definition (Varianz)

$$var(X) := E((X - E(X))^2)$$

heißt Varianz (der Verteilung) von X, falls der Ausdruck auf der rechten Seite (endlich) existiert.

Ħ

8.2.4 Bemerkung

Ähnlich wie sich der Erwartungswert als Schwerpunkt der Verteilung interpretieren lässt, lässt sich die Varianz als die physikalische Größe des Trägheitsmomentes interpretieren.

8.2.5 Satz (Eigenschaften der Varianz)

(i)
$$var(X) = E(X^2) - (E(X))^2$$

(ii)
$$var(aX + b) = a^2 var(X)$$

Beweis

(i)
$$\operatorname{var}(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + (E(X))^2)$$

$$\stackrel{8.1.6(iii)+8.1.5(i)}{=} E(X^2) - 2E(X)E(X) + (E(X))^2$$

$$= E(X^2) - (E(X))^2.$$

(Formal müsste man noch nachprüfen, dass die Existenz von var(X) die von $E(X^2)$ und E(X) nach sich zieht und umgekehrt. Übung!)

(ii) Wegen der Linearität des Erwartungswertes 8.1.6(iii) gilt E(aX - b) = aE(X) + b, also

$$\begin{aligned} \text{var}(aX+b) &= E(((aX+b)-E(aX+b))^2) \\ &= E(((aX+b)-aE(X)-b)^2) = E((a(X-E(X)))^2) \\ &= a^2 E((X-E(X))^2) = a^2 \text{var}(X). \end{aligned}$$

8.2.6 Lemma

Existiert $E(X^2)$ so existieren auch E(X) und var(X).

Beweis

Aus der Existenz von $E(X^2)(<\infty)$ folgt:

$$E(|X|) \le E(X^2 + 1) = E(X^2) + 1 < \infty \implies E(X)$$
 existiert

und damit

$$\begin{aligned} \operatorname{var}(X) &= E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2) \leq E(X^2 + 2|X| \, |E(X)| + E(X)^2) \\ &\leq E(2X^2 + 2(E(X))^2) = 2E(X^2) + 2(E(X))^2 < \infty. \end{aligned}$$

(Tatsächlich gilt immer $var(X) \le E(X^2)$, vgl. Satz 8.2.5(i).) \sharp

8.2.7 Beispiele

- (i) Indikatorfunktionen (Bernoulli-Variable)
- (ii) Poisson-Verteilung
- (iii) Normal-Verteilung

Der Lokations- bzw. Skalenparameter der $\mathcal{N}(\mu, \sigma^2)$ -Verteilung ist gerade der Erwartungswert bzw. die Varianz jeder $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsgröße.

(iv) Rechteck-Verteilung (Gleichförmige Verteilung)

Gilt $X \sim R[a, b]$ $(X \sim U(a, b))$, so gilt

$$E(X) = \frac{a+b}{2}$$
 und $var(X) = \frac{1}{12}(b-a)^2$.

Herleitung ist Übung!

(v) Exponential-Verteilung

8.2.8 Satz (Formel von Bienaymé)

Sind X_1, \ldots, X_N unabhängige Zufallsgrößen, deren Varianzen $var(X_n)$, $n = 1, \ldots, N$, existieren, dann gilt

$$\operatorname{var}\left(\sum_{n=1}^{N}X_{n}\right)=\sum_{n=1}^{N}\operatorname{var}(X_{n}).$$

Beweis

$$\begin{aligned} \operatorname{var}\left(\sum_{n=1}^{N} X_{n}\right) &= E\left(\left(\sum_{n=1}^{N} X_{n} - E\left(\sum_{n=1}^{N} X_{n}\right)\right)^{2}\right) \\ &= E\left(\left(\sum_{n=1}^{N} X_{n} - E(X_{n})\right)\right)^{2}\right) \\ &= E\left(\sum_{n=1}^{N} \sum_{m=1}^{N} (X_{n} - E(X_{n}))(X_{m} - E(X_{m}))\right) \\ &\stackrel{8.1.6(iii)}{=} \sum_{n=1}^{N} E((X_{n} - E(X_{n})^{2}) + 2\sum_{n=1}^{N} \sum_{m=n+1}^{N} E((X_{n} - E(X_{n}))(X_{m} - E(X_{m}))) \\ &\stackrel{8.1.6(iv)}{=} \sum_{n=1}^{N} \operatorname{var}(X_{n}) + 2\sum_{n=1}^{N} \sum_{m=n+1}^{N} \underbrace{E(X_{n} - E(X_{n}))}_{=0} \cdot \underbrace{E(X_{m} - E(X_{m}))}_{=0} \\ &= \sum_{n=1}^{N} \operatorname{var}(X_{n}). \end{aligned}$$

8.2.9 Beispiel (Binomialverteilung)

8.2.10 Bemerkung

Für die hypergeometrische Verteilung läßt sich die Varianz nicht so einfach berechnen, da dort die Summanden X_n (vgl. Beispiel 8.1.8(ii)) nicht unabhängig sind und somit die Formel von Bienaymé nicht angewendet werden kann.

8.2.11 Bemerkung (Abschließende Bemerkung zu den Maßzahlen)

Aus historischen Gründen haben Erwartungswerte von einigen speziellen Funktionen (insbesondere Monome und einfache Polynome) von Zufallsgrößen ihre eigenen Namen behalten:

 $E(X^k)$ heißt **k-tes Moment** von X, k = 1, 2, ...

 $E(|X|^k)$ heißt **k-tes absolutes Moment** von X, k = 1, 2...

 $E((X - E(X))^k)$ heißt **k-tes zentriertes Moment** von X, k = 1, 2, ...

sofern die Erwartungswerte existieren. Wie wir im Verlauf dieses Paragraphen gesehen haben, spielen das erste Moment (Erwartungswert) und das zweite zentrierte Moment (Varianz) eine wichtige Rolle. In Kapitel 9 werden wir sehen, dass die Existenz und Kenntnis dieser beiden Momente bereits genügt, um weiterreichende Resultate für das Grenzverhalten von Folgen und/ oder Partialsummen von Zufallsgrößen zu gewinnen.

8.2.12 Beispiel (Cauchy-Verteilung)

Die Cauchy-Verteilung ist eine Verteilung, bei der der Erwartungswert nicht existiert. Gilt $Y \sim C(\mu, \sigma)$, so haben wir

$$f_Y(x) = \frac{1}{\pi} \frac{\sigma}{\sigma^2 + (x - \mu)^2} = \frac{1}{\pi \sigma} \frac{1}{1 + (\frac{x - \mu}{\sigma})^2}$$

und $Y = \sigma X + \mu$ mit $X \sim C(0, 1)$. Es reicht wieder eine Cauchy-Verteilung mit Parametern 0 und 1 zu betrachten, d. h. $X \sim C(0, 1)$.

Wegen

$$\int_{-\infty}^{\infty} f_X(x) dx = \lim_{n \to \infty} \int_{-n}^{n} \frac{1}{\pi} \frac{1}{1 + x^2} dx = \lim_{n \to \infty} \frac{1}{\pi} [\arctan(x)]_{-n}^{n} = \frac{1}{\pi} \cdot \pi = 1$$

ist f_X und somit auch f_Y eine Dichte.

Wegen

$$\lim_{a \uparrow \infty} \int_{0}^{a} x \, f_X(x) \, dx = \lim_{a \uparrow \infty} \int_{0}^{a} \frac{1}{\pi} \, \frac{x}{1 + x^2} \, dx = \lim_{a \uparrow \infty} \left[\frac{1}{\pi} \cdot \frac{1}{2} \log(1 + x^2) \right]_{0}^{a}$$
$$= \lim_{a \uparrow \infty} \frac{1}{\pi} \frac{1}{2} \log(1 + a^2) = \infty$$

folgt

$$\int_{-\infty}^{\infty} |x| f_X(x) dx = 2 \lim_{a \uparrow \infty} \int_{0}^{\infty} |x| f_X(x) dx = \infty,$$

we shalb E(X) nicht existiert.

8.2.13 Beispiel (Blut-Test-Problem)

Eine große Zahl N von Personen soll auf eine Blutkrankheit untersucht werden. Jede habe mit Wahrscheinlichkeit p die Krankheit. Wird jede Person einzeln getestet, so sind N chemische Analysen notwendig.

Im 2. Weltkrieg wurde alternativ das Blut von k Personen auf einmal chemisch untersucht. Fiel der Test negativ aus, so wurden alle Personen noch einmal einzeln untersucht. Die Anzahl der chemischen Analysen pro Gruppe i ist also eine Zufallsgröße.

Erwartungswert und Varianz einiger Verteilungen

Diskrete Verteilungen						
Verteilung	Bezeichnung	Träger	Erwartungswert	Varianz		
Einpunktverteilung	$arepsilon_a$	$\{a\}$	a	0		
Diskrete Gleichverteilung	$G(x_1,\ldots,x_n)$	$\{x_1,\ldots,x_n\}$	\overline{x}	s_x^2		
Binomialverteilung	Bin(N,p)	$\{0,\ldots,N\}$	Np	Np(1-p)		
Bernoulli-Verteilung	Bin(1,p)	$\{0,\ldots,1\}$	p	p(1-p)		
Hypergeometrische Verteilung	Hyp(N,R,S)	$\{\max\{0, N - S\}, \dots, \min\{R, N\}\}$	$N\frac{R}{R+S}$	$N\frac{R}{R+S}\frac{S}{R+S}\frac{R+S-N}{R+S-1}$		
Geometrische Verteilung	Geo(p)	\mathbb{N}_0	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$		
Poisson-Verteilung	$Poi(\lambda)$	\mathbb{N}_0	λ	λ		

Stetige Verteilungen						
Rechteckverteilung	R[a,b]	[a,b]	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$		
Exponential-Verteilung	$Exp(\lambda)$	$[0,\infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$		
Weibull-Verteilung	$W(\alpha, \beta)$	$[0,\infty)$	$\beta \frac{1}{\alpha} \Gamma \left(\frac{1}{\alpha} \right)$	$\beta^2 \left(\frac{2}{\alpha} \Gamma \left(\frac{2}{\alpha} \right) - \left(\frac{1}{\alpha} \Gamma \left(\frac{1}{\alpha} \right) \right)^2 \right)$		
Gamma-Verteilung	$\Gamma(\alpha,\beta)$	$[0,\infty)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$		
χ^2 -Verteilung	χ_f^2	$[0,\infty)$	f	2f		
F-Verteilung	F_{f_1,f_2}	$[0,\infty)$	$\frac{f_2}{f_2 - 2}$	$\frac{2f_2^2(f_1+f_2-2)}{f_1(f_2-2)^2(f_2-4)}$		
			$(f_2 > 2)$	$(f_2 > 4)$		
Normalverteilung	$\mathcal{N}(\mu, \sigma^2)$	\mathbb{R}	μ	σ^2		
t-Verteilung	t_f	\mathbb{R}	0	$\frac{f}{f-2}$		
				(f > 2)		

8.3 Abhängigkeitsmaßzahlen

8.3.1 Motivation (Abhängigkeitsmaßzahlen)

Werden bei einem Zufallsexperiment zwei verschiedene Zufallsgrößen X und Y beobachtet, die. z.B. Größe und Gewicht eines zufällig ausgewählten Objektes beschreiben, so ist man nicht nur daran interessiert, ob X und Y unabhängig sind, sondern auch, ob spezielle Abhängigkeitsverhältnisse auftreten, z. B. ob große (bzw. kleine) Werte x der Zufallsgröße X eher mit großen (bzw. kleinen) Werten y von Y auftreten oder umgekehrt. Groß bzw. klein muss dabei natürlich immer im Verhältnis zu einer Referenzgröße, einer Lagemaßzahl, gesehen werden. Folgende Definition liefert eine geeignete Definition, die zu einer Beschreibung der Abhängigkeitsstruktur innerhalb eines ganzen Zufallsvektors (X_1, \ldots, X_N) verallgemeinert werden kann:

8.3.2 Definition (Kovarianz)

(i) Sind X, Y Zufallsgrößen mit existierendem Erwartungswert, so heißt

$$cov(X, Y) := E((X - E(X))(Y - E(Y)))$$

Kovarianz von X und Y (sofern der Erwartungswert existiert).

(ii) Ist $X = (X_1, \dots, X_N)^{\top}$ ein N-dimensionaler Zufallsvektor, dessen Komponenten X_1, \dots, X_N die Varianzen $\textit{var}(X_1), \dots, \textit{var}(X_N)$ besitzen, dann heißt

$$extstyle extstyle ext$$

Kovarianzmatrix von X.

8.3.3 Bemerkung

Im Satz 8.3.8 (Cauchy-Schwarz'sche Ungleichung) werden wir sehen, dass die Existenz von $var(X_i)$, $var(X_j)$ die Existenz von $cov(X_i, X_j)$ nach sich zieht.

8.3.4 Satz (Eigenschaften der Kovarianz)

Für Zufallsgrößen X, Y und Konstanten $a, b, c, d \in \mathbb{R}$ gilt:

- (i) cov(X, X) = var(X),
- (ii) cov(X, Y) = cov(Y, X),
- (iii) $cov(X, Y) = E(X \cdot Y) E(X) \cdot E(Y),$
- (iv) $cov(aX + b, cY + d) = a \cdot c \cdot cov(X, Y)$.

Sind X und Y stochastisch unabhängig, so gilt insbesondere

(v)
$$cov(X,Y) = 0$$
.

Beweis

- (i) und (ii) folgen sofort aus der Definition der Kovarianz und der Definition der Varianz (Definition 8.2.3).
- (iii) Wie im Beweis vom Satz 8.2.5(i), wobei wieder unter der Voraussetzung der Existenz von E(X) und E(Y) die Existenz von cov(X,Y) die von $E(X \cdot Y)$ nach sich zieht und umgekehrt, gilt:

$$\begin{aligned} \operatorname{cov}(X,Y) &= E((X-E(X))(Y-E(Y)) \\ &= E(X\cdot Y - X\cdot E(Y) - Y\cdot E(X) + E(X)\cdot E(Y)) \\ &= E(X\cdot Y) - E(X)\cdot E(Y) - E(Y)\cdot E(X) + E(X)\cdot E(Y) \\ &= E(X\cdot Y) - E(X)\cdot E(Y). \end{aligned}$$

$$\begin{split} \operatorname{cov}(aX+b,cY+d) \\ &= E((aX+b-E(aX+b))\cdot(cY+d-E(cY+d))) \\ &= E(a(X-E(X))\cdot c(Y-E(Y))) \\ &= a\cdot c\cdot E((X-E(X))\cdot (Y-E(Y))) \\ &= a\cdot c\cdot \operatorname{cov}(X,Y). \end{split}$$

(v) Folgt aus (iii) und Satz 8.1.6(iv) mit $g = h = \text{Identitat d. h. } g(x) = h(x) = x \text{ für alle } x \in \mathbb{R}.$

8.3.5 Bemerkung

Die Umkehrung von (v) in Satz 8.3.4 gilt im Allgemeinen nicht. D. h. aus cov(X, Y) = 0 folgt im Allgemeinen **nicht** die stochastische Unabhängigkeit von X und Y.

8.3.6 Satz (Eigenschaften der Kovarianzmatrix)

Für einen N-dimensionalen Zufallsvektor $X = (X_1, \ldots, X_N)^{\top}$, Matrix $A \in \mathbb{R}^{M \times N}$, $b \in \mathbb{R}^M$, $a = (a_1, \ldots, a_N)^T \in \mathbb{R}^N$ gilt:

- (i) $Cov(AX + b) = ACov(X)A^{T}$, (wobei A^{T} die Transposition der Matrix A ist),
- (ii) $var(a^T X) = a^T Cov(X) a, d.h.$ $var\left(\sum_{n=1}^N a_n X_n\right) = \sum_{n=1}^N a_n^2 var(X_n) + 2\sum_{n=1}^N \sum_{m=n+1}^N a_n a_m cov(X_n, X_m),$
- (iii) Cov(X) ist positiv semi-definit,
- (iv) Cov(X) singulär

 \Leftrightarrow

Es existiert $a \in \mathbb{R}^N$, $a \neq 0$, $b \in \mathbb{R}$ mit $a^T X = b$ P-fast sicher, d. h. $P(a^T X = b) = 1$.

Beweis

(i) Sei
$$A=\begin{pmatrix}A_1\\\vdots\\A_M\end{pmatrix}$$
 mit $A_i=(A_{i1},\dots,A_{iN}), i=1,\dots,M,$ und $b=\begin{pmatrix}b_1\\\vdots\\b_M\end{pmatrix}.$ Für die

(i, j)'te Komponente von Cov(AX + b) gilt dann

$$\begin{split} & \mathsf{Cov}(AX + b)_{ij} = \mathsf{cov}(A_iX + b_i, \, A_jX + b_j) \\ & \stackrel{8.3.4(iv)}{=} \mathsf{cov}(A_iX, \, A_jX) = \mathsf{cov}\left(\sum_{n=1}^N A_{in}X_n, \, \sum_{m=1}^N A_{jm}X_m\right) \\ & = E\left(\left(\sum_{n=1}^N A_{in}X_n - E\left(\sum_{n=1}^N A_{im}X_n\right)\right) \cdot \left(\sum_{m=1}^N A_{jm}X_m - E\left(\sum_{m=1}^N A_{jm}X_m\right)\right)\right) \\ & = E\left(\left(\sum_{n=1}^N A_{in}(X_n - E(X_n))\right) \cdot \left(\sum_{m=1}^N A_{jm}(X_m - E(X_m))\right)\right) \\ & = E\left(\sum_{n=1}^N \sum_{m=1}^N A_{in}A_{jm}(X_n - E(X_n))(X_m - E(X_m))\right) \\ & = \sum_{n=1}^N \sum_{m=1}^N A_{in}A_{jm}E((X_n - E(X_n))(X_m - E(X_m))) \\ & = \sum_{n=1}^N \sum_{m=1}^N A_{in}A_{jm}\mathsf{cov}(X_n, \, X_m) \\ & = A_i \, \mathsf{Cov}(X) \, A_j^T = (A \, \mathsf{Cov}(X) \, A^T)_{ij} \end{split}$$

für alle $i, j = 1, \ldots, M$.

- (ii) Folgt aus (i) mit M=1 und $a^{\top}X:\Omega\to\mathbb{R}$ unter Ausnutzung von Satz 8.3.4(i)): $\mathsf{var}(a^TX)=\mathsf{cov}(a^TX,a^TX)=\mathsf{Cov}(a^TX)=a^T\mathsf{Cov}(X)a.$
- (iii) Da $\operatorname{var}(a^T X) \geq 0$ nach Definition der Varianz gilt, folgt aus (ii) $a^T \operatorname{Cov}(X) a = \operatorname{var}(a^T X) \geq 0$ für alle $a \in \mathbb{R}^N$, was die Definition von "positiv semi-definit" ist.

#

(iv) Cov(X) singulär

$$\begin{split} &\Leftrightarrow \bigvee_{a \in \mathbb{R}^N: a \neq 0} a^T \mathsf{Cov}(X) a = 0 \\ &\Leftrightarrow \bigvee_{a \in R^N: a \neq 0} \mathsf{var}(a^T X) = 0 \\ &\Leftrightarrow \bigvee_{a \in \mathbb{R}^N: a \neq 0} P(a^T X = a^T E(X) \text{ konstant }) = 1. \end{split}$$

Dabei gilt in (*) die Richtung " \Leftarrow " wegen Beispiel 8.1.5(i) und Beispiel 8.2.7(i). Die Richtung " \Rightarrow " in (*) gilt aus folgendem Grund:

$$\begin{aligned} 0 &= \mathrm{var}(Y) = E((Y - E(Y)^2) = \int (y - E(Y))^2 \, P^Y(dy) \\ &= \begin{cases} \sum_y (y - E(Y))^2 p_Y(y) \\ \sum_y (y - E(Y))^2 \, f_Y(y) \, dy \end{cases} \\ &= \begin{cases} \sum_{y: (y - E(Y))^2 > 0} p_Y(y) \, dy \\ \sum_{y: (y - E(Y))^2 > 0} (y) f_Y(y) \, dy = 0 \end{cases} \\ &\Rightarrow P^Y(\{y; \ (y - E(Y))^2 > 0\}) = 0 \\ &\Rightarrow P^Y(\{y; \ (y - E(Y))^2 = 0\}) = 1 \\ &\Rightarrow P((Y - E(X))^2 = 0) = 1 \\ &\Rightarrow P(Y = E(Y)) = 1 \end{aligned}$$

8.3.7 Bemerkung

- (i) Die Aussage (ii) von Satz 8.3.6 verallgemeinert die Formel von Bienaymé (Satz 8.2.8).
- (ii) Sind die Komponenten X_1, \ldots, X_N eines Zufallsvektors X stochastisch unabhängig, so ist $\mathsf{Cov}(X)$ wegen Satz 8.3.4(v) eine Diagonalmatrix. Sind X_1, \ldots, X_N nicht entartet, d. h. für alle $n = 1, \ldots, N$ gibt es kein $c \in \mathbb{R}$ mit $P(X_n = c) = 1$, dann gilt $\mathsf{var}(X_n) > 0$ für alle $n = 1, \ldots, N$ und $\mathsf{Cov}(X)$ ist nicht singulär. Nach Satz 8.3.6(iv) bedeutet dies, dass für alle $0 \neq a \in \mathbb{R}^N$ und alle $b \in \mathbb{R}$ gilt

$$P(a^T X = b) < 1,$$

so dass fast sicher **nicht** $a^T X = b$ gilt.

Das bedeutet aber, dass die Funktionen $1, X_1, \ldots, X_N : \Omega \to \mathbb{R}$ im "Vektorsinn" linear unabhängig sind. Die stochastische Unabhängigkeit von $1, X_1, \ldots, X_N$ impliziert also die lineare Unabhängigkeit von $1, X_1, \ldots, X_N$ und damit auch von X_1, \ldots, X_N .

Ħ

Die Umkehrung gilt aber nicht. Denn sind $1, X_1, \ldots, X_N$ linear unabhängig, dann kann aber $A \in \mathcal{A}, a \in \mathbb{R}^N, b \in \mathbb{R}$ existieren mit

$$a^T X(\omega) = b$$
 für alle $\omega \in A$.

Gilt dann noch P(A) = 1, dann ist nach Satz 8.3.6(iv) Cov(X) singular und somit nicht von Diagonal-Gestalt (falls X_1, \ldots, X_N nicht entartet sind, was angenommen wird). Damit sind X_1, \ldots, X_N nicht stochastisch unabhängig.

Wir sehen also den Unterschied zwischen stochastischer und linearer Unabhängigkeit.

8.3.8 Satz (Cauchy-Schwarz'sche Ungleichung)

Sind X, Y Zufallsgrößen mit existierenden Varianzen var(X), var(Y), dann existiert cov(X, Y) und es gilt

$$|cov(X,Y)| \le \sqrt{var(X)} \cdot \sqrt{var(Y)}.$$

Beweis

(i) Zur Existenz von cov(X, Y):

$$|(X - E(X))(Y - E(Y))| \le (X - E(X))^2 + (Y - E(Y))^2$$

Folglich

$$E(|(X - E(X))(Y - E(Y))|) \le \operatorname{var}(X) + \operatorname{var}(Y) < \infty.$$

und cov(X, Y) existiert.

(ii) Die Kovarianz-Matrix $\mathsf{Cov}((X,Y)^\top)$ des Zufallsvektors $(X,Y)^\top$ hat die Gestalt

$$\begin{pmatrix} \operatorname{var}(X) & \operatorname{cov}(X,Y) \\ \operatorname{cov}(X,Y) & \operatorname{var}(Y) \end{pmatrix} = \operatorname{Cov}\left(\begin{pmatrix} X \\ Y \end{pmatrix}\right).$$

Nach Satz 8.3.6(iii) ist sie positiv semi-definit, d.h. alle Eigenwerte sind nicht-negativ und insbesondere gilt für die Determinante

$$0 \leq \ \det \ \left(\mathsf{Cov} \left(\binom{X}{Y} \right) \right) = \mathsf{var}(X) \cdot \mathsf{var}(Y) - \mathsf{cov}(X,Y)^2.$$

Daraus folgt die Behauptung.

8.3.9 Bemerkung

Wegen $cov(aX + b, cY + d) = a \cdot c cov(X, Y)$ (Satz 8.3.4(iv)) bleibt die Abhängigkeitsmaßzahl Kovarianz unverändert unter Lokationstransformation, variiert aber mit Skalentransformation. Um diesen Effekt auszuschließen, geht man zu normierten Zufallsgrößen

$$\frac{1}{\sqrt{Var(X)}}(X - E(X))$$
 bzw. $\frac{1}{\sqrt{\mathsf{var}(Y)}}(Y - E(Y))$

über und berechnet für diese die Abhängigkeitsmaßzahl. "Normiert" bedeutet hier, dass die neuen Zufallsgrößen Erwartungswert 0 und Varianz 1 besitzen, denn

$$E\left(\frac{1}{\sqrt{\mathsf{var}(X)}}(X - E(X))\right) = \frac{1}{\sqrt{\mathsf{var}(X)}}\left(E(X) - E(X)\right) = 0$$

und

$$\operatorname{var}\left(\frac{1}{\sqrt{\operatorname{var}(X)}}(X-E(X))\right) = \frac{1}{\operatorname{var}(X)}\operatorname{var}(X-E(X)) = \frac{1}{\operatorname{var}(X)}\operatorname{var}(X) = 1.$$

8.3.10 Definition (Korrelationskoeffizient)

(i) Sind X, Y Zufallsgrößen mit existierenden Varianzen var(X), var(Y) > 0, dann heißt

$$\rho(X,Y) := \textit{cov}\left(\frac{1}{\sqrt{\textit{var}(X)}}(X - E(X)), \ \frac{1}{\sqrt{\textit{var}(Y)}}(Y - E(Y))\right)$$

Korrelationskoeffizient von X und Y.

- (ii) Gilt $\rho(X,Y) = 0$, dann heißen X und Y unkorreliert.
- (iii) Gilt $\rho(X,Y) > 0$, dann heißen X und Y positiv korreliert.
- (iv) Gilt $\rho(X,Y) < 0$, dann heißen X und Y negativ korreliert.

8.3.11 Satz (Eigenschaften des Korrelationskoeffizienten)

(i)
$$\rho(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}}$$

- (ii) $-1 < \rho(X, Y) < 1$
- (iii) X, Y unabhängig $\Rightarrow X, Y$ unkorreliert
- iv) $|\rho(X,Y)| = 1 \Leftrightarrow 1, X, Y$ linear abhängig P-fast sicher, d. h. es gibt $0 \neq a \in \mathbb{R}^2$, $b \in \mathbb{R}$ mit $P\left(a^T\binom{X}{Y} = b\right) = 1$.

Beweis

- (i) Folgt aus der Definition 8.3.10 und Satz 8.3.4 (iv).
- (ii) Folgt mit (i) aus der Cauchy-Schwarz'schen Ungleichung (Satz 8.3.8).
- (iii) Folgt mit (i) aus Satz 8.3.4 (v).
- (iv) Nach dem Beweis der Cauchy-Schwar'zschen Ungleichung (Satz 8.3.8) gilt $|\rho(X,Y)| = 1 \Leftrightarrow \det \left(\mathsf{Cov} \left(\binom{X}{Y} \right) \right) = 0$ $\Leftrightarrow \mathsf{Cov} \left(\binom{X}{Y} \right)$ ist singulär

$$\overset{8.3.6\text{(iv)}}{\Leftrightarrow} \text{ Es existiert } 0 \neq a \in \mathbb{R}^2, \, b \in \mathbb{R} \text{ mit } P\left(a^T\binom{X}{Y} = b\right) = 1.$$

8.3.12 Beispiel

Seien X und Y Zufallsvariablen mit folgender Zähldichte:

$$x_{i} \begin{array}{c|c|c} y_{j} \\ \hline P(X = x_{i}, Y = y_{j}) & 0 & 1 & 2 & P(X = x_{i}) \\ \hline -1 & \frac{1}{10} & \frac{1}{5} & \frac{1}{10} & \frac{2}{5} \\ \hline 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{10} & \frac{3}{5} \\ \hline P(Y = y_{j}) & \frac{7}{20} & \frac{9}{20} & \frac{1}{5} & 1 \\ \hline \end{array}$$

8.3.13 Beispiel (Normalverteilung)

Sind X_1, \ldots, X_N unabhängige, standardnormal-verteilte Zufallsgrößen, so besitzt nach Beispiel 5.3.11 $X = (X_1, \ldots, X_N)$ eine N-dimensionale Normalverteilung mit Parametern

$$\mu_{\circ} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \text{ und } \Sigma_0 = \begin{pmatrix} 1 & 0 \\ \ddots & 0 \\ 0 & 1 \end{pmatrix}.$$

Wegen der Unabhängigkeit und $var(X_n) = 1, n = 1, ..., N$, gilt $Cov(X) = \Sigma_0$.

8.3.14 Lemma

Sei $(X_1, X_2)^{\top} \sim \mathcal{N}_2(\mu, \Sigma)$, d.h. $(X_1, X_2)^{\top}$ hat eine zweidimenionale Normalverteilung mit Erwartungsvektor μ und Kovarianzmatrix Σ . Dann gilt

$$X_1, X_2$$
 sind unabhängig $\iff \rho(X_1, X_2) = 0.$

8.3.15 Beispiel (Hypergeometrische Verteilung)

Wird N-mal aus einer Urne mit R roten und S schwarzen Kugeln gezogen, so sei

 $X_n = 1 =$ in der *n*-ten Ziehung wurde eine rote Kugel gezogen,

 $X_n = 0$ in der *n*-ten Ziehung wurde eine schwarze Kugel gezogen,

 $n=1,\ldots,N.$

Man zeige als Übung:

$$\begin{split} & \operatorname{cov}(X_n, X_{n-1}) = -\frac{R \cdot S}{(R+S)^2 (R+S-1)} < 0, \\ & \rho(X_n, X_{n-1}) = -\frac{1}{R+S-1}, \\ & \operatorname{var}\left(\sum_{n=1}^N X_n\right) = \frac{R+S-N}{R+S-1} \; N \; \frac{R}{R+S} \; \frac{S}{R+S} \; \; . \end{split}$$

9 Konvergenz von Zufallsvariablen

9.1 Das schwache Gesetz der großen Zahlen

9.1.1 Definition (Stochastische Unabhängigkeit von Folgen) Sei (Ω, \mathcal{A}, P) ein Wahrscheinlickeitsraum.

(i) Eine Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen $A_n\in\mathscr{A},\ n\in\mathbb{N}$, heißt (stochastisch) unabhängig.

 \Leftrightarrow

 $\bigwedge_{N\in\mathbb{N}} A_1,\ldots,A_N$ unabhängig (vgl. Definition 5.2.3).

(ii) Eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen X_n auf (Ω, \mathscr{A}, P) , $n\in\mathbb{N}$, heißt (stochastisch) unabhängig

 \Leftrightarrow

 $\bigwedge_{N\in\mathbb{N}} X_1, \dots, X_N \text{ unabhängig (vgl. Definition 5.3.2)}.$

9.1.2 Folgerung

(i) $(A_n)_{n\in\mathbb{N}}$ unabhängig

$$\Leftrightarrow \bigwedge_{\substack{1 < N \in \mathbb{N} \\ i_1 < i_2 < \ldots < i_N}} A_{i_1}, \ldots, A_{i_N}$$
 unabhängig.

(ii) $(X_n)_{n\in\mathbb{N}}$ unabhängig

$$\Leftrightarrow \bigwedge_{\substack{1 < N \in \mathbb{N} \\ i_1 < i_2 < \dots < i_N}} X_{i_1}, \dots, X_{i_N}$$
 unabhängig.

Beweis

"⇐" trivial

" \Rightarrow " folgt für (i) aus Definition 5.2.3 und für (ii) aus Definition 5.3.2.

9.1.3 Bemerkung

- (i) Folgerung 9.1.2 besagt, dass eine Folge von Ereignissen (bzw. Zufallsvariablen) genau dann unabhängig ist, wenn jede endliche Teilmenge dieser Ereignisse (bzw. Zufallsvariablen) unabhängig ist.
- (ii) Die Definition 9.1.1(i) und (ii) sind wieder in dem Sinne verträglich, dass für Indikatorvariablen $X_n = 1_{A_n}$ gilt:

 $(1_{A_n})_{n\in\mathbb{N}}$ unabhängig

 $\Leftrightarrow (A_n)_{n \in \mathbb{N}}$ unabhängig.

(vgl. Beispiel 5.3.4).

- (iii) Die Existenz von einer Folge $(X_n)_{n\in\mathbb{N}}$ von unabhängigen Zufallsvariablen (und damit auch von einer Folge von unabhängigen Ereignissen) wird in der Maßtheorie gezeigt.
- **9.1.4 Definition** (Stochastische und fast sichere Konvergenz)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen auf (Ω, \mathcal{A}, P) und X Zufallsgröße auf (Ω, \mathcal{A}, P) .

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert fast sicher gegen X (kurz $X_n \stackrel{f.s.}{\longrightarrow} X$)

 \Leftrightarrow

$$P(\{\omega \in \Omega; \lim_{n \to \infty} X_n(\omega) = X(\omega)\}) = 1.$$

(ii) $(X_n)_{n\in\mathbb{N}}$ konvergiert stochastisch / in Wahrscheinlichkeit gegen X (kurz $X_n \xrightarrow{\mathcal{P}} X$)

 \Leftrightarrow

$$\bigwedge_{\varepsilon>0} \lim_{n\to\infty} P(\{\omega \in \Omega; |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0.$$

9.1.5 Satz

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert fast sicher gegen X

 \Rightarrow

 $(X_n)_{n\in\mathbb{N}}$ konvergiert stochastisch gegen X.

(ii) Die Umkehrung von (i) gilt im Allgemeinen nicht.

Beweis

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert fast sicher gegen X.

 \rightarrow

$$\bigwedge_{\varepsilon>0} \lim_{n\to\infty} P\left(\left\{\omega; |X_n(\omega) - X(\omega)| > \varepsilon\right\}\right)$$

$$\leq \lim_{n \to \infty} P\left(\bigcup_{k \geq n} \{\omega; |X_k(\omega) - X(\omega)| > \varepsilon\}\right)$$

$$\stackrel{(\sigma\text{-Stetigkeit}}{=} \underbrace{\left(\text{Satz 3.1.14}\right)}_{} P \left(\bigcap_{n \geq 1} \bigcup_{k \geq n} \{\omega; |X_k(\omega) - X(\omega)| > \varepsilon \} \right)$$

9.1 Das schwache Gesetz der großen Zahlen

$$= P\left(\left\{\omega; \bigwedge_{n\geq 1} \bigvee_{k\geq n} |X_k(\omega) - X(\omega)| > \varepsilon\right\}\right)$$

$$= P\left(\left\{\omega; \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| \geq \varepsilon\right\}\right)$$

$$= 1 - P\left(\left\{\omega; \limsup_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon\right\}\right)$$

$$\leq 1 - P\left(\left\{\omega; \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\right\}\right)$$

$$= 1 - P\left(\left\{\omega; \lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0\right\}\right)$$

$$= 0.$$

 \Rightarrow (X_n) konvergiert stochastisch gegen X.

(ii) Wir müssen nur ein Beispiel von $(X_n)_{n\in\mathbb{N}}$ geben, die stochastisch aber nicht fast sicher konvergieren. Sei dazu $(\Omega, \mathscr{A}, P) = ([0,1], \mathscr{B}_{[0,1]}, U(0,1))$ $(\mathscr{B}_{[0,1]} := \{B \cap [0,1]; B \in \mathscr{B}\}$ heißt auch Spur- σ -Algebra von \mathscr{B}) und

$$X_{1} = 1_{[0,1/2)}, X_{2} = 1_{[1/2,1)}$$

$$X_{3} = 1_{[0,1/4)}, X_{4} = 1_{[1/4,1/2)}, X_{5} = 1_{[1/2,3/4)}, X_{6} = 1_{[3/4,1)}$$

$$\vdots$$

$$X_{n} = 1_{\left[\frac{k}{2m}, \frac{k+1}{2m}\right)}$$

falls $n = 2^m + k - 1$ mit $k \in \{0, \dots, 2^m - 1\}$ und $2^m \le n + 1 < 2^{m+1}$.

Sei $\omega \in [0,1)$ beliebig. Für alle $N \in \mathbb{N}$ gibt es dann ein $n \geq N$, $m \in \mathbb{N}$, $k \in \{0,\ldots,2^m-2\}$ mit $2^m \leq n+1 < 2^{m+1}$, $n=2^m+k-1$ und $\omega \in \left[\frac{k}{2^m},\,\frac{k+1}{2^m}\right)$, $\omega \notin \left[\frac{k+1}{2^m},\,\frac{k+2}{2^m}\right)$,

d. h.

$$X_n(\omega) = 1, \ X_{n+1}(\omega) = 0.$$

Das bedeutet

$$\limsup_{n \to \infty} X_n(\omega) = 1, \quad \liminf_{n \to \infty} X_n(\omega) = 0$$

für alle $\omega \in [0,1)$ und somit

$$P\left(\left\{\omega; \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}\right) = P(\emptyset) = 0$$

Ħ

für alle Zufallsgrößen X. D. h. $(X_n)_{n\in\mathbb{N}}$ konvergiert nicht fast sicher. Andererseits gilt für $X(\omega)=0$ für alle $\omega\in[0,1]$:

$$\bigwedge_{\varepsilon>0, n\in\mathbb{N}} P(\{\omega; |X_n(\omega) - X(\omega)| > \varepsilon\}) = P(\{\omega; |X_n(\omega)| > \varepsilon\})$$

$$= P(\{\omega; X_n(\omega) = 1\}) = P\left(\left[\frac{k}{2^m}, \frac{k+1}{2^m}\right)\right) = \frac{1}{2^m} = \frac{2}{2^{m+1}} < \frac{2}{n+1},$$

$$\text{da } 2^m \le n + 1 < 2^{m+1}. \text{ Es folgt}$$

$$\bigwedge_{\varepsilon>0} \lim_{n\to\infty} P(\{\omega; |X_n(\omega) - X(\omega)| > \varepsilon\}) \le \lim_{n\to\infty} \frac{2}{n+1} = 0,$$

und somit die stochastische Konvergenz von $(X_n)_{n\in\mathbb{N}}$.

9.1.6 Satz

- (i) Konvergiert $(X_n)_{n\in\mathbb{N}}$ stochastisch gegen X und $(Y_n)_{n\in\mathbb{N}}$ stochastisch gegen Y, so gilt
 - (a) $(X_n + Y_n)_{n \in \mathbb{N}}$ konvergiert stochastisch gegen X + Y,
 - (b) $(X_n \cdot Y_n)_{n \in \mathbb{N}}$ konvergiert stochastisch gegen $X \cdot Y$.
- (ii) Konvergiert $(X_n)_{n\in\mathbb{N}}$ fast sicher gegen X und $(Y_n)_{n\in\mathbb{N}}$ fast sicher gegen Y, so gilt
 - (a) $(X_n + Y_n)_{n \in \mathbb{N}}$ konvergiert fast sicher gegen X + Y,
 - (b) $(X_n \cdot Y_n)_{n \in \mathbb{N}}$ konvergiert fast sicher gegen $X \cdot Y$.

Beweis

Es wird nur die Behauptung (i) b) gezeigt. Der Rest ist Übungsaufgabe.

$$P(|X_{n} \cdot Y_{n} - X \cdot Y| > \varepsilon)$$

$$\leq P(|X_{n} \cdot Y_{n} - X \cdot Y| > \varepsilon, |X| \leq m, |Y| \leq m) + P(|X| > m) + P(|Y| > m)$$

$$= P(|(X_{n} - X) (Y_{n} - Y) + X (Y_{n} - Y) + (X_{n} - X) Y| > \varepsilon, |X| \leq m, |Y| \leq m)$$

$$+ P(|X| > m) + P(|Y| > m)$$

$$\leq P(|X_{n} - X| |Y_{n} - Y| + |X| |Y_{n} - Y| + |X_{n} - X| |Y| > \varepsilon, |X| \leq m, |Y| \leq m)$$

$$+ P(|X| > m) + P(|Y| > m)$$

$$\leq P(|X_n - X| |Y_n - Y|) > \varepsilon/3 \lor |X| |Y_n - Y| > \varepsilon/3 \lor |X_n - X| |Y| > \varepsilon/3,$$

 $|X| \leq m, |Y| \leq m)$

$$+ P(|X| > m) + P(|Y| > m)$$

$$\leq P\left(|X_n - X| > \sqrt{\varepsilon/3}\right) + P\left(|Y_n - Y| > \sqrt{\varepsilon/3}\right) + P\left(|Y_n - Y| > \frac{\varepsilon}{3m}\right) + P\left(|X_n - X| > \frac{\varepsilon}{3m}\right) + P(|X| > m) + P(|Y| > m)$$

(*m* entsprechend gewählt)

$$\leq P\left(|X_n - X| > \sqrt{\varepsilon/3}\right) + P\left(|Y_n - Y| > \sqrt{\varepsilon/3}\right) + P\left(|Y_n - Y| > \frac{\varepsilon}{3m}\right) + P\left(|X_n - X| > \frac{\varepsilon}{3m}\right) + \delta/6 + \delta/6$$

(*n* entsprechend gewählt)

$$\leq 6 \cdot \delta/6 = \delta.$$

Um das schwache Gesetz der großen Zahlen zeigen zu können, wird die Tschebyscheff-Ungleichung benötigt, die aus der Markoff-Ungleichung folgt.

9.1.7 Satz (Markoff-Ungleichung)

Sei Z eine Zufallsgröße auf (Ω, \mathscr{A}, P) und $g : \mathbb{R}^+ \to \mathbb{R}$ eine nichtnegative monoton wachsende Funktion, so dass E(g(|Z|)) existiert. Dann gilt für alle $\varepsilon > 0$:

$$P(|Z| \ge \varepsilon) \le \frac{1}{g(\varepsilon)} E(g(|Z|)).$$

Beweis

Setze $Y(\omega) = 0$ für $g(|Z(\omega)|) < g(\varepsilon)$ und $Y(\omega) = g(\varepsilon)$ für $g(|Z(\omega)|) \ge g(\varepsilon)$. Dann gilt $Y(\omega) \le g(|Z(\omega)|)$ für alle $\omega \in \Omega$ und aus der Monotonie von g folgt:

$$\begin{array}{lclcl} E(g(|Z|)) & \geq & E(Y) & = & g(\varepsilon) \cdot P(Y = g(\varepsilon)) & = & g(\varepsilon) \cdot P(g(|Z|) \geq g(\varepsilon)) \\ = & g(\varepsilon) \cdot P(|Z| \geq \varepsilon). \end{array}$$

9.1.8 Satz (Tschebyscheff-Ungleichung)

Sei X eine Zufallsgröße auf (Ω, \mathcal{A}, P) für die var(X) existiert. Dann gilt für alle $\varepsilon > 0$:

$$P(|X - E(X)| \ge \varepsilon) \le \frac{1}{\varepsilon^2} \operatorname{var}(X).$$

Beweis

Die Aussage folgt aus Satz 9.1.7, indem man Z = X - E(X) und $g(|Z|) = Z^2$ setzt.

9.1.9 Satz (Schwaches Gesetzt der großen Zahlen)

Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen mit gleichem Erwartungswert und $\operatorname{var}(X_n) \leq c < \infty, n \in \mathbb{N}$, dann gilt für alle $\varepsilon > 0$:

$$\lim_{N \to \infty} P\left(\left| \frac{1}{N} \sum_{n=1}^{N} X_n - E(X_1) \right| > \varepsilon \right) = 0,$$

d.h. $\frac{1}{N} \sum_{n=1}^{N} X_n$ konvergiert stochastisch gegen $E(X_1)$.

Beweis

Sei
$$Y_N = \frac{1}{N} \sum_{n=1}^N X_n$$
. Dann gilt $E(Y_N) = E(X_1)$ und $\operatorname{var}(Y_N) = \frac{1}{N^2} \operatorname{var} \left(\sum_{n=1}^N X_n \right)$

Bienaymé (Satz 8.2.8) $\frac{1}{N^2} \sum_{N=1}^N \operatorname{var}(X_n)$
 $\leq \frac{1}{N^2} N c = \frac{c}{N}$.

Mit der Tschebyscheff-Ungleichung (Satz 9.1.8) folgt für alle $\varepsilon > 0$:

$$\lim_{N \to \infty} P\left(\left|\frac{1}{N} \sum_{n=1}^{N} X_n - E(X_1)\right| > \varepsilon\right) = \lim_{N \to \infty} P(|Y_N - E(Y_N)| > \varepsilon)$$

$$\leq \lim_{N \to \infty} \frac{1}{\varepsilon^2} \text{var}(Y_N) = \lim_{N \to \infty} \frac{c}{\varepsilon^2 N} = 0.$$

9.1.10 Bemerkung

1. Aus dem Beweis des schwachen Gesetzes der großen Zahl ist ersichtlich, dass die Voraussetzung der Unabhängigkeit der Zufallsgrößen X_n abgeschwächt werden kann. Es reicht aus, nur die paarweise Unkorreliertheit der X_n zu fordern. Denn dann gilt weiterhin mit Satz 8.3.6(ii)

$$\operatorname{var}\left(\sum_{n=1}^{N} X_n\right) = \sum_{n=1}^{N} \operatorname{var}(X_n).$$

2. Die Gleichheit der Erwartungswerte kann auch abgeschwächt werden: Sind die Erwartungswerte verschieden, dann konvergiert

$$\frac{1}{N}\sum_{n=1}^{N}(X_n - E(X_n))$$

stochastisch gegen 0.

9.1.11 Folgerung

Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge paarweise unkorrelierter Zufallsgrößen mit $\operatorname{var}(X_n) \leq c < \infty, n \in \mathbb{N}$, dann konvergiert

$$\frac{1}{N}\sum_{n=1}^{N}(X_n - E(X_n))$$

stochastisch gegen 0.

9.1.12 Beispiel

Eine Maschine produziert Schrauben, die im Mittel 50mm lang sind. Die Variabilität um den Mittelwert beträgt 1mm.

a) Was ist die maximale Wahrscheinlichkeit, dass eine Schraube um mehr als 2mm von

der vorgegebenen Länge von 50mm abweicht?

b) Wie hoch ist maximal der langfristige Anteil von Schrauben, die um mehr als 2mm von der vorgegebenen Länge abweichen?

9.2 Das starke Gesetz der großen Zahlen

9.2.1 Satz (Borel-Cantelli-Lemma)

Für eine Folge $(A_n)_{n\in\mathbb{N}}$ von Ereignissen sei $A^* = \limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$.

(i) Ist
$$\sum_{k=1}^{\infty} P(A_k) < \infty$$
, so ist $P(A^*) = 0$.

(ii) Sind
$$(A_n)_{n\in\mathbb{N}}$$
 unabhängig und $\sum_{k=1}^{\infty} P(A_k) = \infty$, so ist $P(A^*) = 1$.

Beweis

(i) Wegen $\bigcap_{n\geq 1}\bigcup_{k\geq n}A_k\subseteq\bigcup_{k\geq N}A_k$ für alle $N\in\mathbb{N}$ gilt

$$P(A^*) \le P\left(\bigcup_{k \ge N} A_k\right) \le \sum_{k \ge N} P(A_k) =: c_N$$

und $\lim_{N\to\infty} c_N = 0$ wegen der Konvergenz der Summe.

(ii) $P(A^*) = 1 - P\left(\bigcap_{n \ge 1} \bigcup_{k \ge n} A_k\right) = 1 - P\left(\bigcup_{n \ge 1} \bigcap_{k \ge n} \overline{A}_k\right)$ $\ge 1 - \sum_{n \ge 1} P\left(\bigcap_{k \ge n} \overline{A}_k\right) = 1 - \sum_{n \ge 1} P\left(\lim_{N \to \infty} \bigcap_{k = n}^N \overline{A}_k\right)$ a Stetick

$$\stackrel{\sigma\mathrm{-Stetigk.}}{=} 1 - \sum_{n=1}^{\infty} \lim_{N \to \infty} P\left(\bigcap_{k=n}^{N} \overline{A}_{k}\right)$$

Unabhängigk.
$$1 - \sum_{n=1}^{\infty} \lim_{N \to \infty} \prod_{k=n}^{N} (1 - P(A_k)).$$

Wegen $1 - x \le e^{-x}$ für $x \ge 0$ und wegen $P(A_k) \ge 0$ folgt:

$$P(A^*) \ge 1 - \sum_{n=1}^{\infty} \lim_{N \to \infty} e^{-\sum_{k=n}^{N} P(A_k)} \stackrel{(*)}{=} 1 - \sum_{n=1}^{\infty} 0 = 1.$$

Dabei gilt (*) wegen
$$\sum_{k=n}^{\infty} P(A_k) = \infty$$
.

9.2.2 Satz (Charakterisierung der Fast-sicheren-Konvergenz)

 X_n konvergiert fast sicher gegen X

$$\Leftrightarrow \bigwedge_{\varepsilon>0} \lim_{N\to\infty} P\left(\bigcap_{n\geq N} \{|X_n - X| < \varepsilon\}\right) = 1.$$

$$\Leftrightarrow \bigwedge_{\varepsilon>0} \lim_{N\to\infty} P\left(\bigcup_{n>N} \{|X_n-X| \ge \varepsilon\}\right) = 0.$$

Beweis

$$\begin{split} E &:= \{\omega; \ \lim_{n \to \infty} X_n(\omega) = X(\omega)\} \\ &= \{\omega; \ \lim_{n \to \infty} |X_n(\omega) - X(\omega)| = 0\} \\ &= \bigcap_{k \ge 1} \bigcap_{N \ge 1} \bigcap_{n \ge N} \left\{\omega; \ |X_n(\omega) - X(\omega)| < \frac{1}{k}\right\} \end{split}$$

"⇐": Insbesondere gilt

$$\bigwedge_{k \in \mathbb{N}} 1 = \lim_{N \to \infty} P\left(\bigcap_{n \ge N} \left\{ |X_n - X| < \frac{1}{k} \right\} \right) \stackrel{(3.1.14)}{=} P\left(\bigcup_{N \ge 1} \bigcap_{n \ge N} \left\{ |X_n - X| < \frac{1}{k} \right\} \right).$$

Daraus folgt

$$P(E) = P\left(\bigcap_{k \ge 1} \bigcup_{N \ge 1} \bigcap_{n \ge N} \{\omega; |X_n(\omega) - X(\omega)| < \frac{1}{k}\}\right)$$

$$\stackrel{(3.1.14)}{=} \lim_{k \to \infty} P\left(\bigcup_{N \ge 1} \bigcap_{n \ge N} \{|X_n - X| < \frac{1}{k}\}\right) = 1.$$

"⇒"

$$1 = P(E) = P\left(\bigcap_{k \ge 1} \bigcap_{N \ge 1} \bigcap_{n \ge N} \{\omega; |X_n(\omega) - X(\omega)| < \frac{1}{k}\}\right)$$

$$\leq P\left(\bigcup_{N \ge 1} \bigcap_{n \ge N} \{\omega; |X_n(\omega) - X(\omega)| < \frac{1}{k}\}\right)$$

$$= \lim_{N \to \infty} P\left(\bigcap_{n \ge N} \{\omega; |X_n(\omega) - X(\omega)| < \frac{1}{k}\}\right)$$

für alle $k \in \mathbb{N}$.

Sei $\varepsilon > 0$ beliebig. Dann gibt es $k \in \mathbb{N}$ mit $\frac{1}{k} < \varepsilon$ und

$$1 = \lim_{N \to \infty} P\left(\bigcap_{n \ge N} \left\{ |X_n - X| < \frac{1}{k} \right\} \right) \le \lim_{N \to \infty} P\left(\bigcap_{n \ge N} \left\{ |X_n - X| < \varepsilon \right\} \right).$$

9.2.3 Satz (Starkes Gesetz der großen Zahlen)

Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von paarweisen unkorrelierten Zufallsgrößen mit $\operatorname{var}(X_n) \leq c < \infty$ für $n \in \mathbb{N}$, so gilt:

$$Z_N = \frac{1}{N} \sum_{n=1}^{N} (X_n - E(X_n))$$

konvergiert fast sicher gegen 0.

Beweis

Sei $\varepsilon > 0$. Für jedes $N \in \mathbb{N}$ gibt es M = M(N) mit $M^2 \le N < (M+1)^2$.

Ohne Beschränkung der Allgemeinheit können wir $E(X_n) = 0$ annehmen.

Setze
$$S_N = \sum_{n=1}^N X_n$$

und

$$A_N = \left\{ \frac{1}{N} |S_N| \ge 2\varepsilon \right\}$$

$$B_M = \left\{ \frac{1}{M} |S_M| \ge \varepsilon \right\}$$

$$C_N = \left\{ \frac{1}{M(N)^2} |S_{M(N)^2} - S_N| \ge \varepsilon \right\}.$$

Dann gilt

$$A_{N} = \{|S_{N}| \geq N2\varepsilon\}$$

$$\subset \{|S_{M(N)^{2}}| \geq N\varepsilon\} \cup \{|S_{M(N)^{2}} - S_{N}| \geq N\varepsilon\}$$

$$\subset B_{M(N)^{2}} \cup C_{N}.$$

Weiterhin gilt mit Satz 8.2.8

$$\operatorname{var}\left(rac{1}{M^2}S_{M^2}
ight) = rac{1}{M^4}\sum_{n=1}^{M^2}\operatorname{var}(X_n) \leq rac{c}{M^2}$$

und

$$\mathrm{var}\left(\tfrac{1}{M(N)^2}(S_{M(N)^2}-S_N)\right) = \mathrm{var}\left(\tfrac{1}{M(N)^2}\sum_{n=M(N)^2+1}^N X_n\right) \leq \tfrac{1}{M(N)^4} \ c \ (N-M(N)^2).$$

Mit der Tschebyscheff-Ungleichung und der paarweisen Unkorreliertheit von $(X_n)_{n\in\mathbb{N}}$ folgt

$$P(B_{M^2}) \le \frac{c}{\varepsilon^2 M^2}$$
 und $P(C_N) \le \frac{c (N - M(N)^2)}{\varepsilon^2 M(N)^4}$.

Summieren wir über N, so erhalten wir

$$\sum_{M=1}^{\infty} P(B_{M^2}) < \infty$$

und

$$\sum_{N=1}^{\infty} P(C_N) \leq \frac{c}{\varepsilon^2} \sum_{M=1}^{\infty} \sum_{N=M^2}^{(M+1)^2 - 1} \frac{N - M^2}{M^4} = \frac{c}{\varepsilon^2} \sum_{M=1}^{\infty} \frac{1}{M^4} (1 + 2 + \dots + 2M)$$
$$= \frac{c}{\varepsilon^2} \sum_{M=1}^{\infty} \frac{(2M)(2M+1)}{M^4 \cdot 2} < \infty.$$

Nach dem Borel-Cantelli-Lemma (i) (Satz 9.2.1) gilt

$$P\left(\bigcap_{n\geq 1}\bigcup_{N\geq n}A_{N}\right) \leq P\left(\bigcap_{n\geq 1}\bigcup_{N\geq n}(B_{M(N)^{2}}\cup C_{N})\right)$$

$$= P\left(\bigcap_{n\geq 1}\bigcup_{N\geq n}B_{M(N)^{2}}\cup\bigcap_{n\geq 1}\bigcup_{N\geq n}C_{N}\right)$$

$$\leq P\left(\bigcap_{n\geq 1}\bigcup_{M\geq n}B_{M^{2}}\right) + P\left(\bigcap_{n\geq 1}\bigcup_{N\geq n}C_{N}\right) \stackrel{(9.2.1)(i)}{=} 0 + 0.$$

Es folgt
$$\lim_{n\to\infty} P\left(\bigcap_{N\geq n} \left\{\frac{1}{N}|S_N| < 2\varepsilon\right\}\right) = P\left(\overline{\bigcap_{n\geq 1} \bigcup_{N\geq n} A_N}\right) = 1$$

für alle $\varepsilon>0$ und mit Satz 9.2.2 die fast sichere Konvergenz von $\frac{1}{N}|S_N|=|Z_N|$ gegen 0. \sharp

9.2.4 Definition (Empirische Verteilungsfunktion)

Für Zufallsgrößen X_1, \ldots, X_N heißt

$$F_N: \mathbb{R} \times \Omega \ni (x, \omega) \to F_N(x, \omega) := \frac{1}{N} \sum_{n=1}^N 1_{(-\infty, x]}(X_n(\omega))$$

die empirische Verteilungsfunktion von X_1, \ldots, X_N .

9.2.5 Folgerung (Punktweise Konvergenz der empirischen Verteilungsfunktion) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen mit Verteilungsfunktion $F_{X_n} = F$ für $n \in \mathbb{N}$. Für alle $x \in \mathbb{R}$ konvergiert dann

$$F_N(x,\cdot) = \frac{1}{N} \sum_{n=1}^{N} 1_{(-\infty,x]}(X_n)$$

fast sicher gegen F(x).

Beweis

Sei $x \in \mathbb{R}$ beliebig. Setze $Y_n = 1_{(-\infty,x]}(X_n)$.

Dann gilt für alle $n \in \mathbb{N}$

$$E(Y_n) = E(1_{(-\infty,x]}(X_n)) = P^{X_n}((-\infty,x]) = F_{X_n}(x) = F(x)$$

und

$$\operatorname{var}(Y_n) = \operatorname{var}(1_{(-\infty,x]}(X_n)) = F(x) \ (1 - F(x))$$

(Varianz eines B(1, F(x))-Experimentes).

Mit dem starken Gesetz der großen Zahlen (Satz 9.2.3) konvergiert

$$F_N(x,\cdot) - F(x) = \frac{1}{N} \sum_{n=1}^{N} (Y_n - E(Y_n))$$

#

fast sicher gegen 0.

9.2.6 Satz (Glivenko-Cantelli-Lemma)

(Gleichmäßige Konvergenz der empirischen Verteilungsfunktion)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen mit Verteilungsfunktionen $F_{X_n}=F$ für alle $n\in\mathbb{N}$ und

$$D_N: \Omega \ni \omega \to D_N(\omega) = \sup_{x \in \mathbb{R}} |F_N(x, \omega) - F(x)|.$$

Dann konvergiert D_N fast sicher gegen 0.

Beweis

Siehe zum Beispiel Billingsley (1995), Probability and Measure, S. 269.

9.2.7 Definition (Quantil, Median)

- (i) Sei X Zufallsgröße mit Verteilungsfunktion F_X und $\alpha \in [0,1]$. Dann heißt
 - (a) $q_X(\alpha) := \inf\{x; F_X(x) \ge \alpha\} \text{ das } \alpha\text{-Quantil der Verteilung von } X$,
 - (b) $med(X) = q_X(1/2) der Median der Verteilung von X.$
- (ii) Sind X_1, \ldots, X_N Zufallsgrößen, $\alpha \in [0, 1]$ und F_N die empirische Verteilungsfunktion von X_1, \ldots, X_N , dann heißt

$$q_N(\alpha) := q_{X_1(\omega),\dots,X_N(\omega)}(\alpha) := \inf\{x; F_N(x,\omega) \ge \alpha\}$$

das empirische α -Quantil bzgl. X_1, \ldots, X_N .

9.2.8 Satz (Starkes Gesetz der großen Zahlen für α -Quantile)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge unabhängiger Zufallsgrößen mit Verteilungsfunktion $F_{X_n} = F$ für $n \in \mathbb{N}$, $\alpha \in (0,1)$ und $F_{X_n}(q_{X_n}(\alpha) + \varepsilon) > \alpha$ für alle $\varepsilon > 0$. Dann konvergiert

$$q_N(\alpha)$$
 fast sicher gegen $q_{X_n}(\alpha)$.

Beweis

Sei $\varepsilon > 0$, $F = F_{X_n}$ und $q = q_{X_n}(\alpha)$. Nach Voraussetzung und Definition von $q_{X_n}(\alpha) = q$ gibt es $\delta > 0$ mit

$$F(q-\varepsilon) \le \alpha - \delta, \quad F(q+\varepsilon) \ge \alpha + \delta.$$

Nach Satz 9.2.5 und Satz 9.2.2 gilt

$$\lim_{M \to \infty} P\left(\bigcup_{N \ge M} \{|F_N(q - \varepsilon) - F(q - \varepsilon)| \ge \delta/2 \text{ oder } |F_N(q + \varepsilon) - F(q + \varepsilon)| \ge \delta/2\}\right)$$

$$\leq \lim_{M \to \infty} P\left(\bigcup_{N \ge M} \{|F_N(q - \varepsilon) - F(q - \varepsilon)| \ge \delta/2\}\right)$$

$$+ \lim_{M \to \infty} P\left(\bigcup_{N \ge M} \{|F_N(q + \varepsilon) - F(q + \varepsilon)| \ge \delta/2\}\right) = 0$$

und somit

$$1 = \lim_{M \to \infty} P\left(\bigcap_{N \ge M} \{|F_N(q - \varepsilon) - F(q - \varepsilon)| < \delta/2 \text{ und } |F_N(q + \varepsilon) - F(q + \varepsilon)| < \delta/2\}\right)$$

$$\leq \lim_{M \to \infty} P\left(\bigcap_{N \ge M} \{F_N(q - \varepsilon) < F(q - \varepsilon) + \delta/2 \text{ und } F_N(q + \varepsilon) > F(q + \varepsilon) - \delta/2\}\right)$$

$$\leq \lim_{M \to \infty} P\left(\bigcap_{N \ge M} \{F_N(q - \varepsilon) < \alpha - \delta/2 \text{ und } F_N(q + \varepsilon) > \alpha + \delta/2\}\right)$$

#

$$\leq \lim_{M \to \infty} P\left(\bigcap_{N \geq M} \{|q_N(\alpha) - q_{X_n}(\alpha)| < \varepsilon\}\right).$$

Satz 9.2.2 liefert dann die Behauptung.

9.2.9 Beispiel (Stochastische Simulation von Erwartungswert und Varianz)

Können Erwartungswert oder Varianz nicht explizit berechnet werden, können sie per Simulation bestimmt werden. Dazu erzeugt man von der Verteilung genügend Zufallszahlen x_1, x_2, \ldots, x_N . Der Erwartungswert wird dann nach dem Gesetz der großen Zahlen (Satz 9.1.9 und Satz 9.2.3) durch das arithmetische Mittel \overline{x} der Zufallszahlen approximiert.

9.2.10 Beispiel (Integration mittels stochastischer Simulation)

Zufallszahlen können mittels des Gesetzes der großen Zahlen auch zur Berechnung von Integralen benutzt werden. Verschiedene Methoden der simulierten Integration können außerdem mittels Erwartungswert und Varianz verglichen werden.

9.3 Verteilungskonvergenz

9.3.1 Definition (Verteilungskonvergenz)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von Zufallsgrößen auf (Ω, \mathcal{A}, P) mit Verteilungsfunktionen F_{X_n} und X Zufallsgröße auf (Ω, \mathcal{A}, P) mit Verteilungsfunktion F_X .

$(X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen X (kurz $X_n \xrightarrow{\mathcal{D}} X$)

 $\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \text{ für alle } x \in \mathbb{R}, \text{ bei denen } F_X \text{ stetig ist.}$

9.3.2 Beispiel

Sei $Y \sim R[-1,1]$ (= U(-1,1)), d. h. Y besitzt eine Rechteckverteilung auf [-1,1], $X_n = \frac{1}{n} Y$ und $X \sim \varepsilon_0$, d. h. X besitzt eine Einpunktverteilung auf 0.

9.3.3 Lemma

Eine Verteilungsfunktion F_X ist an höchstens abzählbar vielen Stellen nicht stetig.

Beweis

Da eine Verteilungsfunktion immer rechtsseitig stetig ist (2.23)(ii), kann höchstens der linksseitige Limes nicht mit dem Funktionswert übereinstimmen. D. h. ist F_X bei x nicht stetig, so gilt

$$P^X(\{x\}) = F_X(x) - \lim_{b \uparrow_x} F_X(b) > 0.$$

Sei

$$A := \{x; F_X(x) - \lim_{b \uparrow x} F_X(b) > 0\}$$

und

$$A_n := \left\{ x; F_X(x) - \lim_{b \uparrow x} F_X(b) > \frac{1}{n} \right\}.$$

Dann gilt

$$A = \bigcup_{n=1}^{\infty} A_n = \lim_{n \to \infty} A_n$$

und

$$1 \ge P^X(A_n) = \sum_{x \in A_n} P^X(\{x\}) > \sum_{x \in A_n} \frac{1}{n} = \frac{\#A_n}{n} ,$$

woraus $\#A_n \leq n$ folgt. Also hat $A = \lim_{n \to \infty} A_n$ höchstens abzählbar unendlich viele Elemente.

9.3.4 Satz (Lemma von Slutzky)

Seien $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ Folgen von Zufallsgrößen, X Zufallsgröße. Konvergiert $(X_n)_{n\in\mathbb{N}}$ in Verteilung gegen X und $(|X_n-Y_n|)_{n\in\mathbb{N}}$ stochastisch gegen 0, dann konvergiert auch $(Y_n)_{n\in\mathbb{N}}$ in Verteilung gegen X.

Kurz:

$$X_n \xrightarrow{\mathcal{D}} X, \quad |X_n - Y_n| \xrightarrow{\mathcal{P}} 0$$

 $\Rightarrow Y_n \xrightarrow{\mathcal{D}} X.$

 $(\mathcal{D} \text{ für "Distribution"} = \text{Verteilung}, \mathcal{P} \text{ für "Probability"}).$

Beweis

Sei x_0 ein beliebiger Stetigkeitspunkt von F_X und $\varepsilon > 0$ beliebig. Wegen der Stetigkeit von F_X bei x_0 gibt es $\delta > 0$, so dass für alle $|x - x_0| < \delta$ gilt:

$$|F_X(x_0) - F_X(x)| < \frac{\varepsilon}{3}.$$

Da F_X höchstens an abzählbar vielen Punkten nicht stetig ist, gibt es δ_1 mit $0 < \delta_1 < \delta$, so dass F_X stetig bei $x_0 \pm \delta_1$ ist. Wegen der Verteilungskonvergenz von $(X_n)_{n \in \mathbb{N}}$ und der stochastischen Konvergenz von $(|X_n - Y_n|)_{n \in \mathbb{N}}$ gibt es $N_0 \in \mathbb{N}$ mit

$$|F_{X_n}(x_0 \pm \delta_1) - F_X(x_0 \pm \delta_1)| < \frac{\varepsilon}{3}$$
$$P(|X_n - Y_n|) > \delta_1) < \frac{\varepsilon}{3}$$

für alle $n \geq N_0$. Für $n \geq N_0$ folgt dann

$$F_{Y_n}(x_0) = P(Y_n \in (-\infty, x_0])$$

$$= P(Y_n \le x_0, |X_n - Y_n| \le \delta_1) + P(Y_n \le x_0, |X_n - Y_n| > \delta_1)$$

$$\le P(X_n \le x_0 + \delta_1) + P(|X_n - Y_n| > \delta_1)$$

$$< F_{X_n}(x_0 + \delta_1) + \frac{\varepsilon}{3} < F_X(x_0 + \delta_1) + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} < F_X(x_0) + \varepsilon$$

und

$$F_{X}(x_{0}) - \varepsilon < F_{X}(x_{0} - \delta_{1}) - \frac{2}{3} \varepsilon$$

$$< F_{X_{n}}(x_{0} - \delta_{1}) - \frac{\varepsilon}{3} = P(X_{n} \le x_{0} - \delta_{1}) - \frac{\varepsilon}{3}$$

$$\le P(X_{n} \le x_{0} - \delta_{1}, |X_{n} - Y_{n}| \le \delta_{1}) + P(|X_{n} - Y_{n}| > \delta_{1}) - \frac{\varepsilon}{3}$$

$$< P(Y_{n} \le x_{0}) = F_{Y_{n}}(x_{0}),$$

also $|F_{Y_n}(x_0) - F_X(x_0)| < \varepsilon$.

9.3.5 Satz

- (i) $(X_n)_{n\in\mathbb{N}}$ konvergiert stochastisch gegen X \Rightarrow $(X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen X.
- (ii) Die Umkehrung von (i) gilt im Allgemeinen nicht.
- (iii) $(X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen $a\in\mathbb{R}$ \Rightarrow $(X_n)_{n\in\mathbb{N}}$ konvergiert stochastisch gegen a.
- (iv) $(\sqrt{n}X_n)_{n\in\mathbb{N}}$ konvergiert in Verteilung gegen X \Rightarrow $(X_n)_{n\in\mathbb{N}}$ konvergiert stochastisch gegen 0.

Beweis

(i) Setze im Lemma von Slutzky (Satz 9.3.4) $\widetilde{X}_n = X$, $\widetilde{Y}_n = X_n$. Dann gilt $\widetilde{X}_n = X \xrightarrow{\mathcal{D}} X$, $|\widetilde{X}_n - \widetilde{Y}_n| = |X - X_n| \xrightarrow{\mathcal{P}} 0$ und somit mit dem Lemma von Slutzky

$$X_n = \widetilde{Y}_n \xrightarrow{\mathcal{D}} X$$
.

(ii) Gegenbeispiel: Sei $X \sim R[-1,1], X_n = (-1)^n X$. Dann gilt $X_n \sim R[-1,1]$ und somit $X_n \xrightarrow{\mathcal{D}} X.$

Aber für ungerades n und $\varepsilon \in (0,1)$ gilt:

$$P(|X_n - X| > \varepsilon) = P(|-X - X| > \varepsilon) = P(2|X| > \varepsilon) = P(|X| > \frac{\varepsilon}{2})$$

$$= P\left(-1 \le X \le -\frac{\varepsilon}{2}\right) + P\left(\frac{\varepsilon}{2} \le X \le 1\right) = \int_{-1}^{-\varepsilon/2} \frac{1}{2} dx + \int_{\varepsilon/2}^{1} \frac{1}{2} dx = 1 - \frac{\varepsilon}{2}.$$

Also konvergiert $P(|X_n - X| > \varepsilon)$ nicht gegen 0, weshalb X_n nicht stochastisch gegen X konvergieren kann.

(iii) Konvergiert $(X_n)_{n\in\mathbb{N}}$ in Verteilung gegen $a\in\mathbb{R}$, so gilt $\lim_{n\to\infty}F_{X_n}(x)=0 \text{ für } x< a \text{ und } \lim_{n\to\infty}F_{X_n}(x)=1 \text{ für } x>a,$ da a als Zufallsgröße die Einpunktverteilung ε_a besitzt. Es folgt für alle $\varepsilon>0$

$$\lim_{n \to \infty} P(|X_n - a| > \varepsilon) = \lim_{n \to \infty} P(X_n < a - \varepsilon \text{ oder } X_n > a + \varepsilon)$$

$$\leq \lim_{n \to \infty} \left(F_{X_n}(a - \varepsilon) + 1 - F_{X_n}(a + \varepsilon) \right) = 0,$$

also konvergiert $(X_n)_{n\in\mathbb{N}}$ stochastisch gegen a.

#

(iv) Übung!

9.3.6 Satz

Seien $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, $(Z_n)_{n\in\mathbb{N}}$ Folgen von Zufallsgrößen mit $X_n \xrightarrow{\mathcal{P}} 0$, $Y_n \xrightarrow{\mathcal{P}} a$ und $Z_n \xrightarrow{\mathcal{P}} Z$. Dann gilt:

- (i) $X_n + Z_n \xrightarrow{\mathcal{D}} Z$,
- (ii) $Y_n + Z_n \xrightarrow{\mathcal{D}} a + Z$,
- (iii) $X_n \cdot Z_n \xrightarrow{\mathcal{P}} 0$,
- (iv) $Y_n \cdot Z_n \xrightarrow{\mathcal{D}} a \cdot Z$,
- (v) $Z_n/Y_n \xrightarrow{\mathcal{D}} Z/a$, falls $a \neq 0$.

Beweis

- (i) Folgt aus (ii).
- (ii) Nach dem Lemma von Slutzky (Satz 9.3.4) gilt $\tilde{Z}_n := Y_n a + Z_n \xrightarrow{\mathcal{D}} Z$. Daraus folgt $Y_n + Z_n = \tilde{Z}_n + a \xrightarrow{\mathcal{D}} Z + a$, denn

$$\lim_{n \to \infty} F_{\tilde{Z}_n + a}(x) = \lim_{n \to \infty} P(\tilde{Z}_n + a \le x)$$

$$= \lim_{n \to \infty} P(\tilde{Z}_n \le x - a) = P(Z \le x - a) = P(Z + a \le x) = F_{Z + a}(x),$$

falls x-a Stetigkeitspunkt von Z ist. x-a ist aber genau dann Stetigkeitspunkt von Z, wenn x Stetigkeitspunkt von Z+a ist.

- (iii) Folgt aus (iv).
- (iv) Wegen $Z_n \xrightarrow{\mathcal{D}} Z$ gibt es für alle $\delta > 0$ ein $k \in \mathbb{N}$ und ein $N_1 \in \mathbb{N}$ mit

$$P(|Z_n| > k) = P(Z_n < -k) + P(Z_n > k) \le F_{Z_n}(-k) + 1 - F_{Z_n}(k)$$

$$\le F_Z(-k) + \delta + 1 - F_Z(k) + \delta < 4\delta \text{ für alle } n \ge N_1.$$

Dann gibt es für alle $\varepsilon > 0$ ein $N_0 \ge N_1$ mit

$$P(|Y_n \cdot Z_n - a Z_n| > \varepsilon) = P(|Z_n| |Y_n - a| > \varepsilon)$$

$$\leq P(|Z_n| |Y_n - a| > \varepsilon, |Z_n| \leq k) + P(|Z_n| > k)$$

$$\leq P(|Y_n - a| > \frac{\varepsilon}{k}) + 4\delta \leq 5\delta$$

für alle $n \geq N_0$. Es folgt

$$|Y_n \cdot Z_n - aZ_n| \xrightarrow{\mathcal{P}} 0$$
.

Außerdem folgt aus $Z_n \xrightarrow{\mathcal{D}} Z$ auch $aZ_n \xrightarrow{\mathcal{D}} aZ$, denn

$$\lim_{n \to \infty} F_{aZ_n}(x) = \lim_{n \to \infty} P(aZ_n \le x) = \lim_{n \to \infty} P\left(Z_n \le \frac{x}{a}\right)$$
$$= \lim_{n \to \infty} F_{Z_n}\left(\frac{x}{a}\right) = F_Z\left(\frac{x}{a}\right) = F_{aZ}(x),$$

falls F_{aZ} bei x stetig ist.

Mit dem Lemma von Slutzky (Satz 9.3.4) folgt dann die Behauptung.

(v) Übung. \sharp

9.3.7 Beispiel (Gegenbeispiel)

Im Allgemeinen gilt nicht, dass aus $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \xrightarrow{\mathcal{D}} Y$ auch $X_n + Y_n \xrightarrow{\mathcal{D}} X + Y$ folgt.

9.4 Zentraler Grenzwertsatz

In diesem Abschnitt wird ein Grenzwertresultat bezüglich der Verteilungskonvergenz vorgestellt. Es behandelt nicht die Konvergenz der Zufallsvariablen selbst, sondern liefert eine Approximation der Verteilungsfunktion der standardisierten Zufallsvariablen S_N^* für große Werte von N. Es besagt, dass für unabhängige X_1, X_2, \ldots

$$\sqrt{N}\left(\frac{1}{N}\sum_{n=1}^{N}\left(X_{n}-E(X_{n})\right)\right)$$

für $N \to \infty$ in Verteilung gegen eine Zufallsvariable mit Normalverteilung konvergiert.

Ein elementarer Beweis für diese Aussage befindet sich in Krengel (1991), Abschnitt 11.3. Setzt man noch voraus, dass die Zufallsgrößen X_n eine Bernoulli-Verteilung besitzen, so dass $\sum_{n=1}^{N} X_n$ eine Binomialverteilung mit Parameter N und p besitzt, so wird der Beweis noch einfacher. Der zentrale Grenzwertsatz für diesen Fall ist auch als Satz von Moivre-Laplace bekannt. Sein Beweis befindet sich in Krengel (1991), Abschnitt 5.2. Da beide Beweise aber ansonsten nicht viele Einsichten liefern, geben wir hier nur die Zentralen Grenzwertsätze für beide Fälle an und veranschaulichen insbesondere den Satz von Moivre-Laplace. Der Beweis des allgemeinen Zentralen Grenzwertsatzes, und damit auch des Spezialfalls des Satzes von Moivre-Laplace, wird in Statistik V mit Mitteln geführt, die auch für sich genommen von Interesse sind, aber nicht elementar sind.

9.4.1 Satz (Zentraler Grenzwertsatz von Lindeberg-Lévy)

Ist $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsgrößen mit $\operatorname{var}(X_n) < \infty$, dann konvergiert

$$\sqrt{N} \, \left(\frac{1}{N} \, \sum_{n=1}^{N} \, \frac{X_n - E(X_n)}{\sqrt{\mathit{var}(X_n)}} \right) \quad \mathit{für} \, \, N \to \infty$$

in Verteilung gegen X, das eine Standard-Normalverteilung besitzt, d.h. X hat eine Normalverteilung mit Erwartungswert 0 und Varianz 1.

9.4.2 Folgerung

Seien X_1, X_2, \ldots unabhängige und identisch verteilte Zufallsvariablen mit $E(X_1) = \mu$, $0 < var(X_1) = \sigma^2 < \infty$ und $S_N = \sum_{n=1}^N X_n$. Dann gilt für $a, b \in \mathbb{R}$, a < b:

1.
$$\lim_{N \to \infty} P\left(a \le \frac{S_N - N\mu}{\sqrt{N}\sigma} \le b\right) = \int_a^b \varphi(t) dt = \Phi(b) - \Phi(a).$$

2.
$$\lim_{N \to \infty} P\left(\frac{S_N - N\mu}{\sqrt{N}\sigma} \le b\right) = \int_{-\infty}^b \varphi(t) dt = \Phi(b).$$

Dabei bezeichnet Φ die Verteilungsfunktion der Standard-Normalverteilung und es gilt $\varphi = \Phi'$, d.h. φ ist die Ableitung von Φ .

Alternativ können die Aussagen auch mit dem arithmetischen Mittel $\overline{X}_N = \frac{1}{N}S_N$ formuliert werden:

1.
$$\lim_{N \to \infty} P\left(a \le \sqrt{N} \frac{\overline{X}_N - \mu}{\sigma} \le b\right) = \int_a^b \varphi(t) dt = \Phi(b) - \Phi(a).$$

2.
$$\lim_{N \to \infty} P\left(\sqrt{N} \frac{\overline{X}_N - \mu}{\sigma} \le b\right) = \int_{-\infty}^b \varphi(t) dt = \Phi(b).$$

Wir motivieren jetzt den Zentralen Grenzwertsatz von de Moivre-Laplace, der ein Spezialfall des Zentralen Grenzwertsatzes von Lindeberg-Lévy ist. Dazu betrachten wir den Fall von stochastisch unabhängigen Bin(1,p)-verteilten Zufallsvariablen X_1, X_2, \ldots und setzen $S_n = \sum_{n=1}^N X_n$. Wegen $E(S_N) = Np$ und $var(S_N) = Np(1-p)$ tendieren sowohl Erwartungswert als auch Varianz gegen unendlich. Beide Effekte sind "unschön" und werden durch die Standardisierung

$$S_N^* = \frac{S_N - E(S_N)}{\sqrt{\operatorname{var}(S_N)}} = \frac{S_N - Np}{\sqrt{Np(1-p)}}$$

verhindert, da $E(S_N^*) = 0$ und $\text{var}(S_N^*) = 1$ für jedes $N \in \mathbb{N}$ gilt. Die Verteilung von S_N^* ist gegeben durch

$$P(S_N^* = s_{N,n}) = P(S_N = n) = \binom{N}{n} p^n (1-p)^{N-n}, \quad n = 0, \dots, N,$$

wobei die Trägerpunkte von $P^{S_N^*}$ gegeben sind durch

$$s_{N,n} = \frac{n - Np}{\sqrt{Np(1-p)}}, \quad n = 0, \dots, N, N \in \mathbb{N}.$$

Im folgenden illustrieren einige Histogramme die Zähldichte von S_N^* für die Werte N=5,10,20,50,100 und p=0.2. Die Balkenbreite $b_n=b$ berechnet sich gemäß der Formel $1/\sqrt{Np(1-p)}$. Insbesondere ist die Balkenbreite unabhängig von n, so dass alle Balken die gleiche Breite besitzen. Gemäß der Definition des Histogramms entsprechen die Flächen der Balken der Wahrscheinlichkeit $P(S_N^*=s_{N,n})$, so dass die Höhen gegeben sind durch

$$h_n = \frac{P(S_N^* = s_{N,n})}{b} = \sqrt{Np(1-p)} \cdot \binom{N}{n} p^n (1-p)^{N-n}, \quad n = 0, \dots, N.$$

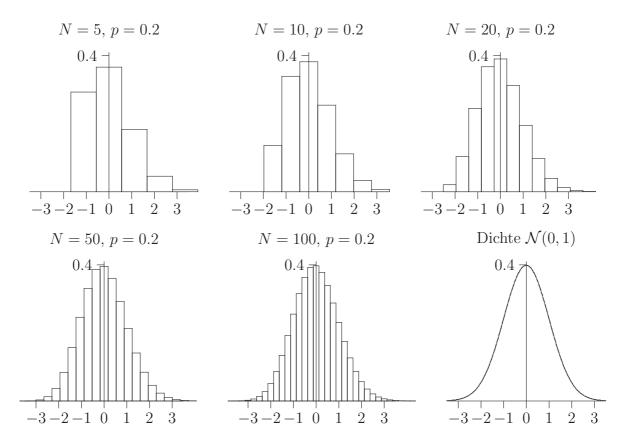
Die Klassenmitten sind $s_{N,n}$, so dass die Klassengrenzen von Klasse K_n bestimmt sind durch $v_n = s_{N,n} - \frac{b}{2}$ und $v_{n+1} = s_{N,n} + \frac{b}{2}$, $n = 0, \dots, N$. Es gilt also:

$$K_0 = \left[s_{N,0} - \frac{b}{2}, s_{N,0} + \frac{b}{2} \right], \quad K_n = \left(s_{N,n} - \frac{b}{2}, s_{N,n} + \frac{b}{2} \right], \quad n = 1, \dots, N.$$

Für N = 5 und p = 0.2 ergeben sich die folgenden Klassen:

$$K_0 = [-1.677, -0.559], \quad K_1 = (-0.559, 0.559], \quad K_2 = (0.559, 1.677],$$

 $K_3 = (1.677, 2.795], \quad K_4 = (2.795, 3.913], \quad K_5 = (3.913, 5.031].$



Die Umrisse der Histogramme nähern sich offensichtlich der Dichte φ der Standardnormalverteilung an, wobei

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

Naheliegend ist auch, dass die Fläche des Histogramms bis zu einer Stelle z die Fläche zwischen der Dichtefunktion und der x-Achse approximiert. Dieses wird zusammengefaßt in der Aussage:

9.4.3 Satz (Zentraler Grenzwertsatz von de Moivre-Laplace) Sei $S_N \sim Bin(N, p), p \in (0, 1)$. Dann gilt für $a, b \in \mathbb{R}, a < b$:

1.
$$\lim_{N \to \infty} P\left(a \le \frac{S_N - Np}{\sqrt{Np(1-p)}} \le b\right) = \int_a^b \varphi(t) dt = \Phi(b) - \Phi(a).$$

2.
$$\lim_{N \to \infty} P\left(\frac{S_N - Np}{\sqrt{Np(1-p)}} \le b\right) = \int_{-\infty}^b \varphi(t) dt = \Phi(b).$$

Die obige Grenzwertaussage wird folgendermaßen angewendet. Sei $S_N \sim \text{Bin}(N, p)$. Dann ist die Verteilungsfunktion von S_N an einer Stelle $k \in \{0, ..., N\}$ gegeben durch

$$F^{S_N}(k) = \sum_{i=0}^k \binom{N}{i} p^i (1-p)^{N-i}.$$

Ist N groß, so müssen i.a. aufwendige Berechnungen durchgeführt werden. Daher liefert der Zentrale Grenzwertsatz eine einfache Approximationsmöglichkeit gemäß:

$$F^{S_N}(k) = P(S_N \le k) = P\left(S_N^* \le \frac{k - Np}{\sqrt{Np(1 - p)}}\right) \approx \Phi\left(\frac{k - Np}{\sqrt{Np(1 - p)}}\right).$$

Für p = 0.3, N = 100 und k = 40 erhält man als Approximation $\Phi(2.18) = 0.985371$. Der exakte Wert ist $P(S_N \le 40) = 0.987502$.

Eine **Faustregel** zur Anwendbarkeit des Zentralen Grenzwertsatzes von de Moivre-Laplace ist: Gilt

$$N \cdot p \cdot (1 - p) \ge 9 = 3^2$$

so sind die Approximationen aus Satz 9.4.3 für Anwendungen brauchbar. Da Np(1-p) die Varianz von S_N ist, bedeutet diese Forderung, dass die Varianz hinreichend groß sein muß. (Man beachte, dass die Dichtefunktion φ außerhalb des Intervalls [-3,3] weniger als 0.3% ihrer Fläche besitzt; siehe 4.14.7)

Abgewandelter Zentraler Grenzwertsatz von Moivre-Laplace: In der Inferenz-Statistik wird der Zentrale Grenzwertsatz von Moivre-Laplace vor allem in der folgenden abgewandelten Form benutzt:

$$P\left(a \le \sqrt{N} \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} \le b\right) \approx \Phi(b) - \Phi(a),$$

wobei $\hat{p} = \frac{1}{N} S_N$ gesetzt wird.

9.4.4 Beispiel

Wir wenden nun den Zentralen Grenzwertsatzes von de Moivre-Laplace auf den 1000fachen Würfelwurf an. Folgende Ereignisse werden betrachtet:

- 1. Die Anzahlen von Sechsen beträgt genau 166.
- 2. Es werden zwischen 150 und 200 Sechsen beobachtet.
- 3. Es werden mehr als 180 Sechsen beobachtet.

Die absolute Häufigkeit $S_{1000} = 1000 \cdot f^{(1000)}$ der Ziffer Sechs folgt einer Bin $(1000, \frac{1}{6})$ -Verteilung. Daher kann wegen $Np(1-p) = 1000 \cdot \frac{1}{6} \cdot \frac{5}{6} = 138.89$ Satz 9.4.3 benutzt werden. Die Wahrscheinlichkeit des ersten Ereignisses lässt sich folgendermaßen approximieren:

$$P(S_{1000} = 166) = P(S_{1000} \le 166) - P(S_{1000} \le 165)$$

$$\approx \Phi\left(\frac{166 - 166.67}{\sqrt{138.89}}\right) - \Phi\left(\frac{165 - 166.67}{\sqrt{138.89}}\right)$$

$$= \Phi(-0.06) - \Phi(-0.14) = \Phi(0.14) - \Phi(0.06)$$

$$= 0.555670 - 0.523922 = 0.031748.$$

Der exakte Wert beträgt 0.069007. Für das Ereignis $\{150 \le S_{1000} \le 200\}$ erhält man die Approximation:

$$P(150 \le S_{1000} \le 200) = P(S_{1000} \le 200) - P(S_{1000} \le 149)$$

$$\approx \Phi\left(\frac{200 - 166.67}{\sqrt{138.89}}\right) - \Phi\left(\frac{149 - 166.67}{\sqrt{138.89}}\right)$$

$$= \Phi(2.83) - \Phi(-1.50) = \Phi(2.83) + \Phi(1.50) - 1$$

$$= 0.997599 + 0.933193 - 1 = 0.930792.$$

Der exakte Wert ist 0.926453. Für das dritte Ereignis berechnet man:

$$P(S_{1000} > 180) = 1 - P(S_{1000} \le 180) \approx 1 - \Phi\left(\frac{180 - 166.67}{\sqrt{138.89}}\right)$$

= 1 - \Phi(1.13) = 1 - 0.870762 = 0.129238.

Der exakte Wert ist 0.120896.

Die folgenden Beispiele betreffen Anwendungen des Zentralen Grenzwertsatz von Lindeberg-Lévy.

9.4.5 Beispiel

Ein Unternehmen produziert Glühbirnen, deren Lebensdauer mit $\mu=10.000$ Stunden angegeben wird. Die Standardabweichung σ wird mit 1.000 Stunden veranschlagt. Dann beträgt die Wahrscheinlichkeit, dass die mittlere Brenndauer von 400 Glühbirnen mindestens 9.900 Stunden beträgt approximativ

$$P(\overline{X}_{400} \ge 9.900) = P\left(\sqrt{400} \frac{\overline{X}_{400} - 10.000}{1000} \ge \sqrt{400} \frac{9.900 - 10.000}{1.000}\right)$$
$$= P\left(\sqrt{400} \frac{\overline{X}_{400} - 10.000}{1000} \ge -2\right) \approx 1 - \Phi(2) = 0.977250.$$

Die durchschnittliche Brenndauer der 400 Glühbirnen beträgt daher mit einer Wahrscheinlichkeit größer als 97.7% über 9.900 Stunden.

9.4.6 Beispiel

Zur Einschätzung der Leistungsfähigkeit von Fließbandarbeitern wird die zur Bearbeitung eines Standardvorganges notwendige Zeit bei 100 Personen gemessen. Es wird angenommen, dass die Bearbeitungszeit im Mittel 60 sec beträgt. Die Standardabweichung wird mit 10 sec angesetzt.

Gesucht ist der Wert der Wahrscheinlichkeit, dass das arithmetische Mittel von 100 Personen höchstens 2 sec vom Sollwert 60 sec abweicht. Mittels der Tschebyscheff-Ungleichung erhält man folgende untere Schranke für die gesuchte Wahrscheinlichkeit:

$$P(|\overline{X}_{100} - 60| \le 2) = 1 - P(|\overline{X}_{100} - 60| > 2) \ge 1 - \frac{\mathsf{var}\overline{X}_{100}}{2^2} = 1 - \frac{1}{4} = 0.75.$$

Man beachte, dass die Abschätzung für eine beliebige Verteilung richtig ist. Die Anwendung des Zentralen Grenzwertsatzes liefert die Näherung:

$$P(|\overline{X}_{100} - 60| \le 2) = P(-2 \le \overline{X}_{100} - 60 \le 2)$$

$$= P\left(-2 \le \sqrt{100} \frac{\overline{X}_{100} - 60}{10} \le 2\right) \approx \Phi(2) - \Phi(-2) = 2\Phi(2) - 1$$

$$= 0.9545.$$

Dieser Wert ist bei zugrundeliegender Normalverteilung exakt, bei einer anderen Verteilung als approximativ anzusehen.