



Stochastik für Informatiker, Physiker, Chemiker und Wirtschaftswissenschaftler

Universität Ulm
Abteilung Stochastik

Vorlesungsskript
Prof. Dr. Volker Schmidt
Stand: Sommersemester 2001

ULM, IM JULI 2001

1 Einleitung

1.1 Was ist Stochastik?

Der Begriff *Stochastik* stammt ursprünglich aus dem Griechischen und bedeutet dort: die Kunst des geschickten Vermutens. Die mathematische Stochastik befaßt sich mit der Beschreibung und Untersuchung von Ereignissen, zeitlichen Entwicklungen bzw. räumlichen Strukturen, die *vom Zufall beeinflusst* werden. Solche Ereignisse, Entwicklungen bzw. Strukturen werden oft durch *Daten* dokumentiert, für deren Analyse die *Statistik* – ein Teilgebiet der Stochastik – geeignete Methoden bereitstellt.

1.2 Typische Fragestellungen und Ergebnisse

Zu den Aufgaben der Stochastik gehört die Bewertung von Ereignissen, Entwicklungen bzw. Strukturen durch die Bestimmung ihrer *Wahrscheinlichkeit*, die eine Maßzahl für die Chance ihres Eintretens ist.

Das Phänomen „Zufall“ kommt in vielfältiger Weise in zahlreichen Bereichen des täglichen Lebens vor, z.B. bei der Vorhersage von zukünftigen Aktienkursen bzw. Zinssätzen, bei der Wettervorhersage, bei Würfel- bzw. Kartenspielen, beim Zahlenlotto, usw.

Stochastische Modelle sind in vielen Disziplinen der Wissenschaft ein wichtiges Hilfsmittel, so in

- Informatik und Ingenieurwissenschaften (z.B. bei der Dimensionierung und Leistungsanalyse von Kommunikations- und Rechnersystemen)
- Physik, Chemie und Materialwissenschaften (z.B. bei der Strukturanalyse von Werkstoffen)
- Wirtschaftswissenschaften (z.B. beim Risikomanagement von Versicherungen und Banken, Analyse von Finanzmärkten)
- Biologie, Medizin, Gesundheitsökonomie und Epidemiologie (z.B. bei Kostenanalyse, Mustererkennung, Bildanalyse)

Es ist üblich, die Stochastik in die folgenden **Teilgebiete** zu unterteilen:

- Wahrscheinlichkeitsrechnung
- Statistik
- stochastische Prozesse und Felder (z.B. Markov-Modelle)
- stochastische Simulation (z.B. Markov-Chain-Monte-Carlo)

Typische Fragestellungen und Ergebnisse der Stochastik sind

- geschlossene Formeln für die Wahrscheinlichkeit von Ereignissen, Entwicklungen bzw. Strukturen (oftmals nur unter restriktiven Modellannahmen möglich)
- Grenzwertsätze (Näherungslösungen, z.B. Gesetz der großen Zahlen, Zentraler Grenzwertsatz)
- Methoden zur Schätzung unbekannter Modellparameter; Tests hypothetischer Modellannahmen (Signifikanztests)
- Kopplung von stochastischer Modellierung, statistischer Datenanalyse und Computer-Simulation

1.3 Beispiel

Betrachten Roulette-Spiel mit 38 möglichen Ausgängen, nämlich 18 rote Felder, 18 schwarze Felder und 2 grüne Felder. Betrachten Spieler, der auf „Rot“ setzt. Er gewinnt 1 Euro mit der Wahrscheinlichkeit $\frac{18}{38}$ und verliert 1 Euro mit der Wahrscheinlichkeit $\frac{20}{38}$. Sei nun X_n der zufällige „Gewinn“ beim n -ten Spiel. Dann gilt:

$$P(X_n = 1) = \frac{9}{19}, \quad P(X_n = -1) = \frac{10}{19}.$$

Die Zufallsgrößen X_1, X_2, \dots sind *unabhängig* und *identisch verteilt*. Betrachten Gesamtgewinn $S_n = X_1 + \dots + X_n$ aus den ersten n Spielen. Die Folge S_1, S_2, \dots heißt *Random Walk*.

Wie groß ist der erwartete Gewinn $\mathbb{E} X_n$ beim n -ten Spiel? Wie groß ist der erwartete Gesamtgewinn $\mathbb{E} S_n$ aus den ersten n Spielen? Es gilt:

$$\begin{aligned} \mathbb{E} X_n &= 1 \cdot \frac{9}{19} + (-1) \cdot \frac{10}{19} = -0.05263 \\ \mathbb{E} S_n &= \mathbb{E} X_1 + \dots + \mathbb{E} X_n = -n \cdot 0.05263 \end{aligned}$$

(Schwaches) Gesetz der großen Zahlen „Für große n ist $\frac{S_n}{n}$ nahe bei $\mathbb{E} X_1$ mit hoher Wahrscheinlichkeit.“

Zentraler Grenzwertsatz „Für große n läßt sich die Wahrscheinlichkeit $P\left(a_n \leq \frac{S_n}{\sqrt{n}} \leq b_n\right)$ durch die Normalverteilung approximieren.“

Beachte Den Begriffsbildungen „Zufallsgröße“, „unabhängig“, „identisch verteilt“, „für große n “, „nahe bei $\mathbb{E} X_1$ “, „mit hoher Wahrscheinlichkeit“ bzw. „Normalverteilung“ liegen mathematische Definitionen zugrunde. Sie gehören zu den Grundbegriffen der Stochastik, die in den folgenden Abschnitten detailliert erläutert werden.

2 Ereignisse und Wahrscheinlichkeiten

2.1 Ereignisse als Mengen

Wir modellieren Ereignisse als Mengen. Dabei ist eine *Menge* eine Zusammenfassung von wohldefinierten und unterscheidbaren Dingen (Elemente) zu einem Ganzen.

Schreibweise Ω Grundmenge, ω Element

$\omega \in \Omega$: ω ist Element von Ω

$\omega \notin \Omega$: ω ist nicht Element von Ω

$A = \{a, b, c, \dots\}$: Die Menge A besteht aus den Elementen a, b, c, \dots

$A = \{\omega : \omega \in \Omega, \omega \text{ hat Eigenschaft } E\}$: A besteht aus denjenigen Elementen ω von Ω , die die Eigenschaft E haben.

Beispiel $\Omega = \mathbb{N}$, $A = \{2, 4, 6, \dots\} = \{n : n \in \mathbb{N}, n \text{ ist durch } 2 \text{ teilbar}\}$

Der Vergleich von Ereignissen erfolgt durch den Vergleich der Mengen, durch die die Ereignisse modelliert werden.

Definition 2.1

1. $A_1 \subset A_2$ bedeutet, A_1 ist Teilmenge von A_2 , d.h. aus $\omega \in A_1$ folgt $\omega \in A_2$
2. $A_1 = A_2$, falls $A_1 \subset A_2$ und $A_2 \subset A_1$.

Betrachten Ereignisse, die bei einem Zufallsexperiment (z.B. Münzwurf, Werfen eines Würfels, Roulette-Spiel, Erzeugen einer Pseudozufallszahl mit einem Zufallszahlengenerator) eintreten können. Dann ist

- $\Omega =$ Menge aller möglichen Versuchsergebnisse (Grundmenge, Grundgesamtheit, Merkmalraum, Stichprobenraum);
- $A_1, A_2 \subset \Omega$: Ereignisse = Teilmengen von Versuchsergebnissen mit bestimmten Eigenschaften;
- $\{\omega\} \subset \Omega$: Elementarereignis = ein (einzelnes) Versuchsergebnis;
- Angenommen: bei einem Versuch wird das Ergebnis ω erzielt. Dann sagen wir: Das Ereignis A tritt ein, falls $\omega \in A$.
- Für $A_1 \subset A_2$ gilt: Wenn A_1 eintritt, dann tritt auch A_2 ein.

Beispiel (einmaliges Würfeln): $\Omega = \{1, 2, 3, 4, 5, 6\}$; Elementarereignisse $\{1\}, \{2\}, \dots, \{6\}$.

Das Ereignis $A_1 = \{2\}$ tritt genau dann ein, wenn die Zahl 2 gewürfelt wird.

Das Ereignis $A_2 = \{2, 4, 6\}$ tritt genau dann ein, wenn eine gerade Zahl gewürfelt wird.

Also gilt: $A_1 \subset A_2$, d.h., wenn A_1 eintritt, dann tritt auch A_2 ein.

Definition 2.2 Diejenige Teilmenge von Ω , die *kein* Element enthält, heißt *leere Menge* und wird mit \emptyset bezeichnet.

Beachte Das Ereignis \emptyset tritt niemals ein und wird deshalb *unmögliches Ereignis* genannt.

2.2 Ereignissysteme

Aus gegebenen Ereignissen A_1, A_2, \dots kann man durch deren „Verknüpfung“ weitere Ereignisse bilden. Dies wird durch die folgenden Mengenoperationen modelliert.

Definition 2.3

1. *Vereinigungsmenge* $A_1 \cup A_2$: Menge aller Elemente, die zu A_1 oder A_2 gehören.
 $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \dots$: Menge aller Elemente, die zu *mindestens einer* der Mengen A_i gehören.
2. *Schnittmenge* $A_1 \cap A_2$: Menge aller Elemente, die zu A_1 und A_2 gehören.
 $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \dots$: Menge aller Elemente, die zu *jeder* der Mengen A_i gehören.
3. *Differenzmenge* $A_1 \setminus A_2$: Menge aller Elemente von A_1 , die *nicht* zu A_2 gehören.
 Spezialfall: $A^c = \Omega \setminus A$ (*Komplement*)

Beispiel Sei $\Omega = \{a, b, c, d\}$, $A_1 = \{a, b, c\}$, $A_2 = \{b, d\}$.

Dann gilt $A_1 \cup A_2 = \{a, b, c, d\}$; $A_1 \cap A_2 = \{b\}$; $A_1 \setminus A_2 = \{a, c\}$; $A_1^c = \{d\}$.

Beachte

1. Das Ereignis $A_1 \cup A_2$ tritt genau dann ein, wenn A_1 oder A_2 oder beide eintreten.
2. Das Ereignis $A_1 \cap A_2$ tritt genau dann ein, wenn A_1 und A_2 eintreten.
3. Das Ereignis $A_1 \setminus A_2$ tritt genau dann ein, wenn A_1 eintritt und A_2 nicht eintritt.

Lemma 2.4 Für beliebige Mengen $A_1, A_2 \subset \Omega$ gilt: $A_1 \setminus A_2 = A_1 \cap A_2^c$.

Beweis klar

Definition 2.5 Die Mengen $A_1, A_2, \dots \subset \Omega$ heißen *paarweise disjunkt*, falls $A_i \cap A_j = \emptyset$ für beliebige $i \neq j$.

Bemerkung Jede beliebige Folge von Mengen $A_1, A_2, \dots \subset \Omega$ kann man in eine Folge von paarweise disjunkten Mengen A'_1, A'_2, \dots überführen: Sei $A'_1 = A_1$, $A'_2 = A_2 \setminus A'_1$, $A'_3 = A_3 \setminus (A'_1 \cup A'_2)$, ...

Dann gilt: $A'_i \cap A'_j = \emptyset$ für $i \neq j$, und $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} A'_n$

Weitere Eigenschaften Sei $A, B, C \subset \Omega$.

1. *Eindeutigkeitsgesetze*: $A \cup \emptyset = A$, $A \cap \emptyset = \emptyset$, $A \cup \Omega = \Omega$, $A \cap \Omega = A$
 (allgemein: falls $A \subset B$, dann gilt $A \cap B = A$, $A \cup B = B$)
2. *de Morgansche Gesetze*: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$
3. *Assoziativ-Gesetze*: $A \cup (B \cup C) = (A \cup B) \cup C$, $A \cap (B \cap C) = (A \cap B) \cap C$
4. *Distributiv-Gesetze*: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Es ist oft nicht zweckmäßig, *alle* möglichen Teilmengen von Ω in die Modellierung einzubeziehen, sondern man betrachtet nur die Familie derjenigen Teilmengen von Ω , deren Wahrscheinlichkeiten tatsächlich von Interesse sind. Diese Mengenfamilie soll jedoch abgeschlossen sein bezüglich der Operationen \cup, \cap, \setminus , was durch die folgende Begriffsbildung erreicht wird.

Definition 2.6 Eine nichtleere Familie \mathcal{F} von Teilmengen von Ω heißt *Algebra*, falls

$$(A1) \quad A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$

$$(A2) \quad A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$$

Beispiel $\Omega = \{a, b, c, d\}$, $\mathcal{F}_1 = \{\emptyset, \{a\}, \{b, c, d\}, \Omega\}$ ist eine Algebra, $\mathcal{F}_2 = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ ist dagegen keine Algebra.

Lemma 2.7 Sei \mathcal{F} eine Algebra und $A_1, A_2, \dots, A_n \in \mathcal{F}$. Dann gilt

1. $\emptyset, \Omega \in \mathcal{F}$
2. $A_1 \cap A_2 \in \mathcal{F}$
3. $A_1 \setminus A_2 \in \mathcal{F}$
4. $\bigcup_{i=1}^n A_i \in \mathcal{F}, \bigcap_{i=1}^n A_i \in \mathcal{F}$

Beweis

1. Weil \mathcal{F} nicht leer, gibt es ein $A \in \mathcal{F}, A \subset \Omega$. Also gilt $A^c \in \mathcal{F}$ wegen (A1) bzw. $\underbrace{A \cup A^c}_{=\Omega} \in \mathcal{F}$ wegen (A2) bzw. $\underbrace{\Omega^c}_{=\emptyset} \in \mathcal{F}$ wegen (A1)
2. $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{F}$
3. $A_1 \setminus A_2 = A_1 \cap A_2^c = (A_1^c \cup A_2)^c \in \mathcal{F}$
4. Induktion

Um Grenzwerte bilden zu können, ist es erforderlich, daß das Mengensystem \mathcal{F} nicht nur abgeschlossen ist bezüglich Vereinigung bzw. Durchschnitt von *endlich* vielen Mengen (vgl. Teilaussage 4 von Lemma 2.7), sondern auch bezüglich Vereinigung bzw. Durchschnitt von *abzählbar unendlich* vielen Mengen. Dies wird durch die Hinzunahme der folgenden Bedingung erreicht.

Definition 2.8 Eine Algebra \mathcal{F} heißt σ -Algebra, falls zusätzlich

$$(A3) \quad A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \text{ gilt.}$$

Beachte

1. Das Paar (Ω, \mathcal{F}) heißt *Meßraum*, falls \mathcal{F} eine σ -Algebra ist.
2. Sei $\Omega = \mathbb{N}$, und sei \mathcal{F} die Familie derjenigen Teilmengen A von \mathbb{N} , so daß entweder A oder A^c nur endlich viele Elemente hat. Das Mengensystem \mathcal{F} ist eine Algebra, jedoch keine σ -Algebra.
3. Für jedes Ω ist die Potenzmenge \mathcal{P} , d.h. die Familie *aller* Teilmengen von Ω , stets eine σ -Algebra.
4. Falls Ω endlich oder abzählbar unendlich ist, kann $\mathcal{F} = \mathcal{P}$ gewählt werden. Falls Ω nicht abzählbar ist (z.B. $\Omega = \mathbb{R}$ oder $\Omega = [0, 1]$), dann muß eine kleinere σ -Algebra betrachtet werden (*nicht* \mathcal{P}), vgl. Abschnitt 3.1.

2.3 Wahrscheinlichkeitsmaße

Gegeben sei ein Meßraum (Ω, \mathcal{F}) . Betrachten eine Mengenfunktion, d.h. eine Abbildung $P : \mathcal{F} \rightarrow [0, 1]$, die jeder Menge $A \in \mathcal{F}$ eine Zahl $P(A) \in [0, 1]$ zuordnet. Dann heißt $P(A)$ Wahrscheinlichkeit des Ereignisses $A \in \mathcal{F}$.

Definition 2.9 Die Mengenfunktion $P : \mathcal{F} \rightarrow [0, 1]$ heißt *Wahrscheinlichkeitsmaß* auf \mathcal{F} , falls

$$(P1) \quad P(\Omega) = 1 \text{ („Normiertheit“)}$$

$$(P2) \quad P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \text{ für paarweise disjunkte } A_1, A_2, \dots \in \mathcal{F} \text{ („}\sigma\text{-Additivität“)}$$

Falls (Ω, \mathcal{F}) ein Meßraum und P ein Wahrscheinlichkeitsmaß auf \mathcal{F} ist, dann heißt das Tripel (Ω, \mathcal{F}, P) *Wahrscheinlichkeitsraum*.

Theorem 2.10 Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und $A, A_1, A_2, \dots \in \mathcal{F}$. Dann gilt

1. $P(A^c) = 1 - P(A)$
2. $A_1 \subset A_2 \Rightarrow P(A_1) \leq P(A_2)$
3. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
4. $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$

Beweis

1. $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$
2. $P(A_2) = P(A_1 \cup (A_2 \setminus A_1)) = P(A_1) + \underbrace{P(A_2 \setminus A_1)}_{\geq 0} \Rightarrow P(A_2) \geq P(A_1)$
3. Wegen $A_1 \cup A_2 = (A_1 \setminus (A_2 \cap A_1)) \cup (A_1 \cap A_2) \cup (A_2 \setminus (A_1 \cap A_2))$ gilt
 $P(A_1 \cup A_2) = P(A_1 \setminus (A_2 \cap A_1)) + P(A_1 \cap A_2) + P(A_2 \setminus (A_1 \cap A_2))$.
 Weil $P(A \setminus B) = P(A) - P(B)$ für $A, B \in \mathcal{F}$ mit $A \supset B$ gilt, folgt hieraus die Behauptung.
4. Übungsaufgabe

2.4 Endliche Wahrscheinlichkeitsräume

2.4.1 Laplacescher Wahrscheinlichkeitsraum

Für jedes $A \subset \Omega$ bezeichne $|A|$ die Anzahl der Elemente, die zu A gehören. Es gelte $|\Omega| < \infty$ (und damit auch $|A| < \infty$ für jedes $A \subset \Omega$). Ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit $|\Omega| < \infty$ heißt *endlicher Wahrscheinlichkeitsraum*.

Definition 2.11 Ein endlicher Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$, bei dem alle Elementarereignisse *gleichwahrscheinlich* sind, d.h. $P(\{\omega\}) = \frac{1}{|\Omega|}, \forall \omega \in \Omega$, heißt *Laplacescher Wahrscheinlichkeitsraum*.

Beachte Sei $(\Omega, \mathcal{P}(\Omega), P)$ ein Laplacescher Wahrscheinlichkeitsraum. Für jedes $A \subset \Omega$ gilt dann $P(A) = |A|/|\Omega|$ wegen der σ -Additivität von Wahrscheinlichkeitsmaßen. Die so gegebene Wahrscheinlichkeit $P(A) = |A|/|\Omega|$ heißt *Laplacesche Wahrscheinlichkeit*.

Beispiel (zweimaliges Würfeln)

$\omega = (i, j)$ ($i =$ Augenzahl beim 1. Wurf; $j =$ Augenzahl beim 2. Wurf)

$\Omega = \{(i, j) : 1 \leq i, k \leq 6\}; |\Omega| = 36; \mathcal{F} = \mathcal{P}(\Omega)$.

Sei $p : \Omega \rightarrow [0, 1]$ mit $p(\omega_1) = p(\omega_2) \forall \omega_1, \omega_2 \in \Omega$ und $\sum_{\omega \in \Omega} p(\omega) = 1$. Hieraus folgt, daß $p(\omega) = \frac{1}{36} \forall \omega \in \Omega$.

Für ein beliebiges Ereignis $A \subset \Omega$ definieren wir $P(A) = \sum_{\omega \in A} p(\omega)$. D.h. $P(A) = |A|/|\Omega|$.

Sei beispielsweise $A = \{\text{Gesamtaugenzahl} \geq 10\} = \{(6, 6), (6, 5), (5, 6), (6, 4), (4, 6), (5, 5)\}$.

Dann gilt $|A| = 6$ und somit $P(A) = \frac{6}{36} = \frac{1}{6}$.

2.4.2 Einfache Urnenmodelle

Gegeben sei eine Urne mit N Elementen, die mit den Zahlen $1, 2, \dots, N$ numeriert werden. Aus dieser Urne werden n Elemente „zufällig“ entnommen. Ergebnis des gesamten Losvorganges ist ein n -Tupel (i_1, \dots, i_n) . Dabei gibt i_j die Nummer des Elementes an, das bei der j -ten Ziehung entnommen wird. Wir betrachten vier verschiedene Arten von Losvorgängen, die sich durch Kombination der folgenden Vorgehensweisen ergeben:

- *mit* Zurücklegen (d.h., Mehrfachziehungen sind möglich)
- *ohne* Zurücklegen (d.h., jedes Element kann maximal einmal gezogen werden)

- mit Reihenfolge (d.h. $(1, 1, 4, 2) \neq (1, 2, 4, 1)$)
- ohne Reihenfolge (d.h. $(1, 1, 4, 2) = (1, 2, 4, 1)$)

Sei $D = \{1, 2, \dots, N\}$ die Menge der Elemente, die sich zu Beginn des Zufallsexperimentes in der Urne befinden. Die Grundmengen $\Omega_I - \Omega_{IV}$, die die vier verschiedenen Arten von Losvorgängen modellieren, haben die folgende Gestalt:

1. Stichprobe mit Reihenfolge und mit Zurücklegen
 $\Omega_I = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D \text{ für } i = 1, \dots, n\} = D^n$
 $|\Omega_I| = N^n$
2. Stichprobe mit Reihenfolge und ohne Zurücklegen
 $\Omega_{II} = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in D, \omega_i \neq \omega_j \text{ für } i \neq j\}$
 Stufe 1: N Möglichkeiten
 Stufe 2: $N - 1$ Möglichkeiten
 \vdots
 Stufe n : $(N - n + 1)$ Möglichkeiten
 Also: $|\Omega_{II}| = N \cdot (N - 1) \cdot \dots \cdot (N - n + 1) = \frac{N!}{(N-n)!}$
 Wichtiger Spezialfall: $n = N$ (Permutationen) $\Rightarrow |\Omega_{II}| = N!$
3. Stichprobe ohne Reihenfolge und ohne Zurücklegen
 $\Omega_{III} = \{\omega = \{\omega_1, \dots, \omega_n\}, \omega_i \in D, \omega_1 < \omega_2 < \dots < \omega_n\}$
 Also: $|\Omega_{III}| = \frac{N!}{(N-n)! \cdot n!} = \binom{N}{n}$ (Binomialkoeffizient)
4. Stichprobe ohne Reihenfolge und mit Zurücklegen
 $\Omega_{IV} = \{\omega = \{\omega_1, \dots, \omega_n\}, \omega_i \in D, \omega_1 \leq \omega_2 \leq \dots \leq \omega_n\}$
 $|\Omega_{IV}| = \binom{N+n-1}{n} = \frac{(N+n-1)!}{n! \cdot (N-1)!}$

Zusammenfassung

Stichprobe vom Umfang n aus $\{1, 2, \dots, N\}$	mit Zurücklegen	ohne Zurücklegen	
mit Reihenfolge	$ \Omega_I = N^n$	$ \Omega_{II} = \frac{N!}{(N-n)!}$	unterscheidbare Marken
ohne Reihenfolge	$ \Omega_{IV} = \binom{N+n-1}{n}$	$ \Omega_{III} = \binom{N}{n}$	ununterscheidbare Marken
	mit Mehrfachbelegung	ohne Mehrfachbelegung	Verteilung von n Marken auf N Zellen

Beispiele

1. Von den 16 Mannschaften, die am UEFA-Cup eines bestimmten Jahrganges teilnehmen, seien 2 Mannschaften aus Deutschland. Wie groß ist die Wahrscheinlichkeit einer rein deutschen Paarung in der ersten Runde?
Lösung: $N = 16, n = 2$ (mit Reihenfolge und ohne Zurücklegen) $8 \cdot 2 \cdot 14! / 16! = 1/15$.
2. Angenommen: 4 identische Würfel werden gleichzeitig geworfen. Wie groß ist die Wahrscheinlichkeit, daß die vier Augenzahlen voneinander verschieden sind?

Erste Lösungsidee: Stichprobe ohne Reihenfolge und mit Zurücklegen. $\Rightarrow |\Omega_{IV}| = \binom{6+4-1}{4} = 126$.

Problem: keine Chancengleichheit der Elementarereignisse, denn $\{1, 1, 1, 1\}$ ist $(4!)$ -mal weniger wahrscheinlich als $\{1, 2, 3, 4\}$ (wegen der Permutationen).

besser: Wir nehmen an, daß wir die 4 Würfel *nacheinander* werfen und dabei auf die Reihenfolge der erzielten Augenzahlen achten.

\Rightarrow Laplacescher Wahrscheinlichkeitsraum.

$|\Omega_I| = 6^4$ (Anzahl der möglichen Fälle), $A \subset \Omega_I$, $A = \{\text{Stichproben, bestehend aus 4 unterschiedlichen Augenzahlen}\}$, $|A| = 6 \cdot 5 \cdot 4 \cdot 3$ (Anzahl der günstigen Fälle)

$$P(A) = \frac{|A|}{|\Omega_I|} = \frac{6 \cdot 5 \cdot 4 \cdot 3}{6^4} = \frac{5}{18}$$

3. *Zahlenlotto:* $n = 6$ aus $N = 49$ (ohne Reihenfolge und ohne Zurücklegen)

Wie groß ist die Wahrscheinlichkeit, mindestens 4 Richtige zu haben?

Lösung: $D = \{1, \dots, 49\}$, $|\Omega_{III}| = \binom{49}{6}$, $\Omega_{III} = \{\omega = \{\omega_1, \dots, \omega_6\}, \omega_i \in D, \omega_1 < \dots < \omega_6\}$

$A_i = \{\text{genau } i \text{ Richtige}\}$, $P(A_4 \cup A_5 \cup A_6) = ?$

Weil $A_i \cap A_j = \emptyset \forall i \neq j$, gilt $P(A_4 \cup A_5 \cup A_6) = P(A_4) + P(A_5) + P(A_6)$.

Dabei ist $|A_4| = \binom{6}{4} \binom{43}{2}$, $|A_5| = \binom{6}{5} \binom{43}{1}$, $|A_6| = \binom{6}{6} \binom{43}{0} = 1$

$$\Rightarrow P(A) = \frac{\binom{6}{4} \binom{43}{2} + \binom{6}{5} \binom{43}{1} + 1}{\binom{49}{6}} = 0.000987$$

Beachte Dieses Beispiel ist ein Spezialfall der *hypergeometrischen Verteilung*.

Hypergeometrische Verteilung Betrachten Urne mit N Elementen, wobei zwei Typen von Elementen vorhanden seien (S schwarze Kugeln, R rote Kugeln); $N = S + R$. Sei n =Anzahl der insgesamt entnommenen Kugeln; s =Anzahl der entnommenen schwarzen Kugeln; $P(s, n; S, N)$ =Wahrscheinlichkeit, daß s schwarze Kugeln bei der Entnahme von insgesamt n Kugeln gezogen werden, falls von den N vorhandenen Kugeln S schwarz sind.

$$\Rightarrow P(s, n; S, N) = \frac{\binom{S}{s} \binom{N-S}{n-s}}{\binom{N}{n}}$$

3 Zufallsvariable und ihre Charakteristiken

3.1 Zufallsvariable

Betrachten einen beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) und ein beliebiges Element $\omega \in \Omega$, wobei wir so wie bisher $\{\omega\}$ als Elementarereignis bzw. Versuchsergebnis interpretieren.

Beachte Häufig interessiert nicht ω selbst, sondern eine (quantitative oder qualitative) Kennzahl $X(\omega)$ von ω , d.h., wir betrachten die Abbildung $\omega \rightarrow X(\omega)$.

Beispiele

1. $\Omega =$ Menge von Eintragungen in einem Telefonbuch
 $\omega =$ Familienname, $X(\omega) =$ Anzahl der Buchstaben von ω
oder
 $\omega =$ Telefonnummer, $X(\omega) =$ Anzahl der Ziffer „1“ in ω
2. zweimaliges Würfeln $\Omega = \{\omega = (\omega_1; \omega_2), \omega_i \in \{1, \dots, 6\}\}$
Augensumme $\omega \rightarrow X(\omega) = \omega_1 + \omega_2$
Sei $A = \{\omega : X(\omega) = 10\} = \{(6, 4), (5, 5), (4, 6)\}$ bzw. allgemeiner $A = \{\omega : X(\omega) = k\}$, wobei $k \in \{2, \dots, 12\}$.
Gesucht ist die Wahrscheinlichkeit $P(A)$. Hierfür ist es erforderlich, daß $A \in \mathcal{F}$.
Allgemein muß also $\{\omega : \omega \in \Omega, X(\omega) = k\} \in \mathcal{F}$ für jedes $k = 2, \dots, 12$ gelten.
Bei diesem Beispiel ist das gleichbedeutend mit $\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathcal{F}$ für jedes $x \in \mathbb{R}$.

Das führt zu der folgenden Begriffsbildung.

Definition 3.1 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum. Die Abbildung $X : \Omega \rightarrow \mathbb{R}$ heißt *Zufallsvariable* (bzw. *Zufallsgröße*), falls

$$\{\omega : \omega \in \Omega, X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}. \quad (1)$$

Beachte

1. Die Regularitätsbedingung (1) wird *Meßbarkeit* der Abbildung X bezüglich der σ -Algebra \mathcal{F} genannt.
2. In vielen Fällen interessiert nicht nur die Wahrscheinlichkeit, daß die Werte $X(\omega)$ der Zufallsvariable X einen vorgegebenen *Schwellenwert* x nicht überschreiten, d.h., daß X Werte im Intervall $B = (-\infty, x]$ annimmt.
3. Oftmals interessiert auch die Wahrscheinlichkeit, daß X Werte in einer allgemeineren Teilmenge $B \subset \mathbb{R}$ annimmt, wobei B beispielsweise die Vereinigung von mehreren disjunkten Intervallen sein kann.
4. Deshalb wird nicht nur im Grundraum Ω , sondern auch im *Bildraum* \mathbb{R} ein System von Teilmengen von \mathbb{R} betrachtet, das abgeschlossen bezüglich der Mengenoperationen \cup, \cap, \setminus ist, d.h. eine σ -Algebra ist.
5. Dabei wird oft die sogenannte *Borel- σ -Algebra* $\mathcal{B}(\mathbb{R})$ betrachtet, die definiert ist als die kleinste σ -Algebra von Teilmengen von \mathbb{R} , die alle offenen Intervalle (a, b) enthält; $-\infty < a < b < \infty$. D.h.

$$\mathcal{B}(\mathbb{R}) = \sigma\left(\underbrace{\{(a, b), -\infty < a < b < \infty\}}_{\text{Erzeugersystem}}\right)$$

6. Insbesondere enthält $\mathcal{B}(\mathbb{R})$ auch alle halboffenen bzw. abgeschlossenen Intervalle, denn es gilt

$$(a, b] = \bigcap_{n=1}^{\infty} (a, b + n^{-1}) \in \mathcal{B}(\mathbb{R}), \quad [a, b) = \bigcap_{n=1}^{\infty} (a - n^{-1}, b) \in \mathcal{B}(\mathbb{R}), \quad [a, b] = \bigcap_{n=1}^{\infty} (a - n^{-1}, b + n^{-1}) \in \mathcal{B}(\mathbb{R}).$$

7. Für jede abzählbare Teilmenge $C = \{x_1, x_2, \dots\}$ von \mathbb{R} gilt $C \in \mathcal{B}(\mathbb{R})$, denn für jedes $x \in \mathbb{R}$ gilt $\{x\} = \bigcap_{n=1}^{\infty} (x - n^{-1}, x + n^{-1}) \in \mathcal{B}(\mathbb{R})$ und damit auch $C = \bigcup_{i=1}^{\infty} \{x_i\} \in \mathcal{B}(\mathbb{R})$.

3.2 Verteilung von Zufallsvariablen

Die Regularitätsbedingung (1) in Definition 3.1 kann durch die folgende (scheinbar schärfere, in Wirklichkeit jedoch äquivalente) Bedingung ersetzt werden: Für jede Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ gilt

$$\{\omega : \omega \in \Omega, X(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}).$$

Dies führt zu der folgenden Begriffsbildung.

Definition 3.2 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum, und $X : \Omega \rightarrow \mathbb{R}$ sei eine beliebige Zufallsvariable. Die *Verteilung* der Zufallsvariable X ist die Mengenfunktion $P_X : \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$ mit

$$P_X(B) = P(\omega : \omega \in \Omega, X(\omega) \in B) \quad \forall B \in \mathcal{B}(\mathbb{R}). \quad (2)$$

Beachte

1. Die in (2) definierte Mengenfunktion P_X ist ein Wahrscheinlichkeitsmaß auf dem Meßraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
2. Die Abbildung $P \rightarrow P_X$ nennt man *Maßtransport* vom Meßraum (Ω, \mathcal{F}) in den Meßraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Die folgende Kurzschreibweise ist üblich: $P(X \in B) = P(\{\omega : \omega \in \Omega, X(\omega) \in B\}) \quad \forall B \in \mathcal{B}(\mathbb{R})$

Speziell: $P(X \leq x) = P(\{\omega : \omega \in \Omega, X(\omega) \leq x\}) \quad \forall x \in \mathbb{R}$

3.2.1 Diskrete Zufallsvariable; Wahrscheinlichkeitsfunktion

Wir unterscheiden 2 (Grund-) Typen von Zufallsvariablen: *diskrete* und *absolutstetige* Zufallsvariable.

Definition 3.3 Die Zufallsvariable X (bzw. ihre Verteilung) heißt *diskret*, falls es eine abzählbare Teilmenge $C \subset \mathbb{R}$ gibt, so daß $P(X \in C) = 1$.

Beachte Der Begriff der absolutstetigen Zufallsvariable wird später in Definition 3.6 eingeführt. Wir erwähnen jedoch schon jetzt ein wichtiges Unterscheidungsmerkmal:

1. diskrete Zufallsvariable haben einen abzählbaren Wertebereich, z.B. $\mathbb{N}, \mathbb{Z}, \mathbb{Q} \subset \mathbb{R}$
z.B. sei $X =$ Augensumme bei zweimaligem Würfeln $\Rightarrow X : \Omega \rightarrow \{2, 3, \dots, 12\}$
2. absolutstetige Zufallsvariable haben einen überabzählbaren Wertebereich, z.B. $[a, b], [a, \infty), (-\infty, b], \mathbb{R}$
z.B. Roulette mit drehbarem Zeiger und „kontinuierlicher“ Skala, wobei
 $X =$ Wert des Spiels = Winkel des Zeigers, d.h. $X : \Omega \rightarrow [0, 2\pi)$

Definition 3.4 Sei X eine diskrete Zufallsvariable, d.h., es gebe eine abzählbare Menge $C = \{x_1, x_2, \dots\}$, so daß $P(X \in C) = 1$. Dann heißt die Folge p_1, p_2, \dots mit $p_k = P(X = x_k)$ *Wahrscheinlichkeitsfunktion* (bzw. *Zähldichte*) von X .

Beachte

1. Für jede Wahrscheinlichkeitsfunktion $\{p_k\}$ gilt $p_k \geq 0$ für jedes $k = 1, 2, \dots$ und $\sum_{k=1}^{\infty} p_k = 1$.
2. Die Verteilung einer diskreten Zufallsvariablen X wird eindeutig durch die Wahrscheinlichkeitsfunktion $\{p_k\}$ bestimmt. Für jedes $x_k \in C$ heißt die Zahl $p_k = P(X = x_k)$ *Einzelwahrscheinlichkeit* von X .

Beispiele

1. *zweimaliges Würfeln*
 - Sei $X =$ Summe der Augenzahlen beim zweimaligen Würfeln.

- Dann gilt $P(X \in C) = 1$ mit $C = \{2, 3, \dots, 12\}$, d.h. $x_k = k + 1$.
- Die Wahrscheinlichkeitsfunktion von X ist gegeben durch:

x_k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

2. Bernoulli-Schema

- Einmaliger Münzwurf: $\Omega_1 = \{0, 1\}$, „0“ = Wappen, „1“ = Zahl
- n -maliger Münzwurf: Für $i = 1, \dots, n$ setzen wir $\Omega_i = \{0, 1\}$, wobei $P_i(\{\omega_i\}) = \frac{1}{2}$ im Fall einer fairen Münze bzw. allgemein $P_i(\{1\}) = a_i$ und $P_i(\{0\}) = 1 - a_i$.
- Bei *identischen Versuchsbedingungen* nehmen wir an, daß $a_1 = a_2 = \dots = a_n = p$, $p \in [0, 1]$.
- *Unabhängige Versuche* modellieren wir durch den Ansatz (Ω, \mathcal{F}, P) mit

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}, \quad \mathcal{F} = \mathcal{P}(\Omega),$$

wobei P das sogenannte *Produktmaß* auf \mathcal{F} ist, für das gilt: $P(\{\omega\}) = \prod_{i=1}^n P_i(\{\omega_i\})$.

- Sei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ ein beliebiges Elementarereignis. Dann gilt

$$P(\{\omega\}) = \prod_{i:\omega_i=1} a_i \prod_{j:\omega_j=0} (1 - a_j)$$

- Für $\omega \in \Omega$ mit $|\{i : \omega_i = 1\}| = k$ und $a_1 = a_2 = \dots = a_n = p$ gilt dann insbesondere

$$P(\{\omega\}) = p^k (1 - p)^{n-k}.$$

- Deuten „1“ als Erfolg und „0“ als Mißerfolg.
- Sei $X =$ Anzahl der Erfolge bei n Versuchen. Falls $a_1 = a_2 = \dots = a_n = p$, dann gilt

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

- Sprechweise: X ist *binomialverteilt* mit den Parametern n und p .
- Spezialfall: Falls $n = 1$, dann sagen wir, daß X *Bernoulli-verteilt* ist.

3.2.2 Grundlegende Klassen diskreter Verteilungen (Zusammenfassung)

1. Hypergeometrische Verteilung (Urnenmodell)

- Betrachten Urne mit N Elementen (S schwarze, R rote Kugeln), d.h. $N = S + R$;
- n Elemente, $n \leq N$, sollen insgesamt ausgewählt werden;
- $X =$ zufällige Anzahl von schwarzen Kugeln bei n Entnahmen

$$X : \Omega \rightarrow \{0, 1, \dots, \min\{n, S\}\}, \quad p_k = P(X = k) = \frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}}, \quad k = 1, \dots, \min\{S, n\}$$

- 3 Parameter: $N, S \leq N, n \leq N$;
- Berechnung der Einzelwahrscheinlichkeiten p_k ist *schwierig*, falls N und S groß sind;
- *Ausweg*: Binomialverteilung liefert Näherungsformel, falls $N, S \rightarrow \infty, \frac{S}{N} \xrightarrow{S, N \rightarrow \infty} p$

2. Binomialverteilung (Bernoulli-Schema)

- n -maliger Münzwurf;
- identische Erfolgswahrscheinlichkeiten $a_1 = \dots = a_n = p$;
- $X =$ Anzahl der Erfolge bei n Versuchen

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

- Für jedes $k = 0, 1, \dots, n$ gilt

$$\frac{\binom{S}{k} \binom{N-S}{n-k}}{\binom{N}{n}} \rightarrow \binom{n}{k} p^k (1-p)^{n-k},$$

falls $N, S \rightarrow \infty, \frac{S}{N} \xrightarrow{S, N \rightarrow \infty} p$.

- 2 Parameter: n und p (Schreibweise: $\text{Bin}(n, p)$);
- Berechnung der Einzelwahrscheinlichkeiten p_k ist *schwierig*, falls n groß und p klein ist;
- *Ausweg*: Poisson-Verteilung liefert Näherungsformel, falls $n \rightarrow \infty$ und $p \rightarrow 0$ mit $n \cdot p \rightarrow \lambda$, so daß $0 < \lambda < \infty$.

3. Poisson-Verteilung (Gesetz der seltenen Ereignisse)

- Betrachten diskrete Zufallsvariable $X : \Omega \rightarrow \{0, 1, \dots\}$;
- die Einzelwahrscheinlichkeiten $p_k = P(X = k)$ seien gegeben durch

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots;$$

- 1 Parameter: λ (Schreibweise: $\text{Poi}(\lambda)$);
- *Gesetz der seltenen Ereignisse*: Für jedes $k = 0, 1, \dots$ gilt

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda},$$

falls $n \rightarrow \infty$ und $p \rightarrow 0$ mit $n \cdot p \rightarrow \lambda$.

4. geometrische Verteilung

- Sei $X =$ die Anzahl der Versuche bis zum ersten Erfolg im Bernoulli-Schema (mit $n = \infty$), d.h. $X : \Omega \rightarrow \{1, 2, \dots\}$;
- die Einzelwahrscheinlichkeiten $p_k = P(X = k)$ sind dann gegeben durch

$$p_k = p(1-p)^{k-1}, \quad k = 1, 2, \dots;$$

- 1 Parameter: $p \in [0, 1]$.

3.2.3 Verteilungsfunktion; absolutstetige Zufallsvariable

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

Definition 3.5 Die Funktion $F_X : \mathbb{R} \rightarrow [0, 1]$ mit $F_X(x) = P(X \leq x)$ heißt *Verteilungsfunktion* von X .

Eigenschaften von Verteilungsfunktionen

1. *Asymptotisches Verhalten im Unendlichen*: $F_X(-\infty) = \lim_{x \rightarrow -\infty} F_X(x) = 0, \quad F_X(\infty) = \lim_{x \rightarrow \infty} F_X(x) = 1$

2. *Monotonie:* $F_X(x) \leq F_X(x+h) \quad \forall x \in \mathbb{R} \text{ und } h \geq 0$
3. *Rechtsstetigkeit:* $F_X(x)$ ist rechtsseitig stetig, d.h. für jede Folge $\{h_n\}$ mit $h_n \geq 0$ und $\lim_{n \rightarrow \infty} h_n = 0$ gilt

$$\lim_{n \rightarrow \infty} F_X(x+h_n) = F_X(x) \quad \forall x \in \mathbb{R}.$$

Beachte

1. Mit Hilfe der Verteilungsfunktion F_X lassen sich auch die folgenden Wahrscheinlichkeiten ausdrücken

$$P(a \leq X \leq b), \quad P(a < X \leq b), \quad P(a < X < b), \quad P(a \leq X < b),$$

denn es gilt beispielsweise

$$\begin{aligned} P(a \leq X \leq b) &= P(\{X \leq b\} \setminus \{X < a\}) \\ &= P(X \leq b) - P(X < a) \\ &= F_X(b) - \lim_{h \downarrow 0} F_X(a-h). \end{aligned}$$

2. Im allgemeinen gilt jedoch nicht $F(a) = \lim_{h \downarrow 0} F_X(a-h)$.
3. Für die Verteilungsfunktion F_X einer diskreten Zufallsvariable X gilt für jedes $x \in \mathbb{R}$:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\bigcup_{k: x_k \leq x} \{X = x_k\}\right) \\ &= \sum_{k: x_k \leq x} P(X = x_k) \\ &= \sum_{k: x_k \leq x} p_k, \end{aligned}$$

wobei $p_k = P(X = x_k)$.

4. Die Verteilungsfunktion F_X einer diskreten Zufallsvariablen X ist eine sogenannte *Treppenfunktion*, d.h. eine stückweise konstante Funktion mit der Sprunghöhe p_k im Punkt x_k .

Definition 3.6 Die Zufallsvariable $X : \Omega \rightarrow \mathbb{R}$ (bzw. ihre Verteilung) heißt *absolutstetig*, falls die Verteilungsfunktion F_X von X die folgende *Integraldarstellung*

$$F_X(x) = \int_{-\infty}^x f_X(y) dy \quad \forall x \in \mathbb{R} \quad (3)$$

besitzt, wobei $f_X : \mathbb{R} \rightarrow [0, \infty)$ eine (integrierbare) Funktion mit nichtnegativen Werten ist, die *Dichte* (bzw. *Wahrscheinlichkeitsdichte*) von X genannt wird.

Beachte

1. Die Verteilungsfunktion F_X (und damit auch die Verteilung P_X) einer absolutstetigen Zufallsgröße X wird *eindeutig* durch die Dichte f_X bestimmt.
2. Bei vielen Anwendungen ist die Dichte f_X eine (zumindest stückweise) stetige Funktion. Das Integral in der Definitionsgleichung (3) ist dann ein gewöhnliches Riemann-Integral.
3. Falls X absolutstetig ist, dann hat die Verteilungsfunktion F_X keine Sprünge, d.h., F_X ist eine (im üblichen Sinne) stetige Funktion. Hieraus folgt insbesondere, daß

$$P(X = x) = 0 \quad \forall x \in \mathbb{R}. \quad (4)$$

Beweis von (4). Es gilt

$$\begin{aligned} P(X = x) &= \lim_{h \downarrow 0} P(x - h < X \leq x + h) = \lim_{h \downarrow 0} (P(X \leq x + h) - P(X \leq x - h)) \\ &= \lim_{h \downarrow 0} (F_X(x + h) - F_X(x - h)) = \lim_{h \downarrow 0} F_X(x + h) - \lim_{h \downarrow 0} F_X(x - h) \\ &= F_X(x) - F_X(x) = 0. \end{aligned}$$

4. Die Verteilungsfunktion F_X einer absolutstetigen Zufallsvariablen X ist jedoch im allgemeinen *nicht* überall differenzierbar. Und zwar ist F_X dort nicht differenzierbar, wo die Dichte f_X Sprungstellen hat.

Beispiele Um die Verteilung einer absolutstetigen Zufallsvariable X zu beschreiben, genügt es, die Dichte f_X zu betrachten, weil durch f_X die Verteilungsfunktion F_X und damit auch die Verteilung P_X von X eindeutig bestimmt wird.

1. *Normalverteilung* $N(\mu, \sigma^2)$ mit den Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \forall x \in \mathbb{R} \quad (5)$$

Spezialfall: *Standardnormalverteilung* $N(0, 1)$. Dann nimmt die Dichte $f_X(x)$ in (5) die folgende Form an:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \forall x \in \mathbb{R} \quad (6)$$

2. *Exponentialverteilung* $\text{Exp}(\lambda)$ mit Parameter $\lambda > 0$:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{falls } x \geq 0 \\ 0, & \text{falls } x < 0 \end{cases}$$

3. *Gleichverteilung* $U(a, b)$ mit den Parametern $a, b \in \mathbb{R}$, wobei $a < b$:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$$

Beachte

- Absolutstetige Verteilungen treten oft als (asymptotische) Näherungslösungen beim Übergang zu Grenzwerten auf.
- So läßt sich beispielsweise die Binomialverteilung mit den Parametern n und p durch die Standardnormalverteilung approximieren, falls n groß ist. Denn für beliebige $p \in (0, 1)$ und $a, b \in \mathbb{R}$ mit $a < b$ gilt

$$\lim_{n \rightarrow \infty} P\left(a < \frac{X_n - np}{\sqrt{np(1-p)}} \leq b\right) = P(a < X \leq b), \quad (7)$$

wobei X_n binomialverteilt (mit den Parametern n und p) und X standardnormalverteilt ist.

- Dies ist der sogenannte *zentrale Grenzwertsatz von DeMoivre-Laplace*. Weitere Grenzwertsätze werden in Abschnitt 4 der Vorlesung diskutiert.
- Wenn für die Schranken a und b in (7)

$$a = \frac{k-1-np}{\sqrt{np(1-p)}} \quad \text{und} \quad b = \frac{k-np}{\sqrt{np(1-p)}}$$

eingesetzt wird, dann ergibt sich aus (7), daß für große n

$$\begin{aligned} P(X_n = k) &\approx P\left(\frac{k-1-np}{\sqrt{np(1-p)}} < X \leq \frac{k-np}{\sqrt{np(1-p)}}\right) \\ &\approx \frac{1}{\sqrt{np(1-p)}} f_X\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

für jedes $k = 0, 1, \dots, n$, wobei f_X die in (6) gegebene Dichte der Standardnormalverteilung ist.

- Die Näherungsformel

$$P(X_n = k) \approx \frac{1}{\sqrt{np(1-p)}} f_X\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \quad (8)$$

wird im Internet durch zahlreiche *JAVA-Applets* illustriert, vgl. beispielsweise

<http://medweb.uni-muenster.de/institute/imib/lehre/skripte/biomaethe/binorm.html>

wobei in diesem Applet die rechte Seite von (8) durch eine blaue Kurve dargestellt wird (und das Symbol N anstelle von n benutzt wird).

3.3 Zufallsvektoren

3.3.1 Definition

Bei Anwendungen besteht oft die Notwendigkeit, nicht nur eine Kennzahl $X(\omega)$, sondern gleichzeitig mehrere Kennzahlen $X_1(\omega), \dots, X_n(\omega)$ von $\omega \in \Omega$ zu betrachten.

Beispiele

1. zweimaliges Würfeln

- Als Grundraum wählen wir so wie bisher $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$, vgl. Abschnitt 2.4.1.
- Sei $X : \Omega \rightarrow \{0, 1, 2\}$ bzw. $Y : \Omega \rightarrow \{0, 1, 2\}$ die (zufällige) Anzahl, mit der die Augenzahl „6“ bzw. „1“ beim zweimaligen Würfeln erzielt wird.
- Dann gilt für die Wahrscheinlichkeiten $P(X = x, Y = y)$ bzw. für die Einzelwahrscheinlichkeiten $P(X = x)$ und $P(Y = y)$ von X und Y :

$P(X = x, Y = y)$		y			$P(X = x)$
		0	1	2	
x	0	$\frac{16}{36}$	$\frac{8}{36}$	$\frac{1}{36}$	$\frac{25}{36}$
	1	$\frac{8}{36}$	$\frac{2}{36}$	0	$\frac{10}{36}$
	2	$\frac{1}{36}$	0	0	$\frac{1}{36}$
$P(Y = y)$		$\frac{25}{36}$	$\frac{10}{36}$	$\frac{1}{36}$	

- Aus der Tabelle kann man auch die Einzelwahrscheinlichkeiten der Summe $X + Y$ erhalten. Beispielsweise gilt

$$P(X + Y = 1) = P(X = 1, Y = 0) + P(X = 0, Y = 1) = \frac{8}{36} + \frac{8}{36} = \frac{16}{36}.$$

- Analog ergibt sich

$$P(X + Y = 0) = \frac{16}{36}, \quad P(X + Y = 2) = \frac{4}{36}$$

bzw.

$$P(X \cdot Y = 1) = \frac{2}{36}, \quad P(X \cdot Y = 0) = \frac{34}{36}.$$

2. Analyse von Kommunikationsnetzen

- Betrachten ein Kommunikationsnetz mit n Komponenten.
- Sei $\Omega = \Omega_1 \times \dots \times \Omega_n$, wobei Ω die Menge aller möglichen Momentanzustände $\omega = (\omega_1, \dots, \omega_n)$ des Netzes und Ω_i die Menge aller möglichen Momentanzustände ω_i der i -ten Komponente bezeichnet; $i = 1, \dots, n$.
- Dann kann beispielsweise durch die Abbildung $\omega \rightarrow X_i(\omega)$ die Belastung $X_i(\omega)$ der i -ten Komponente in Abhängigkeit vom Momentanzustand ω des Netzes modelliert werden.
- Die (globale) Belastung des gesamten Netzes kann dann durch den Vektor $(X_1(\omega), \dots, X_n(\omega))$ beschrieben werden.

Die gleichzeitige Betrachtung mehrerer Kennzahlen $X_1(\omega), \dots, X_n(\omega)$ von $\omega \in \Omega$ führt zum Begriff des Zufallsvektors.

Definition 3.7 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum, und sei X_1, \dots, X_n eine beliebige Folge von Zufallsvariablen $X_i : \Omega \rightarrow \mathbb{R}$; $i = 1, \dots, n$. Die Abbildung $X = (X_1, \dots, X_n)$ von Ω nach \mathbb{R}^n heißt dann n -dimensionaler *Zufallsvektor* mit den Komponenten X_1, \dots, X_n . Die Funktion $F_X : \mathbb{R}^n \rightarrow [0, 1]$ mit

$$F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (9)$$

heißt (gemeinsame bzw. multivariate) *Verteilungsfunktion* des Zufallsvektors $X = (X_1, \dots, X_n)$.

3.3.2 Eigenschaften multivariater Verteilungsfunktionen

1. *Asymptotisches Verhalten im Unendlichen:* Für beliebige $i \in \{1, \dots, n\}$ und $x_1, \dots, x_n \in \mathbb{R}$ gilt (i) $F_X(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = 0$, wobei

$$F_X(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = \lim_{x_i \rightarrow -\infty} F_X(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n);$$

- (ii) $F_X(\infty, \dots, \infty) = 1$, wobei $F_X(\infty, \dots, \infty) = \lim_{x_1, \dots, x_n \rightarrow \infty} F_X(x_1, \dots, x_n)$;

(iii) $F_X(\infty, \dots, \infty, x_i, \infty, \dots, \infty) = F_{X_i}(x_i)$, wobei $F_X(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$ analog zu den in (i)–(ii) betrachteten Grenzwerten definiert wird und F_{X_i} *Randverteilungsfunktion* von X genannt wird.

2. *Monotonie:* $F_X(x_1 + h_1, \dots, x_n + h_n) \geq F_X(x_1, \dots, x_n) \forall x_1, \dots, x_n \in \mathbb{R}, h_1, \dots, h_n \geq 0$
3. *Rechtsstetigkeit:* $F_X(x_1, \dots, x_n) = \lim_{h_1, \dots, h_n \downarrow 0} F_X(x_1 + h_1, \dots, x_n + h_n) \forall x_1, \dots, x_n \in \mathbb{R}$

Definition 3.8

1. Der Zufallsvektor $X = (X_1, \dots, X_n)$ heißt *diskret*, falls es eine abzählbare Menge $C \subset \mathbb{R}^n$ gibt, so daß $P(X \in C) = 1$.
2. Sei $X = (X_1, \dots, X_n)$ ein diskreter Zufallsvektor. Dann heißt $\{P(X = x), x \in C\}$ *Wahrscheinlichkeitsfunktion* von X .

Beachte Falls $X = (X_1, \dots, X_n)$ ein diskreter Zufallsvektor ist, dann sind auch seine Komponenten X_1, \dots, X_n diskrete Zufallsvariablen. Für die Wahrscheinlichkeitsfunktion $\{P(X_i = x_i), x_i \in C_i\}$ von X_i gilt

$$\begin{aligned} P(X_i = x_i) &= \sum_{y_1 \in C_1} \dots \sum_{y_{i-1} \in C_{i-1}} \sum_{y_{i+1} \in C_{i+1}} \dots \sum_{y_n \in C_n} P(X_1 = y_1, \dots, X_{i-1} = y_{i-1}, X_i = x_i, X_{i+1} = y_{i+1}, \dots, X_n = y_n). \end{aligned}$$

Definition 3.9 Der Zufallsvektor $X = (X_1, \dots, X_n)$ heißt *absolutstetig*, falls es eine (integrierbare) Funktion $f_X : \mathbb{R}^n \rightarrow [0, \infty)$ gibt, so daß

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \dots \int_{-\infty}^{x_1} f_X(y_1, \dots, y_n) dy_1 \dots dy_n \quad \forall x_1, \dots, x_n \in \mathbb{R}. \quad (10)$$

Die Funktion f_X heißt (gemeinsame) *Dichte* von X .

Beachte Falls $X = (X_1, \dots, X_n)$ ein absolutstetiger Zufallsvektor mit der Dichte f_X ist, dann sind auch seine Komponenten X_1, \dots, X_n absolutstetige Zufallsvariablen. Für die Dichte $f_{X_i} : \mathbb{R} \rightarrow [0, \infty)$ von X_i gilt

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{(n-1)\text{-mal}} f_X(y_1, \dots, y_{i-1}, x_i, y_{i+1}, \dots, y_n) dy_1 \dots dy_{i-1} dy_{i+1} \dots dy_n. \quad (11)$$

Definition 3.10 Die in (11) betrachtete Funktion f_{X_i} heißt *Randdichte* von X .

3.3.3 Weitere Beispiele

1. n -maliger Münzwurf

- Betrachten den in Abschnitt 3.2.1 eingeführten Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) mit der Grundmenge

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}$$

- bei identischen Erfolgswahrscheinlichkeiten $a_1 = \dots = a_n = p$;
- Sei $X =$ Anzahl der Erfolge bei n Versuchen. Dann ist X binomialverteilt, d.h.

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

- Außerdem betrachten wir die (zufällige) Nummer Y desjenigen Versuches, bei dem zum ersten Mal ein Erfolg eintritt, d.h.

$$Y(\omega) = \begin{cases} \inf\{i \geq 1 : \omega_i = 1\}, & \text{falls } X(\omega) \geq 1 \\ n+1, & \text{falls } X(\omega) = 0 \end{cases}$$

- Gesucht sind die Einzelwahrscheinlichkeiten des Zufallsvektors (X, Y) :

$$p(x, y) = P(X = x, Y = y) \quad \text{für } 0 \leq x \leq n, 1 \leq y \leq n+1$$

- Offenbar gilt $p(0, n+1) = (1-p)^n$ und $p(0, y) = 0$ für $1 \leq y \leq n$.
- Sei jetzt $X(\omega) = x \geq 1$ und $Y(\omega) = y$, d.h., $\omega_i = 0$ für $i < y$ und $\omega_y = 1$, und es müssen genau $x-1$ Einsen unter $\omega_{y+1}, \dots, \omega_n$ sein.
- Dafür gibt es $\binom{n-y}{x-1}$ Möglichkeiten, falls $x-1 \leq n-y$, und 0 Möglichkeiten, falls $x-1 > n-y$.
- Also ist

$$p(x, y) = \begin{cases} \binom{n-y}{x-1} p^x (1-p)^{n-x}, & \text{falls } x-1 \leq n-y, \\ 0, & \text{falls } x-1 > n-y. \end{cases}$$

- Spezialfall. Sei jetzt $n = 3$ und $p = \frac{1}{2}$. Dann sind die Wahrscheinlichkeiten $P(X = x, Y = y)$ bzw. die (Rand-)Wahrscheinlichkeiten $P(X = x)$ und $P(Y = y)$ in der folgenden Tabelle gegeben:

$P(X = x, Y = y)$		y				$P(X = x)$
		1	2	3	4	
x	0	0	0	0	$\frac{1}{8}$	$\frac{1}{8}$
	1	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
	2	$\frac{2}{8}$	$\frac{1}{8}$	0	0	$\frac{3}{8}$
	3	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$P(Y = y)$		$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	

2. zweidimensionale integrierte Gleichverteilung

- Sei $X = (X_1, X_2)$ ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte

$$f_X(x_1, x_2) = \begin{cases} 4x_1x_2, & \text{falls } 0 \leq x_1, x_2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

- Für die Randdichten gilt dann

$$f_{X_1}(x_1) = \int_0^1 f_X(x_1, x_2) dx_2 = 4x_1 \cdot \frac{1}{2} = 2x_1 \quad \forall x_1 \in [0, 1]$$

und analog

$$f_{X_2}(x_2) = \int_0^1 f_X(x_1, x_2) dx_1 = 2x_2 \quad \forall x_2 \in [0, 1]$$

- Wir wollen nun noch die Wahrscheinlichkeit $P(X_1 \leq 2X_2)$ berechnen. Es gilt

$$\begin{aligned} P(X_1 \leq 2X_2) &= \int_0^{\frac{1}{2}} \int_0^{2x_2} 4x_1x_2 dx_1 dx_2 + \int_{\frac{1}{2}}^1 \int_0^1 4x_1x_2 dx_1 dx_2 \\ &= \int_0^{\frac{1}{2}} 4x_2 \left[\frac{x_1^2}{2} \right]_0^{2x_2} dx_2 + \int_{\frac{1}{2}}^1 4x_2 \left[\frac{x_1^2}{2} \right]_0^1 dx_2 \\ &= 8 \left[\frac{x_2^4}{4} \right]_0^{\frac{1}{2}} + 2 \left[\frac{x_2^2}{2} \right]_{\frac{1}{2}}^1 = \frac{7}{8}. \end{aligned}$$

3.4 Bedingte Wahrscheinlichkeiten

3.4.1 Definition und Multiplikationssatz

Häufig verfügen wir bei der Durchführung von Experimenten über Vorinformationen, die bei der Berechnung von Wahrscheinlichkeiten interessierender Ereignisse berücksichtigt werden sollen.

Bei manchen Untersuchungen wird jedoch lediglich (hypothetisch) angenommen, daß eine bestimmte Vorinformation vorliegt, wobei dann unter dieser hypothetischen Annahme gerechnet wird. Diese sogenannte *Bayessche Methodik* wird im weiteren Verlauf der Vorlesung noch genauer diskutiert.

Beispiele

1. Skatspiel

- Die Kenntnis der eigenen 10 Karten soll als Vorinformation über die Verteilung der übrigen 22 Karten genutzt werden.
- Markieren die 32 Karten mit den Zahlen $1, 2, \dots, 32$.
- Betrachten Laplaceschen Wahrscheinlichkeitsraum, wobei Ω die Menge aller Permutationen von 32 Elementen ist ($N = n = 32$; mit Reihenfolge und ohne Zurücklegen)
- Gesucht sei die Wahrscheinlichkeit des Ereignisses $A_2 \cap A_3$, wobei $A_2 = \{\text{Spieler 2 hat } x \text{ Asse}\}$, $A_3 = \{\text{Spieler 3 hat } y \text{ Asse}\}$, unter der Bedingung, daß das Ereignis $A_1 = \{\text{Spieler 1 hat die Karten mit den Nummern } k_1, \dots, k_{10}\}$ eintritt.
- *Lösungsansatz:* Beziehen die Anzahl der Permutationen, bei denen $A_2 \cap A_3$ eintritt, nicht auf die Gesamtanzahl $32!$ aller möglichen Permutationen, sondern lediglich auf diejenigen Permutationen, bei denen das Ereignis A_1 eintritt.
- D.h., die gesuchte Wahrscheinlichkeit ist die (bedingte) relative Häufigkeit $|(A_2 \cap A_3) \cap A_1|/|A_1|$
- Dabei benutzen wir die Schreibweise:

$$P(A_2 \cap A_3 | A_1) = |(A_2 \cap A_3) \cap A_1|/|A_1|$$

und nennen diese Größe *bedingte Wahrscheinlichkeit* des Ereignisses $A_2 \cap A_3$ unter der Bedingung, daß das Ereignis A_1 eintritt.

2. Urnenmodell

- Betrachten Urne mit N Elementen (S schwarze, R rote Kugeln), d.h. $N = S + R$, vgl. Abschnitt 3.2.2;
- 2 Elemente, $2 \leq N$, sollen insgesamt ausgewählt werden (ohne Zurücklegen);
- Sei A das Ereignis, beim zweiten Versuch „schwarz“ zu ziehen, und sei B das Ereignis, beim ersten Versuch „rot“ zu ziehen.
- Gesucht ist die bedingte Wahrscheinlichkeit $P(A | B)$, beim zweiten Versuch „schwarz“ zu ziehen, falls beim ersten Versuch „rot“ gezogen wird.
- Es gilt

$$P(A | B) = \frac{|A \cap B|}{|B|} = \frac{\frac{R}{N} \frac{S}{N-1}}{\frac{R}{N} \frac{S}{N-1} + \frac{R}{N} \frac{R-1}{N-1}} = \frac{S}{N-1}.$$

Dies führt zu der folgenden (allgemeineren) Begriffsbildung.

Definition 3.11 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum, und $A, B \in \mathcal{F}$ seien beliebige Ereignisse mit $P(B) > 0$. Dann heißt

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad (12)$$

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

Beachte Die Definitionsgleichung (12) kann in der Form $P(A \cap B) = P(B)P(A | B)$ geschrieben werden. Durch Iteration dieser Überlegung ergibt sich die folgende Aussage.

Theorem 3.12 (Multiplikationssatz)

Seien $A_1, A_2, \dots, A_n \in \mathcal{F}$ Ereignisse mit $P(A_1 \cap A_2 \cap \dots \cap A_n) > 0$. Dann gilt:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) \dots P(A_n | A_1 \cap A_2 \cap \dots \cap A_{n-1}). \quad (13)$$

Beispiel (Skatspiel)

- Betrachten das Ereignis $A_i = \{\text{Spieler } i \text{ erhält genau ein As}\}$; $i = 1, 2, 3$.
- Gesucht ist die Wahrscheinlichkeit $P(A_1 \cap A_2 \cap A_3)$, daß jeder der drei Spieler genau ein As erhält?
- *Lösung:* Es gilt

$$P(A_1) = \frac{\binom{4}{1} \binom{28}{9}}{\binom{32}{10}} = 0.428, \quad P(A_2 | A_1) = \frac{\binom{3}{1} \binom{19}{9}}{\binom{22}{10}} = 0.42857, \quad P(A_3 | A_1 \cap A_2) = \frac{\binom{2}{1} \binom{10}{9}}{\binom{12}{10}} = 0.303.$$

- Hieraus und aus (13) ergibt sich

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2 | A_1)P(A_3 | A_1 \cap A_2) = 0.0556.$$

3.4.2 Formel der totalen Wahrscheinlichkeit; Bayessche Formel

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

Bei der Berechnung der Wahrscheinlichkeit $P(A)$ eines Ereignisses $A \in \mathcal{F}$ ist es manchmal nützlich, die (unbedingte) Wahrscheinlichkeit $P(A)$ als gewichtete Summe von bedingten Wahrscheinlichkeiten darzustellen.

Hierfür ist es erforderlich, den Grundraum Ω wie folgt in (meßbare) Teilmengen zu zerlegen.

Definition 3.13 Sei $n \in \mathbb{N}$ eine beliebige natürliche Zahl, und sei $B_1, B_2, \dots, B_n \in \mathcal{F}$ eine (endliche) Folge von Ereignissen mit den Eigenschaften

- (Z1) $B_i \cap B_j = \emptyset$ für $i \neq j$,
- (Z2) $\bigcup_{i=1}^n B_i = \Omega$,
- (Z3) $P(B_i) > 0$ für alle $i = 1, \dots, n$.

Dann heißt B_1, B_2, \dots, B_n *meßbare Zerlegung* von Ω .

Theorem 3.14 Sei $A \in \mathcal{F}$ ein beliebiges Ereignis und B_1, B_2, \dots, B_n eine meßbare Zerlegung von Ω . Dann gilt

1. *Formel der totalen Wahrscheinlichkeit*

$$P(A) = \sum_{j=1}^n P(B_j)P(A | B_j), \quad (14)$$

2. *Bayessche Formel*

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)} \quad (15)$$

für jedes $i = 1, \dots, n$, wobei in (15) vorausgesetzt wird, daß $P(A) > 0$.

Beweis Aus (Z1)–(Z3) und aus der Additivität des Wahrscheinlichkeitsmaßes P ergibt sich, daß

$$\begin{aligned} P(A) &= P(A \cap \Omega) = P\left(A \cap \left(\bigcup_{i=1}^n B_i\right)\right) \\ &= P\left(\bigcup_{i=1}^n (A \cap B_i)\right) = \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(B_i) \frac{P(A \cap B_i)}{P(B_i)} = \sum_{i=1}^n P(B_i)P(A | B_i), \end{aligned}$$

wobei im letzten Schritt die Definitionsgleichung (12) benutzt wird. Damit ist (14) bewiesen. Aus (12) und (14) ergibt sich nun

$$P(B_i | A) = \frac{P(B_i \cap A)}{P(A)} = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^n P(B_j)P(A | B_j)}.$$

Beachte Die Aussagen von Theorem 3.14 bleiben gültig, wenn anstelle einer Zerlegung von Ω in endlich viele Teilmengen eine *unendliche* Folge $B_1, B_2, \dots \in \mathcal{F}$ von Ereignissen mit den Eigenschaften

(Z'1) $B_i \cap B_j = \emptyset$ für $i \neq j$,

(Z'2) $\bigcup_{i=1}^{\infty} B_i = \Omega$,

(Z'3) $P(B_i) > 0$ für alle $i = 1, 2, \dots$

betrachtet wird. Die Formeln (14) und (15) sind dann lediglich wie folgt zu modifizieren:

$$P(A) = \sum_{j=1}^{\infty} P(B_j)P(A | B_j), \quad (16)$$

bzw.

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_{j=1}^{\infty} P(B_j)P(A | B_j)} \quad (17)$$

für jedes $i = 1, 2, \dots$, wobei in (17) erneut vorausgesetzt wird, daß $P(A) > 0$.

Beispiel

- Betrachten eine Fußballmannschaft, deren Siegeschance je Bundesliga-Spiel bei 75% liegt, falls ihr Kapitän in guter Form ist.
- Falls ihr Kapitän jedoch nicht in guter Form ist, dann betrage ihre Siegeschance nur 40%.
- Bei 70% aller Bundesliga-Spiele seiner Mannschaft sei der Kapitän in guter Form.
- Gesucht ist die Wahrscheinlichkeit, daß
 1. die Mannschaft ein Bundesliga-Spiel gewinnt,
 2. der Kapitän bei einem Bundesliga-Spiel in guter Form ist, obwohl die Mannschaft das Spiel nicht gewinnt.
- *Lösung:* Zerlegen den Grundraum Ω auf zwei verschiedene Weisen in zwei Komponenten.
- Sei $A = \{\text{Kapitän ist in guter Form}\}$, $A^c = \{\text{Kapitän ist nicht in guter Form}\}$
bzw.
 $B = \{\text{Mannschaft gewinnt Bundesliga-Spiel}\}$, $B^c = \{\text{Mannschaft gewinnt Bundesliga-Spiel nicht}\}$
- Dann gilt $P(B | A) = 0.75$, $P(B | A^c) = 0.40$, $P(A) = 0.70$
- Aus (14) bzw. (15) ergibt sich nun

$$\begin{aligned} P(B) &= P(B | A)P(A) + P(B | A^c)P(A^c) \\ &= 0.75 \cdot 0.70 + 0.40 \cdot 0.30 = 0.645 \end{aligned}$$

bzw.

$$\begin{aligned} P(A | B^c) &= \frac{P(B^c | A)P(A)}{P(B^c | A)P(A) + P(B^c | A^c)P(A^c)} \\ &= \frac{0.25 \cdot 0.70}{0.25 \cdot 0.70 + 0.60 \cdot 0.30} = 0.493. \end{aligned}$$

3.4.3 Bedingte Verteilung; bedingte Dichte

In Übereinstimmung mit Definition 3.11 ist die folgende Sprechweise üblich.

Definition 3.15

1. Sei (X, Y) ein diskreter Zufallsvektor mit $P((X, Y) \in C) = 1$ für eine abzählbare Menge $C = \{(x_i, y_j) : i, j \in \mathbb{N}\}$. Für jedes $j \in \mathbb{N}$ mit $P(Y = y_j) > 0$ heißt dann $\{P(X = x_i | Y = y_j), i \in \mathbb{N}\}$ die *bedingte Wahrscheinlichkeitsfunktion* von X unter der Bedingung $\{Y = y_j\}$. Sie bestimmt die *bedingte Verteilung* $\{P(X \in B | Y = y_j), B \in \mathcal{B}(\mathbb{R})\}$ von X unter der Bedingung $\{Y = y_j\}$ eindeutig.
2. Analog heißt $\{P(Y = y_j | X = x_i), j \in \mathbb{N}\}$ bzw. $\{P(Y \in B | X = x_i), B \in \mathcal{B}(\mathbb{R})\}$ für jedes $i \in \mathbb{N}$ mit $P(X = x_i) > 0$ die bedingte Wahrscheinlichkeitsfunktion bzw. bedingte Verteilung von Y unter der Bedingung $\{X = x_i\}$.
3. Wenn (X, Y) ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte $f_{(X,Y)}(x, y)$ ist, dann heißt die Funktion $f_{X|Y=y} : \mathbb{R} \rightarrow [0, \infty)$ mit

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

die *bedingte Dichte* von X unter der Bedingung $\{Y = y\}$, wobei $f_Y(y) > 0$ vorausgesetzt wird.

4. Analog heißt die Funktion $f_{Y|X=x} : \mathbb{R} \rightarrow [0, \infty)$ mit

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

die bedingte Dichte von Y unter der Bedingung $\{X = x\}$, wobei $f_X(x) > 0$ vorausgesetzt wird.

Beispiele

1. *zweimaliges Würfeln*

Sei $X : \Omega \rightarrow \{0, 1, 2\}$ bzw. $Y : \Omega \rightarrow \{0, 1, 2\}$ die (zufällige) Anzahl, mit der die Augenzahl „6“ bzw. „1“ beim zweimaligen Würfeln erzielt wird; vgl. Abschnitt 3.3.1. Dann ergeben sich die folgenden bedingten Wahrscheinlichkeitsfunktionen von X unter der Bedingung $\{Y = j\}$:

		i		
		0	1	2
j	$P(X = i Y = j)$	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$
	0	$\frac{16}{25}$	$\frac{8}{25}$	$\frac{1}{25}$
	1	$\frac{8}{10}$	$\frac{2}{10}$	0
	2	1	0	0

2. *integrierte Gleichverteilung*

Sei (X, Y) ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_{(X,Y)}(x, y) = \begin{cases} 4xy, & \text{falls } 0 \leq x, y \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Für $y \in [0, 1]$ gilt dann für die bedingte Dichte von X unter der Bedingung $\{Y = y\}$:

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)} = \begin{cases} 2x, & \text{falls } x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$$

Die bedingte Dichte $f_{X|Y=y}(x)$ stimmt also bei diesem Beispiel mit der (unbedingten Rand-)Dichte $f_X(x)$ von X überein; vgl. Abschnitt 3.3.3.

3.5 Stochastische Unabhängigkeit

3.5.1 Unabhängige Ereignisse

Der Begriff der stochastischen Unabhängigkeit zweier Ereignisse $A, B \in \mathcal{F}$ ist mit der intuitiven Vorstellung verbunden, daß die bedingte Wahrscheinlichkeit $P(A | B)$ des Ereignisses A unter der Bedingung B mit der „unbedingten“ Wahrscheinlichkeit $P(A)$ von A übereinstimmt, d.h. $P(A | B) = P(A)$, wobei $P(B) > 0$ vorausgesetzt wird.

Es ist jedoch zweckmäßiger, die folgende (äquivalente) Gleichung $P(A \cap B) = P(A)P(B)$ zu betrachten, weil durch sie auch der Fall $P(B) = 0$ erfaßt wird.

Definition 3.16

1. Die Ereignisse $A, B \in \mathcal{F}$ heißen *unabhängig*, falls $P(A \cap B) = P(A)P(B)$.
2. Sei $A_1, A_2, \dots, A_n \in \mathcal{F}$ eine beliebige Folge von Ereignissen. Dann sagt man, daß A_1, A_2, \dots, A_n *unabhängige Ereignisse* sind, falls für jede Teilfolge $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots, n\}$ gilt:

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j}). \quad (18)$$

Beachte

1. Der Begriff der Unabhängigkeit wird auch für *unendliche* Folgen von Ereignissen benötigt. Man sagt, daß $A_1, A_2, \dots \in \mathcal{F}$ *unabhängige Ereignisse* sind, falls für jede *endliche* Teilfolge $\{i_1, i_2, \dots, i_k\} \subset \{1, 2, \dots\}$ die Bedingung (18) erfüllt ist.
2. Das folgende Beispiel zeigt, daß aus der Unabhängigkeit von Ereignis-Paaren A_{i_1}, A_{i_2} im allgemeinen *nicht* die (vollständige) Unabhängigkeit der gesamten Folge A_1, A_2, \dots, A_n impliziert wird.

Beispiel

- Sei (Ω, \mathcal{F}, P) gegeben durch $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = \mathcal{P}(\Omega)$, $P(\{k\}) = \frac{1}{4}$ für jedes $k \in \Omega$.
- Sei $A_k = \{k, 4\}$ für $k = 1, 2, 3$.
- Dann sieht man leicht, daß
 1. die Paare A_1, A_2 bzw. A_2, A_3 bzw. A_1, A_3 jeweils unabhängig sind,
 2. die Ereignisse A_1, A_2, A_3 jedoch *nicht* unabhängig sind.
- Denn es gilt

$$P(A_i \cap A_{i+1}) = P(\{4\}) = \frac{1}{4} = P(A_i) \cdot P(A_{i+1}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \quad \text{für } i = 1, 2$$

bzw.

$$P(A_1 \cap A_3) = P(\{4\}) = \frac{1}{4} = P(A_1) \cdot P(A_3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

- Jedoch

$$P(A_1 \cap A_2 \cap A_3) = P(\{4\}) = \frac{1}{4} \neq P(A_1) \cdot P(A_2) \cdot P(A_3) = \frac{1}{8}.$$

3.5.2 Unabhängige Zufallsvariable

Die Unabhängigkeit von Zufallsvariablen wird durch die Unabhängigkeit von Ereignissen ausgedrückt.

So heißen zwei Zufallsvariable X_1, X_2 unabhängig, wenn die Ereignisse $\{X_1 \leq x_1\}$ und $\{X_2 \leq x_2\}$ für beliebige $x_1, x_2 \in \mathbb{R}$ unabhängig im Sinne von Definition 3.16 sind.

Für Folgen von Zufallsvariablen wird der Begriff der Unabhängigkeit folgendermaßen gebildet.

Definition 3.17 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum.

1. Die Zufallsvariablen $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ heißen *unabhängig*, falls

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (19)$$

2. Sei $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ eine beliebige (unendliche) Folge von Zufallsvariablen. Dann sagt man, daß X_1, X_2, \dots *unabhängige Zufallsvariablen* sind, falls jede endliche Teilfolge X_{i_1}, \dots, X_{i_k} von X_1, X_2, \dots aus unabhängigen Zufallsvariablen besteht.

Beachte

1. Aus den Definitionen 3.5 und 3.7 der Verteilungsfunktionen $F_{(X_1, \dots, X_n)}$ und F_{X_1}, \dots, F_{X_n} ergibt sich sofort, daß die Definitionsgleichung (19) äquivalent ist mit

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \dots P(X_n \leq x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}. \quad (20)$$

2. Darüber hinaus kann man zeigen, daß (20) und damit auch (19) äquivalent ist mit

$$P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \dots P(X_n \in B_n) \quad \forall B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}).$$

Hieraus und aus den Definitionsgleichungen (3) und (10) der Dichten $f_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$ und $f_{X_1}(x_1) \dots f_{X_n}(x_n)$ ergibt sich unmittelbar die folgende Charakterisierung der Unabhängigkeit von diskreten bzw. absolutstetigen Zufallsvariablen.

Theorem 3.18

1. Sei $X = (X_1, \dots, X_n)$ ein diskreter Zufallsvektor mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}^n$. Seine Komponenten X_1, \dots, X_n sind genau dann unabhängige Zufallsvariable, wenn

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \dots P(X_n = x_n) \quad \forall (x_1, \dots, x_n) \in C. \quad (21)$$

2. Sei $X = (X_1, \dots, X_n)$ ein absolutstetiger Zufallsvektor. Seine Komponenten X_1, \dots, X_n sind genau dann unabhängige Zufallsvariable, wenn

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n) \quad \forall x_1, \dots, x_n \in \mathbb{R}.$$

3.5.3 Beispiele

1. *n*-maliger Münzwurf

- Betrachten den in Abschnitt 3.2.1 eingeführten Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ mit der Grundmenge

$$\Omega = \Omega_1 \times \dots \times \Omega_n = \{\omega : \omega = (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}$$

und dem Wahrscheinlichkeitsmaß P , das durch

$$P(\{\omega\}) = \prod_{i:\omega_i=1} a_i \prod_{j:\omega_j=0} (1 - a_j) \quad (22)$$

gegeben ist, wobei $0 \leq a_1, \dots, a_n \leq 1$.

- Betrachten die Zufallsvariablen $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$, die gegeben seien durch die Projektion $X_i(\omega) = \omega_i$ für $i = 1, \dots, n$.
- Aus (21) und (22) folgt unmittelbar, daß X_1, \dots, X_n unabhängige Zufallsvariable sind.

- Man kann jedoch auch in diesem Modell Zufallsvariablen konstruieren, die nicht unabhängig sind. Sei nämlich $a_1 = 0.5$, und sei $Y_1 : \Omega \rightarrow \mathbb{R}$ gegeben durch

$$Y_1(\omega) = \begin{cases} 1 & \text{falls } \omega_1 = 0 \\ 0 & \text{falls } \omega_1 = 1 \end{cases}$$

- Dann sind X_1 und Y_1 nicht unabhängig, denn es gilt

$$P(X_1 = 1, Y_1 = 1) = 0 \quad \neq \quad P(X_1 = 1)P(Y_1 = 1) = \frac{1}{4}.$$

2. zweidimensionale Normalverteilung

- Betrachten zwei unabhängige Zufallsvariable X_1, X_2 , die standardnormalverteilt sind. D.h.,

$$f_{X_i}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \quad \forall x \in \mathbb{R}; i = 1, 2.$$

- Für die (gemeinsame) Dichte $f_X(x_1, x_2)$ des Zufallsvektors $X = (X_1, X_2)$ gilt

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right).$$

- Man sagt dann, daß auch der Zufallsvektor $X = (X_1, X_2)$ *standardnormalverteilt* ist.
- *Verallgemeinerung*: Sei $\rho \in [0, 1)$, und sei $X = (X_1, X_2)$ ein absolutstetiger Zufallsvektor mit der Dichte

$$f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right) \quad \forall x_1, x_2 \in \mathbb{R}.$$

- Dann gilt für die (Rand-)Dichte $f_{X_1}(x_1)$ von X_1

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{1-\rho^2}\right) dx_2 \\ &\stackrel{\text{quadratische Ergänzung}}{=} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} \frac{x_1^2(1-\rho^2)}{1-\rho^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \frac{(x_2 - \rho x_1)^2}{1-\rho^2}\right) dx_2. \end{aligned}$$

- Durch die Substitution

$$u = \frac{x_2 - \rho x_1}{\sqrt{1-\rho^2}} \quad \text{bzw.} \quad du = \frac{1}{\sqrt{1-\rho^2}} dx_2$$

ergibt sich also

$$f_{X_1}(x_1) = \frac{1}{2\pi} \exp\left(-\frac{1}{2}x_1^2\right) \underbrace{\int_{\mathbb{R}} \exp\left(-\frac{1}{2}u^2\right) du}_{=\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right).$$

- Analog gilt

$$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right).$$

- Die Komponenten X_1, X_2 des Zufallsvektors $X = (X_1, X_2)$ sind also für *jedes* $\rho \in [0, 1)$ standardnormalverteilt.
- Beachte jedoch, daß X_1 und X_2 nur dann unabhängig sind, wenn $\rho = 0$. Denn für $\rho > 0$ gilt

$$f_X(x_1, x_2) \neq f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

3.6 Funktionen von Zufallsvariablen

3.6.1 Zusammengesetzte Abbildungen

Beispiel (Klassifikation)

- Manchmal ist es zweckmäßig, die Werte von Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}$ zu klassifizieren. Dabei wird der Wertebereich \mathbb{R} von X in m Klassen zerlegt, die wir mit der Menge der ersten m natürlichen Zahlen $\{1, 2, \dots, m\}$ identifizieren.
- Mit anderen Worten: Außer der Zufallsvariablen X betrachten wir noch eine weitere Abbildung $\varphi : \mathbb{R} \rightarrow \{1, 2, \dots, m\}$.
- Durch Nacheinanderausführung der Abbildungen X und φ ergibt sich dann die Abbildung $\varphi(X) : \Omega \rightarrow \{1, 2, \dots, m\}$ mit $\varphi(X)(\omega) = \varphi(X(\omega))$, die jedem $\omega \in \Omega$ die Klasse $\varphi(X)(\omega)$ zuordnet.
- Um die Wahrscheinlichkeit bestimmen zu können, daß die Zufallsvariable X Werte in Klasse i annimmt, muß gewährleistet sein, daß

$$\{\omega : \omega \in \Omega, \varphi(X)(\omega) = i\} \in \mathcal{F}. \quad (23)$$

- Die Abbildung $\varphi(X)$ muß also die Regularitätseigenschaft einer Zufallsvariablen besitzen, d.h. der Meßbarkeitsbedingung (1) genügen.
- Um dies zu erreichen, wird über die Abbildung φ vorausgesetzt, daß $\{x : x \in \mathbb{R}, \varphi(x) = i\}$ eine Teilmenge von \mathbb{R} aus $\mathcal{B}(\mathbb{R})$, d.h. eine Borel-Menge ist, für jedes $i = 1, 2, \dots, m$.
- Man kann zeigen, daß dann (23) für jedes $i \in \{1, 2, \dots, m\}$ bzw. $\{\omega : \omega \in \Omega, \varphi(X)(\omega) \leq x\} \in \mathcal{F}$ für jedes $x \in \mathbb{R}$ gilt, d.h., $\varphi(X)$ ist eine Zufallsvariable.

Beachte

- Es interessieren auch Funktionen $\varphi : \mathbb{R}^n \rightarrow \{1, 2, \dots, m\}$ von Zufallsvektoren $X : \Omega \rightarrow \mathbb{R}^n$.
- Damit auch in diesem Fall eine Bedingung an φ formuliert werden kann, so daß die zusammengesetzte Abbildung $\varphi(X) : \Omega \rightarrow \{1, 2, \dots, m\}$ eine Zufallsvariable ist, wird die *Borel- σ -Algebra* $\mathcal{B}(\mathbb{R}^n)$ von Teilmengen des \mathbb{R}^n betrachtet.
- Sie ist definiert als die kleinste σ -Algebra von Teilmengen des \mathbb{R}^n , die alle offenen Quader $\times_{i=1}^n (a_i, b_i)$ enthält; $-\infty < a_i < b_i < \infty$. D.h.

$$\mathcal{B}(\mathbb{R}^n) = \sigma \left(\underbrace{\left\{ \times_{i=1}^n (a_i, b_i), -\infty < a_i < b_i < \infty \right\}}_{\text{Erzeugersystem}} \right).$$

- Die Abbildung $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ heißt *Borel-meßbar*, wenn

$$\{x : x \in \mathbb{R}^n, \varphi(x) \leq y\} \in \mathcal{B}(\mathbb{R}^n) \quad \forall y \in \mathbb{R}. \quad (24)$$

Allgemein gilt für zusammengesetzte Abbildungen

Lemma 3.19 Sei (Ω, \mathcal{F}, P) ein beliebiger Wahrscheinlichkeitsraum, und $X : \Omega \rightarrow \mathbb{R}^n$ sei ein beliebiger Zufallsvektor. Falls $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Borel-meßbare Abbildung ist, d.h., falls die Bedingung (24) erfüllt ist, dann ist die zusammengesetzte Abbildung $\varphi(X) : \Omega \rightarrow \mathbb{R}$ mit $\varphi(X)(\omega) = \varphi(X(\omega))$ eine (reellwertige) Zufallsvariable, d.h., es gilt

$$\{\omega : \omega \in \Omega, \varphi(X)(\omega) \leq z\} \in \mathcal{F} \quad \forall z \in \mathbb{R}.$$

Beispiel

- Sei $X = (X_1, \dots, X_n)$ ein beliebiger Zufallsvektor.

- Betrachten die folgende Abbildung $\varphi : \mathbb{R}^n \rightarrow \{0, 1, \dots, m\}$, die als Klassifikation der Werte von X aufgefaßt werden kann.
- Sei $\varepsilon > 0$ eine beliebige, jedoch fest vorgegebene Toleranzgrenze.
- Für jedes $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ sei

$$\varphi(x) = |\{(i, j) : 1 \leq i < j \leq n, -\varepsilon \leq x_i - x_j \leq \varepsilon\}|, \quad (25)$$

d.h., $\varphi(x)$ ist die Anzahl derjenigen Paare von Komponenten x_i, x_j von x , die sich um nicht mehr als ε voneinander unterscheiden.

- Der Wertebereich \mathbb{R}^n von X wird dabei in $m + 1 = \binom{n}{2} + 1$ Klassen zerlegt.
- Es ist nicht schwierig zu zeigen, daß die in (25) gegebene Abbildung die Regularitätsbedingung (24) erfüllt.

Beachte

- Außerdem kann man zeigen, daß jede stetige Funktion $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ die Bedingung (24) erfüllt, d.h., jede stetige Funktion ist Borel-meßbar.
- Es ist klar, daß φ insbesondere dann stetig ist, wenn φ ein *Polynom* ist, d.h.

$$\varphi(x_1, \dots, x_n) = \sum_{i=0}^k a_i x_1^{m_{i1}} \dots x_n^{m_{in}} \quad (26)$$

für ein $k \in \mathbb{N}$ und für gewisse Konstanten $a_0, a_1, \dots, a_k \in \mathbb{R}$, $m_{i1}, \dots, m_{in} \geq 0$.

- In den folgenden Abschnitten wird eine Reihe von Spezialfällen diskutiert, bei denen φ die in (26) gegebene Form hat.

3.6.2 Lineare Transformation

Ein wichtiger Spezialfall einer zusammengesetzten Abbildung ist die lineare Transformation von Zufallsvariablen, wobei $n = 1$ und $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit $\varphi(x) = ax + b$; $a, b \in \mathbb{R}$.

Theorem 3.20 Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable und $a, b \in \mathbb{R}$ beliebige Zahlen mit $a \neq 0$. Dann ist $aX + b$ eine Zufallsvariable, und

1. die Verteilungsfunktion von $aX + b$ ist gegeben durch

$$F_{aX+b}(x) = \begin{cases} F_X\left(\frac{x-b}{a}\right), & \text{falls } a > 0, \\ 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right), & \text{falls } a < 0. \end{cases} \quad (27)$$

2. falls X absolutstetig ist mit der Dichte f_X , dann ist auch $aX + b$ absolutstetig mit der Dichte

$$f_{aX+b}(x) = \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right). \quad (28)$$

Beweis Falls $a > 0$, dann gilt

$$F_{aX+b}(x) = P(aX + b \leq x) = P\left(X \leq \frac{x-b}{a}\right) = F_X\left(\frac{x-b}{a}\right).$$

Analog ergibt sich für $a < 0$

$$\begin{aligned} F_{aX+b}(x) &= P(aX + b \leq x) \\ &= P\left(X \geq \frac{x-b}{a}\right) \\ &= 1 - P\left(X < \frac{x-b}{a}\right) \\ &= 1 - F_X\left(\frac{x-b}{a}\right) + P\left(X = \frac{x-b}{a}\right). \end{aligned}$$

Damit ist (27) bewiesen. Sei nun X absolutstetig. Falls die Dichte $f_X(x)$ von X eine stetige Funktion ist, dann ergibt sich (28) durch beidseitiges Differenzieren von (27). Ansonsten nutzt man die Tatsache, daß zwischen Verteilungsfunktion und Dichte einer Zufallsvariablen eine eindeutige Zuordnung besteht, und zeigt, daß das Integral von $|a|^{-1} f_X((x-b)/a)$ die Verteilungsfunktion von $aX + b$ ergibt.

Beispiel

- Sei X standardnormalverteilt, und $\sigma > 0$, $\mu \in \mathbb{R}$ seien beliebige Konstanten.
- Gemäß Theorem 3.20 ist dann die Zufallsvariable $Y = \sigma X + \mu$ absolutstetig, und es gilt

$$f_Y(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \quad (29)$$

- Eine Zufallsvariable Y , deren Dichte durch (29) gegeben ist, heißt *normalverteilt* mit den Parametern μ und σ^2 ; vgl. auch Abschnitt 3.2.3. Dabei verwenden wir die Schreibweise $Y \sim N(\mu, \sigma^2)$.
- Umgekehrt ergibt sich, ausgehend von einer normalverteilten Zufallsvariablen $Y \sim N(\mu, \sigma^2)$, durch die lineare Transformation $X = (Y - \mu)/\sigma$ eine standardnormalverteilte Zufallsvariable $X \sim N(0, 1)$.

3.6.3 Quadrierung

Theorem 3.21 Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsgröße. Dann gilt

1. für die Verteilungsfunktion von X^2

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

2. falls X absolutstetig ist mit der Dichte f_X , dann ist auch X^2 absolutstetig, und es gilt

$$f_{X^2}(x) = \begin{cases} \frac{1}{2\sqrt{x}}(f_X(\sqrt{x}) + f_X(-\sqrt{x})), & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0. \end{cases} \quad (30)$$

Beweis analog zum Beweis von Theorem 3.20, wobei jetzt $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit $\varphi(x) = x^2$.

Beispiel Falls $X \sim N(0, 1)$, dann ergibt sich aus (30):

$$f_{X^2}(x) = \begin{cases} \frac{1}{\sqrt{2\pi x}} \exp\left(-\frac{x}{2}\right) & \text{falls } x \geq 0, \\ 0, & \text{falls } x < 0. \end{cases} \quad (31)$$

Beachte Die Summe von unabhängigen (und identisch verteilten) Zufallsvariablen, deren Dichte durch (31) gegeben ist, heißt χ^2 -verteilt. Die χ^2 -Verteilung ist eine sogenannte *statistische Prüfverteilung*, die im weiteren Verlauf der Vorlesung noch genauer diskutiert wird.

3.6.4 Summe, Produkt und Quotient von unabhängigen Zufallsvariablen

Theorem 3.22 Sei $X = (X_1, X_2) : \Omega \rightarrow \mathbb{R}^2$ ein absolutstetiger Zufallsvektor mit der (gemeinsamen) Dichte f_X . Dann ist auch die Zufallsvariable $X_1 + X_2$ absolutstetig, und ihre Dichte ist gegeben durch

$$f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} f_X(t, z-t) dt \quad \forall z \in \mathbb{R}. \quad (32)$$

Falls die Zufallsvariablen X_1, X_2 unabhängig sind, dann gilt insbesondere die sogenannte *Faltungsformel*

$$f_{X_1+X_2}(z) = \int_{-\infty}^{\infty} f_{X_1}(t)f_{X_2}(z-t) dt \quad \forall z \in \mathbb{R}. \quad (33)$$

Beweis Wir benutzen die Tatsache, daß zwischen Verteilungsfunktion und Dichte einer Zufallsvariablen eine eindeutige Zuordnung besteht, und zeigen, daß das Integral der Funktion in (32) die Verteilungsfunktion von $X_1 + X_2$ ergibt. Und zwar gilt für jedes $z \in \mathbb{R}$

$$\begin{aligned} \int_{-\infty}^z \int_{-\infty}^{\infty} f_X(t, v-t) dt dv &= \int_{-\infty}^{\infty} \int_{-\infty}^z f_X(t, \underbrace{v-t}_{=u}) dv dt \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-t} f_X(t, u) du dt \\ &= \int \int_{\{(t, u): t+u \leq z\}} f_X(t, u) du dt \\ &= P(X_1 + X_2 \leq z). \end{aligned}$$

Die Faltungsformel (33) ergibt sich unmittelbar (32), weil $f_{(X_1, X_2)}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$, falls X_1 und X_2 unabhängig sind.

Beachte

- Falls die Zufallsvariablen X_1, X_2 unabhängig sind mit $X_1 \sim N(\mu_1, \sigma_1^2)$ bzw. $X_2 \sim N(\mu_2, \sigma_2^2)$, dann ergibt sich aus (33), daß auch die Summe $X_1 + X_2$ normalverteilt ist mit

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

- Dieses Additionstheorem für unabhängige und normalverteilte Zufallsvariablen wird auch *Faltungsstabilität* der Normalverteilung genannt.
- Durch Iteration ergibt sich für jede beliebige Anzahl $n \geq 2$ von unabhängigen Zufallsvariablen X_1, \dots, X_n mit $X_i \sim N(\mu_i, \sigma_i^2)$ für alle $i \in \{1, \dots, n\}$, daß

$$X_1 + \dots + X_n \sim N(\mu_1 + \dots + \mu_n, \sigma_1^2 + \dots + \sigma_n^2).$$

Völlig analog zu Theorem 3.22 ergibt sich

Theorem 3.23 Die Zufallsvariablen X_1 und X_2 seien unabhängig und absolutstetig mit den Dichten f_{X_1} und f_{X_2} . Dann sind die Zufallsvariablen $X_1 \cdot X_2$ und X_1/X_2 absolutstetig, und ihre Dichten sind gegeben durch

$$f_{X_1 \cdot X_2}(z) = \int_{-\infty}^{\infty} \frac{1}{|t|} f_{X_1}(t) f_{X_2}\left(\frac{z}{t}\right) dt \quad \forall z \in \mathbb{R} \quad (34)$$

bzw.

$$f_{X_1/X_2}(z) = \int_{-\infty}^{\infty} |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \quad \forall z \in \mathbb{R}. \quad (35)$$

Beachte

- Der Fall $X_2(\omega) = 0$, der bei der Bildung des Quotienten X_1/X_2 zur Division durch Null führen würde, tritt nur mit Wahrscheinlichkeit Null auf (weil X_2 absolutstetig ist).
- Deshalb kann $X_1(\omega)/X_2(\omega)$ für solche $\omega \in \Omega$ gesondert definiert werden (z.B. können wir dann $X_1(\omega)/X_2(\omega) = 0$ setzen).

Beispiel Falls X_1 und X_2 unabhängig sind mit $X_1 \sim N(0,1)$ und $X_2 \sim N(0,1)$, dann gilt

$$f_{X_1/X_2}(z) = \frac{1}{\pi(z^2 + 1)} \quad \forall z \in \mathbb{R}. \quad (36)$$

denn aus (35) ergibt sich, daß

$$\begin{aligned} f_{X_1/X_2}(z) &= \int_{-\infty}^{\infty} |t| f_{X_1}(z \cdot t) f_{X_2}(t) dt \\ &= \int_{-\infty}^{\infty} |t| \frac{1}{2\pi} \exp\left(-\frac{1}{2}((zt)^2 + t^2)\right) dt \\ &\stackrel{\text{Symmetrie}}{=} \frac{1}{\pi} \int_0^{\infty} t \exp\left(-\underbrace{\frac{t^2}{2}(z^2 + 1)}_{=v}\right) dt \\ &= \frac{1}{\pi(z^2 + 1)} \underbrace{\int_0^{\infty} e^{-v} dv}_{=1} = \frac{1}{\pi(z^2 + 1)}. \end{aligned}$$

Beachte Eine absolutstetige Zufallsvariable mit der in (36) gegebenen Dichte heißt *Cauchy-verteilt*.

4 Weitere Charakteristiken von Zufallsvariablen; Grenzwertsätze

4.1 Momente von Zufallsvariablen

Weitere wichtige Charakteristiken von Zufallsvariablen sind deren Momente, insbesondere der Erwartungswert und die Varianz.

4.1.1 Erwartungswert

Bevor wir zur allgemeinen Definition des Erwartungswertes kommen, wollen wir die intuitive Bedeutung dieses Begriffes anhand des folgenden Beispiels erläutern.

Beispiel (wiederholtes Würfeln)

- Betrachten den Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$ mit der Grundmenge

$$\Omega = \Omega_1 \times \Omega_2 \times \dots = \{\omega : \omega = (\omega_1, \omega_2, \dots), \omega_i \in \{1, 2, \dots, 6\}\}$$

und dem Wahrscheinlichkeitsmaß P , das durch

$$P(\{\omega : \omega \in \Omega, \omega_{i_1} = j_1, \dots, \omega_{i_k} = j_k\}) = \frac{1}{6^k}$$

gegeben ist; $k \in \mathbb{N}$; $1 \leq i_1 < \dots < i_k$; $j_1, \dots, j_k \in \{1, 2, \dots, 6\}$.

- Betrachten die Zufallsvariablen $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$, die gegeben seien durch die Projektion $X_i(\omega) = \omega_i$ für $i = 1, 2, \dots$. D.h., X_i ist die (zufällige) Augenzahl, die beim i -ten Würfeln erzielt wird.
- Es ist nicht schwierig zu zeigen, daß X_1, X_2, \dots unabhängige (und identisch verteilte) Zufallsvariable sind.
- Betrachten die Zufallsvariable $Y_n = n^{-1}(X_1 + \dots + X_n)$, d.h. die *mittlere Augenzahl* bei n -maligem Würfeln.
- Man kann zeigen, daß es eine „nichtzufällige“ Zahl $c \in \mathbb{R}$ gibt, so daß

$$P(\{\omega : \lim_{n \rightarrow \infty} Y_n(\omega) = c\}) = 1, \quad (1)$$

wobei

$$c = \sum_{i=1}^6 i P(X_1 = i) = \frac{1}{6} \sum_{i=1}^6 i = 3.5. \quad (2)$$

- Die Formeln (1) und (2) bedeuten: Falls die Anzahl n der durchgeführten Versuche immer größer wird, dann
 1. werden die Werte $Y_n(\omega)$ der mittleren Augenzahl Y_n immer weniger von der jeweiligen Ausprägung ω des Zufalls beeinflusst,
 2. strebt das „Zeitmittel“ Y_n bei n Versuchen gegen das „Scharmittel“ c jedes (einzelnen) Versuches.
- Die Formeln (1) und (2) sind ein Spezialfall des sogenannten *Gesetzes der großen Zahlen*, das im weiteren Verlauf der Vorlesung noch genauer diskutiert wird.
- Das Scharmittel c in (2) wird *Erwartungswert* der Zufallsvariablen X_1 genannt und mit $\mathbb{E} X_1$ bezeichnet.

Auf analoge Weise wird der Begriff des Erwartungswertes für beliebige (diskrete bzw. absolutstetige) Zufallsvariable eingeführt.

Definition 4.1 Betrachten einen beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

1. Sei $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$. Dann heißt das gewichtete Mittel

$$\mathbb{E} X = \sum_{x \in C} x P(X = x) \quad (3)$$

der *Erwartungswert* von X , wobei vorausgesetzt wird, daß

$$\sum_{x \in C} |x| P(X = x) < \infty. \quad (4)$$

2. Sei $X : \Omega \rightarrow \mathbb{R}$ eine absolutstetige Zufallsvariable mit der Dichte $f_X(x)$. Dann heißt das Integral

$$\mathbb{E} X = \int_{-\infty}^{\infty} x f_X(x) dx \quad (5)$$

der *Erwartungswert* von X , wobei vorausgesetzt wird, daß

$$\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty. \quad (6)$$

Beachte

- Die Summe in (4) bzw. das Integral in (6) kann man als Erwartungswert $\mathbb{E}|X|$ der Zufallsvariablen $|X|$ auffassen.
- Man kann sich leicht Beispiele überlegen, bei denen die Summierbarkeitsbedingung (4) bzw. die Integrierbarkeitsbedingung (6) verletzt ist.
- Falls die Zufallsgröße X nur nichtnegative Werte annimmt, d.h. $P(X \geq 0) = 1$, dann kann man den Begriff des Erwartungswertes auch einführen, wenn die Bedingung (4) bzw. (6) *nicht* erfüllt ist.
- In diesem Fall wird $\mathbb{E} X = \infty$ gesetzt.
- Sei X beispielsweise eine absolutstetige Zufallsvariable mit der Dichte

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

d.h. X ist Cauchy-verteilt, vgl. Abschnitt 3.6.4.

- Dann gilt zwar $P(0 < |X| < \infty) = 1$, jedoch

$$\begin{aligned} \mathbb{E}|X| &= \int_{-\infty}^{\infty} |x| \frac{1}{\pi(1+x^2)} dx \\ &> \int_0^{\infty} \frac{x}{\pi(1+x^2)} dx \\ &= \frac{1}{\pi} \lim_{t \rightarrow \infty} \left[\frac{1}{2} \log(1+x^2) \right]_0^t = \frac{1}{2\pi} \lim_{t \rightarrow \infty} \log(1+t^2) = \infty. \end{aligned}$$

Beispiele Wir zeigen nun anhand zweier Beispiele, wie die Formeln (3) und (5) zur Bestimmung des Erwartungswertes genutzt werden können.

1. *Binomialverteilung*

Sei X binomialverteilt mit den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann ergibt sich aus (3), daß

$$\begin{aligned} \mathbb{E} X &= \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} \\ &= n p \sum_{i=1}^n \binom{n-1}{i-1} p^{i-1} (1-p)^{(n-1)-(i-1)} \\ &= n p \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\ &= n p (p + (1-p))^{n-1} = n p. \end{aligned}$$

2. *Normalverteilung*

Sei X normalverteilt mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$. Dann ergibt sich aus (5), daß

$$\begin{aligned} \mathbb{E} X &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(-\frac{1}{2} \underbrace{\left(\frac{x-\mu}{\sigma}\right)^2}_{=v}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma v + \mu) \exp\left(-\frac{1}{2} v^2\right) dv \\ &= \underbrace{\frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v \exp\left(-\frac{1}{2} v^2\right) dv}_{=0} + \underbrace{\frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} v^2\right) dv}_{=\sqrt{2\pi}} = \mu. \end{aligned}$$

Beachte

- Die in (1) und (2) angegebenen Formeln für den Erwartungswert von diskreten bzw. absolutstetigen Zufallsvariablen sind Spezialfälle eines allgemeineren (und damit einheitlichen) Zuganges zum Begriff des Erwartungswertes für beliebige Zufallsvariablen.
- Dieser allgemeinere Zugang beruht jedoch auf einem (im Vergleich zum Riemann-Integral) modifizierten Integral-Begriff, der über den Rahmen der einführenden Stochastik-Vorlesung hinausgeht.
- Sei nämlich $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable mit der Verteilungsfunktion F_X . Dann heißt

$$\mathbb{E} X = \int_{-\infty}^{\infty} x dF_X(x) \quad (7)$$

der *Erwartungswert* von X , wobei das Integral in (7) ein sogenanntes *Stieltjes-Integral* ist.

- Schließlich sei vermerkt, daß es noch eine weitere (mathematische äquivalente) Definitionsmöglichkeit des Erwartungswertes $\mathbb{E} X$ von X gibt, und zwar als ein sogenanntes *Lebesgue-Integral*

$$\mathbb{E} X = \int_{\Omega} X(\omega) P[d\omega] \quad (8)$$

über dem Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) .

- Insbesondere kann man zeigen, daß die Integrale in (7) und (8) übereinstimmen.

4.1.2 Varianz und höhere Momente

Es ist klar, daß der Erwartungswert $\mathbb{E}X$ einer (diskreten bzw. absolutstetigen) Zufallsvariablen X eindeutig durch die Wahrscheinlichkeitsfunktion bzw. die Dichte von X bestimmt wird; vgl. Definition 4.1.

Umgekehrt ist jedoch im allgemeinen die Wahrscheinlichkeitsfunktion bzw. die Dichte einer (diskreten bzw. absolutstetigen) Zufallsvariablen X *nicht eindeutig* durch den Erwartungswert $\mathbb{E}X$ von X festgelegt.

Beispiel (symmetrische diskrete Gleichverteilung)

- Sei $n \in \mathbb{N}$ eine beliebige natürliche Zahl und $X : \Omega \rightarrow \{-n, \dots, -1, 0, 1, \dots, n\}$ eine diskrete Zufallsvariable mit $P(X = i) = (2n + 1)^{-1}$ für jedes $i \in \{-n, \dots, n\}$.
- Dann gilt

$$\mathbb{E}X = \frac{1}{2n + 1} \sum_{i=-n}^n i = \frac{1}{2n + 1} \cdot 0 = 0.$$

- Während der Erwartungswert $\mathbb{E}X$ also nicht von n abhängt, sind die Werte von X mit wachsendem n immer breiter um $\mathbb{E}X$ gestreut.
- Eine Charakteristik, die den Streuungsgrad der Werte von X um den Erwartungswert $\mathbb{E}X$ mißt, ist die *erwartete quadratische Abweichung* $\mathbb{E}((X - \mathbb{E}X)^2)$ vom Erwartungswert $\mathbb{E}X$, genannt *Varianz* von X , die bei diesem Beispiel gegeben ist durch

$$\begin{aligned} \mathbb{E}((X - \mathbb{E}X)^2) &= \sum_{i=-n}^n i^2 \frac{1}{2n + 1} \\ &= 2 \frac{n(n + 1)(2n + 1)}{6} \frac{1}{2n + 1} \\ &= \frac{n(n + 1)}{3}. \end{aligned}$$

Der Erwartungswert $\mathbb{E}X$ der Zufallsvariablen X wird manchmal auch das *erste Moment* von X genannt. Völlig analog lassen sich die Begriffe der Varianz bzw. der höheren Momente einer beliebigen Zufallsvariable X einführen, und zwar durch die Betrachtung des Erwartungswertes $\mathbb{E}\varphi(X)$ entsprechend gewählter Funktionen $\varphi(X)$ von X .

In diesem Zusammenhang ist der folgende Hilfssatz nützlich.

Lemma 4.2 Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable, und sei $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ eine stetige Abbildung. Dann läßt sich der Erwartungswert $\mathbb{E}\varphi(X)$ der Zufallsvariablen $\varphi(X) : \Omega \rightarrow \mathbb{R}$ wie folgt darstellen.

1. Falls X diskret ist mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$, dann gilt

$$\mathbb{E}\varphi(X) = \sum_{x \in C} \varphi(x)P(X = x), \quad (9)$$

wobei vorausgesetzt wird, daß $\sum_{x \in C} |\varphi(x)|P(X = x) < \infty$.

2. Falls X absolutstetig ist mit der Dichte $f_X(x)$, dann gilt

$$\mathbb{E}\varphi(X) = \int_{-\infty}^{\infty} \varphi(x) f_X(x) dx, \quad (10)$$

wobei vorausgesetzt wird, daß $\int_{-\infty}^{\infty} |\varphi(x)| f_X(x) dx < \infty$.

Beweis Wir zeigen nur die Gültigkeit von (9). Die Herleitung von (10) erfordert mathematische Hilfsmittel, die über den Rahmen dieser einführenden Vorlesung hinausgehen. Sei also X diskret ist mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$. Dann ist auch $\varphi(X)$ eine diskrete Zufallsvariable mit

$$P(\{\omega : \omega \in \Omega, \varphi(X)(\omega) \in \varphi(C)\}) = 1,$$

wobei $\varphi(C) = \{\varphi(x) : x \in C\}$. Aus der Definition (3) für den Erwartungswert diskreter Zufallsvariablen ergibt sich somit

$$\begin{aligned} \mathbb{E} \varphi(X) &= \sum_{y \in \varphi(C)} y P(\varphi(X) = y) \\ &= \sum_{y \in \varphi(C)} y \left(\sum_{x: \varphi(x)=y} P(X = x) \right) \\ &= \sum_{x \in C} \varphi(x) P(X = x). \end{aligned}$$

Definition 4.3

- Sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable mit $\mathbb{E}(X^2) < \infty$.
- Betrachten die Abbildung $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit $\varphi(x) = (x - \mathbb{E} X)^2$.
- Dann heißt der Erwartungswert $\mathbb{E} \varphi(X)$ der Zufallsvariablen $\varphi(X)$ die *Varianz* von X (Schreibweise: $\text{Var } X$).
- Für die Varianz von X gilt also

$$\text{Var } X = \mathbb{E} \left((X - \mathbb{E} X)^2 \right). \quad (11)$$

Beispiele

1. Binomialverteilung

- Sei X binomialverteilt mit den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$.
- In Abschnitt 4.1.1 wurde gezeigt, daß $\mathbb{E} X = np$.
- Analog ergibt sich, daß

$$\begin{aligned} \sum_{i=1}^n i^2 P(X = i) &= \sum_{i=1}^n (i(i-1) + i) P(X = i) \\ &= \sum_{i=1}^n i(i-1) P(X = i) + \sum_{i=1}^n i P(X = i) \\ &= n(n-1)p^2 + np \end{aligned}$$

- Also gilt

$$\begin{aligned} \text{Var } X &= \sum_{i=0}^n (i - np)^2 P(X = i) \\ &= \sum_{i=0}^n (i^2 - 2npi + (np)^2) P(X = i) \\ &= \sum_{i=1}^n i^2 P(X = i) - 2np \sum_{i=1}^n i P(X = i) + (np)^2 \sum_{i=0}^n P(X = i) \\ &= n(n-1)p^2 + np - 2(np)^2 + (np)^2 = np(1-p). \end{aligned}$$

2. Normalverteilung

- Sei X normalverteilt mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$.
- In Abschnitt 4.1.1 wurde gezeigt, daß $\mathbb{E} X = \mu$.
- Somit gilt

$$\begin{aligned}
 \text{Var } X &= \mathbb{E}((X - \mu)^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 \exp\left(-\frac{1}{2}\underbrace{\left(\frac{x - \mu}{\sigma}\right)^2}_{=t}\right) dx \\
 &= \sigma^2 \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 \exp\left(-\frac{1}{2}t^2\right) dt \\
 &= \sigma^2 \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \sqrt{2u} \exp(-u) du \\
 &= \sigma^2 \frac{2}{\sqrt{\pi}} \underbrace{\int_0^{\infty} \sqrt{u} \exp(-u) du}_{= \Gamma(3/2)} = \sigma^2. \\
 &= \Gamma(3/2) = \frac{1}{2}\Gamma(1/2) = \sqrt{\pi}/2
 \end{aligned}$$

Beachte

- Die Parameter (und damit die Wahrscheinlichkeitsfunktion bzw. die Dichte) von Binomial- bzw. Normalverteilung sind jeweils eindeutig durch den Erwartungswert und die Varianz dieser Verteilungen festgelegt.
- Für die Normalverteilung ist dies offensichtlich. Denn, wie soeben gezeigt, gilt $\mathbb{E} X = \mu$ und $\text{Var } X = \sigma^2$, falls $X \sim N(\mu, \sigma^2)$.
- Sei nun $Y \sim \text{Bin}(n, p)$. Dann kann man sich leicht überlegen, daß

$$n = \frac{\mu^2}{\mu - \sigma^2}, \quad p = \frac{\mu - \sigma^2}{\mu},$$

wobei

$$\mu = \mathbb{E} Y = np \quad \text{und} \quad \sigma^2 = \text{Var } Y = np(1 - p). \quad (12)$$

- Im allgemeinen wird die Verteilung einer Zufallsvariablen jedoch *nicht eindeutig* durch den Erwartungswert und die Varianz bestimmt.
- *Beispiel.* Der Erwartungswert und die Varianz von $X \sim N(\mu, \sigma^2)$ stimmen mit dem Erwartungswert und der Varianz von $Y \sim \text{Bin}(n, p)$ überein, falls μ und σ^2 durch (12) gegeben sind.

Beachte Sei $k \in \mathbb{N}$, und sei $X : \Omega \rightarrow \mathbb{R}$ eine beliebige Zufallsvariable mit $\mathbb{E}(|X|^k) < \infty$.

- Betrachten die Abbildung $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit $\varphi(x) = x^k$. Dann heißt der Erwartungswert $\mathbb{E} \varphi(X) = \mathbb{E}(X^k)$ von $\varphi(X)$ das *k-te Moment* von X .
- Betrachten die Abbildung $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ mit $\varphi(x) = (x - \mathbb{E} X)^k$. Dann heißt der Erwartungswert $\mathbb{E} \varphi(X) = \mathbb{E}((X - \mathbb{E} X)^k)$ von $\varphi(X)$ das *k-te zentrale Moment* von X .
- Die Varianz $\text{Var } X$ ist also das 2. zentrale Moment von X .
- $\sqrt{\text{Var } X}$ heißt *Standardabweichung* von X .

4.1.3 Eigenschaften von Erwartungswert und Varianz

Wir diskutieren nun einige nützliche Eigenschaften von Erwartungswert und Varianz.

Bei der Herleitung dieser Eigenschaften wird die folgende Verallgemeinerung von Lemma 4.2 für Zufallsvektoren benötigt.

Lemma 4.4 Sei $X : \Omega \rightarrow \mathbb{R}^n$ ein beliebiger Zufallsvektor, und sei $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ eine stetige Abbildung. Dann läßt sich der Erwartungswert $\mathbb{E} \varphi(X)$ der Zufallsvariablen $\varphi(X) : \Omega \rightarrow \mathbb{R}$ wie folgt darstellen.

1. Falls X diskret ist mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}^n$, dann gilt

$$\mathbb{E} \varphi(X) = \sum_{x \in C} \varphi(x) P(X = x), \quad (13)$$

wobei vorausgesetzt wird, daß

$$\sum_{x \in C} |\varphi(x)| P(X = x) < \infty.$$

2. Falls X absolutstetig ist mit der (gemeinsamen) Dichte $f_X(x)$, dann gilt

$$\mathbb{E} \varphi(X) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (14)$$

wobei vorausgesetzt wird, daß

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} |\varphi(x_1, \dots, x_n)| f_X(x_1, \dots, x_n) dx_1 \dots dx_n < \infty.$$

Beweis analog zum Beweis von Lemma 4.2.

Theorem 4.5 Seien X_1 und X_2 beliebige Zufallsvariable mit $\mathbb{E}|X_1| < \infty$, $\mathbb{E}|X_2| < \infty$. Dann gilt für beliebige Konstanten $a, b \in \mathbb{R}$

$$\mathbb{E}(aX_1 + bX_2) = a\mathbb{E}X_1 + b\mathbb{E}X_2. \quad (15)$$

Außerdem gilt für jede Zufallsvariable Y mit $\mathbb{E}(Y^2) < \infty$

$$\text{Var} Y = \mathbb{E}(Y^2) - (\mathbb{E}Y)^2 \quad (16)$$

und für beliebige $a, b \in \mathbb{R}$

$$\text{Var}(aY + b) = a^2 \text{Var} Y. \quad (17)$$

Beweis

- Falls X_1 und X_2 diskrete Zufallsvariablen sind, dann ergibt sich (15) aus (13), wobei ähnlich wie im Beweis von Lemma 4.2 vorgegangen werden kann.

- Wir führen den Beweis von (15) nur für den Fall, daß der Zufallsvektor (X_1, X_2) absolutstetig ist. Für $\varphi(x_1, x_2) = a x_1 + b x_2$ ergibt sich dann aus (14), daß

$$\begin{aligned}
 \mathbb{E}(a X_1 + b X_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (a x_1 + b x_2) f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a x_1 f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b x_2 f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\
 &= \int_{-\infty}^{\infty} a x_1 \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_2 dx_1 + \int_{-\infty}^{\infty} b x_2 \int_{-\infty}^{\infty} f_{(X_1, X_2)}(x_1, x_2) dx_1 dx_2 \\
 &= \int_{-\infty}^{\infty} a x_1 f_{X_1}(x_1) dx_1 + \int_{-\infty}^{\infty} b x_2 f_{X_2}(x_2) dx_2 \\
 &= a \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 + b \int_{-\infty}^{\infty} x_2 f_{X_2}(x_2) dx_2 \\
 &= a \mathbb{E} X_1 + b \mathbb{E} X_2.
 \end{aligned}$$

- Durch Ausmultiplizieren der quadratischen Form in (11) ergibt sich

$$\begin{aligned}
 \text{Var } Y &= \mathbb{E} \left((Y - \mathbb{E} Y)^2 \right) \\
 &= \mathbb{E} \left(Y^2 - 2Y\mathbb{E} Y + (\mathbb{E} Y)^2 \right) \\
 &= \mathbb{E} (Y^2) - 2\mathbb{E} Y \mathbb{E} Y + (\mathbb{E} Y)^2 \\
 &= \mathbb{E} (Y^2) - (\mathbb{E} Y)^2,
 \end{aligned}$$

wobei sich die vorletzte Gleichung aus (15) ergibt. Damit ist (16) bewiesen.

- Für $X_1 = Y$ und $X_2 = b$ ergibt sich aus (15), daß

$$\begin{aligned}
 \text{Var}(aY + b) &= \mathbb{E} (aY + b - \mathbb{E}(aY + b))^2 \\
 &= \mathbb{E} (aY + b - \mathbb{E}(aY) - b)^2 \\
 &= \mathbb{E} (aY - \mathbb{E}(aY))^2 \\
 &= \mathbb{E} ((aY)^2) - (\mathbb{E}(aY))^2 \\
 &= a^2 \mathbb{E} (Y^2) - a^2 (\mathbb{E} Y)^2 \\
 &= a^2 \text{Var } Y,
 \end{aligned}$$

wobei sich die letzten drei Gleichungen aus (16) bzw. erneut aus (15) ergeben. Damit ist auch (17) bewiesen.

Korollar 4.6 Sei $n \in \mathbb{N}$ fest vorgegeben, und seien X_1, \dots, X_n Zufallsvariable mit $\mathbb{E}|X_1| < \infty, \dots, \mathbb{E}|X_n| < \infty$. Dann gilt für beliebige Konstanten $a_1, \dots, a_n \in \mathbb{R}$

$$\mathbb{E}(a_1 X_1 + \dots + a_n X_n) = a_1 \mathbb{E} X_1 + \dots + a_n \mathbb{E} X_n. \quad (18)$$

Beweis folgt aus (15) mittels vollständiger Induktion

Beispiel (wiederholtes Würfeln)

- Betrachten den Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(\Omega), P)$, der in Abschnitt 4.1.1 eingeführt worden ist.
- Betrachten die Zufallsvariablen $X_1, X_2, \dots : \Omega \rightarrow \{1, 2, \dots, 6\}$, wobei X_i die zufällige Augenzahl ist, die beim i -ten Würfeln erzielt wird.
- Dann gilt $\mathbb{E} X_i = \sum_{j=1}^6 j P(X_i = j) = \frac{1}{6} \sum_{j=1}^6 j = 3.5$.

- Aus Korollar 4.6 ergibt sich dann für den Erwartungswert $\mathbb{E} Y_n$ der mittleren Augenzahl $Y_n = n^{-1}(X_1 + \dots + X_n)$ bei n -maligem Würfeln

$$\begin{aligned}\mathbb{E} Y_n &= \mathbb{E} \left(\frac{1}{n} (X_1 + \dots + X_n) \right) \\ &= \frac{1}{n} (\mathbb{E} X_1 + \dots + \mathbb{E} X_n) \\ &= \frac{1}{n} n 3.5 = 3.5.\end{aligned}$$

4.2 Gemischte Momente

Neben Erwartungswert, Varianz bzw. höheren Momenten einer (einzelnen) Zufallsvariablen werden außerdem sogenannte *gemischte Momente* für Familien von (endlich vielen) Zufallsvariablen betrachtet. Sie sind ein wichtiges Hilfsmittel, um Zusammenhänge zwischen zwei oder mehreren Zufallsvariablen zu quantifizieren.

4.2.1 Multiplikationsformel und Kovarianz

Beachte

- Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariable mit $\mathbb{E}(|X_i|^n) < \infty$ für jedes $i \in \{1, \dots, n\}$.
- Man kann zeigen, daß dann $\mathbb{E}|X_1 \dots X_n| < \infty$.
- Der Erwartungswert $\mathbb{E}(X_1 \dots X_n)$ des Produktes $X_1 \dots X_n$ heißt *gemischtes Moment* der Zufallsvariablen X_1, \dots, X_n .

Zunächst diskutieren wir die folgende *Multiplikationsformel* für den Erwartungswert des Produktes von n unabhängigen Zufallsvariablen.

Theorem 4.7 Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariable mit $\mathbb{E}(|X_i|^n) < \infty$ für jedes $i \in \{1, \dots, n\}$. Falls X_1, \dots, X_n unabhängig sind, dann gilt

$$\mathbb{E} \left(\prod_{i=1}^n X_i \right) = \prod_{i=1}^n \mathbb{E} X_i. \quad (19)$$

Beweis

- Wir zeigen die Gültigkeit von (19) nur für die beiden (grundlegenden) Fälle, daß X diskret bzw. absolutstetig ist.
- Die Komponenten des Zufallsvektors $X = (X_1, \dots, X_n)$ seien unabhängige Zufallsvariable.
- Falls X diskret ist, dann gibt es für jedes $i \in \{1, \dots, n\}$ eine abzählbare Menge C_i mit $P(X_i \in C_i) = 1$.
- Darüber hinaus gilt $P(X \in C) = 1$, wobei $C = C_1 \times \dots \times C_n$ eine abzählbare Teilmenge des R^n ist.

- Aus (13) ergibt sich dann für $\varphi(x) = x_1 \cdot \dots \cdot x_n$

$$\begin{aligned}
\mathbb{E} \left(\prod_{i=1}^n X_i \right) &= \sum_{x \in C} \varphi(x) P(X = x) \\
&= \sum_{x_1 \in C_1} \dots \sum_{x_n \in C_n} x_1 \cdot \dots \cdot x_n P(X_1 = x_1, \dots, X_n = x_n) \\
&\stackrel{(3.21)}{=} \sum_{x_1 \in C_1} \dots \sum_{x_n \in C_n} x_1 \cdot \dots \cdot x_n P(X_1 = x_1) \dots P(X_n = x_n) \\
&= \left(\sum_{x_1 \in C_1} x_1 P(X_1 = x_1) \right) \dots \left(\sum_{x_n \in C_n} x_n P(X_n = x_n) \right) \\
&= \prod_{k=1}^n \mathbb{E} X_k.
\end{aligned}$$

- Falls X absolutstetig ist, dann ergibt sich (19) auf völlig analoge Weise aus (14).

Korollar 4.8 Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ unabhängige Zufallsvariable mit $\mathbb{E}(X_i^2) < \infty$ für jedes $i \in \{1, \dots, n\}$. Dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \text{Var} X_1 + \dots + \text{Var} X_n. \quad (20)$$

Beweis

- Wir zeigen die Gültigkeit von (20) zunächst für den Fall $n = 2$.
- Aus (15), (16) und (19) ergibt sich, daß

$$\begin{aligned}
\text{Var}(X_1 + X_2) &\stackrel{(16)}{=} \mathbb{E}((X_1 + X_2)^2) - (\mathbb{E}(X_1 + X_2))^2 \\
&\stackrel{(15)}{=} \left(\mathbb{E}(X_1^2) + 2\mathbb{E}(X_1 X_2) + \mathbb{E}(X_2^2) \right) - \left((\mathbb{E} X_1)^2 + 2\mathbb{E} X_1 \mathbb{E} X_2 + (\mathbb{E} X_2)^2 \right) \\
&\stackrel{(19)}{=} \mathbb{E}(X_1^2) - (\mathbb{E} X_1)^2 + \mathbb{E}(X_2^2) - (\mathbb{E} X_2)^2 \\
&\stackrel{(16)}{=} \text{Var} X_1 + \text{Var} X_2.
\end{aligned}$$

- Für beliebiges $n \in \mathbb{N}$ ergibt sich die Gültigkeit von (20) mittels vollständiger Induktion.

Wir diskutieren nun Eigenschaften des gemischten Momentes $\mathbb{E}(X_1 X_2)$ von zwei beliebigen (nicht notwendig unabhängigen) Zufallsvariablen X_1, X_2 .

In diesem Zusammenhang führen wir zunächst die Begriffe der Kovarianz und des Korrelationskoeffizienten ein.

Definition 4.9 Seien X_1, X_2 beliebige Zufallsvariable mit $\mathbb{E}(X_i^2) < \infty$ für $i = 1, 2$.

- Der Erwartungswert $\mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2))$ heißt die *Kovarianz* von X_1 und X_2 . Schreibweise: $\text{Cov}(X_1, X_2) = \mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2))$
- X_1, X_2 heißen *unkorreliert*, falls $\text{Cov}(X_1, X_2) = 0$.
- Falls $\text{Var} X_1 > 0$ und $\text{Var} X_2 > 0$, dann heißt die Größe

$$\varrho(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var} X_1 \cdot \text{Var} X_2}} \quad (21)$$

der *Korrelationskoeffizient* von X_1, X_2 .

Beachte

- Es ist klar, daß Kovarianz und Korrelationskoeffizient die folgende *Symmetrieeigenschaft* besitzt:

$$\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1), \quad \varrho(X_1, X_2) = \varrho(X_2, X_1). \quad (22)$$

- Außerdem gilt

$$\text{Cov}(X, X) = \text{Var } X, \quad \varrho(X, X) = 1. \quad (23)$$

Darüber hinaus gelten weitere nützliche Rechenregeln und Abschätzungen für Kovarianz bzw. Korrelationskoeffizient.

Theorem 4.10 Seien X_1, X_2 beliebige Zufallsvariable mit $\mathbb{E}(X_i^2) < \infty$ für $i = 1, 2$. Dann gilt

$$\text{Cov}(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2 \quad (24)$$

und für beliebige Zahlen $a, b, c, d \in \mathbb{R}$

$$\text{Cov}(aX_1 + b, cX_2 + d) = a c \text{Cov}(X_1, X_2). \quad (25)$$

Außerdem gilt die *Cauchy-Schwarzsche Ungleichung*

$$|\mathbb{E}(X_1, X_2)| \leq \sqrt{\mathbb{E}(X_1^2) \mathbb{E}(X_2^2)}, \quad (26)$$

und

$$|\text{Cov}(X_1, X_2)| \leq \sqrt{\text{Var } X_1 \text{Var } X_2}. \quad (27)$$

Beweis

- Die Formel (24) ergibt sich unmittelbar aus (15), denn es gilt

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{E}((X_1 - \mathbb{E} X_1)(X_2 - \mathbb{E} X_2)) \\ &= \mathbb{E}(X_1 X_2 - X_1 \mathbb{E} X_2 - X_2 \mathbb{E} X_1 + \mathbb{E} X_1 \mathbb{E} X_2) \\ &\stackrel{(15)}{=} \mathbb{E}(X_1 X_2) - \mathbb{E} X_1 \mathbb{E} X_2. \end{aligned}$$

- Die Formel (25) ergibt sich durch eine ähnliche einfache Rechnung aus (15).
- Wir zeigen nun die Gültigkeit der Ungleichung (26).
- Falls $\mathbb{E}(X_1^2) = 0$, dann gilt $P(X_1 = 0) = 1$ und somit auch $\mathbb{E}(X_1 X_2) = 0 \leq \sqrt{\mathbb{E}(X_1^2) \mathbb{E}(X_2^2)}$.
- Sei jetzt $\mathbb{E}(X_1^2) > 0$. Dann gilt für jede Zahl $a \in \mathbb{R}$

$$\begin{aligned} 0 &\leq \mathbb{E}((a X_1 + X_2)^2) = \mathbb{E}(a^2 X_1^2 + 2a X_1 X_2 + X_2^2) \\ &= a^2 \mathbb{E}(X_1^2) + 2a \mathbb{E}(X_1 X_2) + \mathbb{E}(X_2^2). \end{aligned}$$

- Durch beidseitige Multiplikation mit $\mathbb{E}(X_1^2)$ bzw. quadratische Ergänzung ergibt sich hieraus, daß

$$\begin{aligned} 0 &\leq a^2 (\mathbb{E}(X_1^2))^2 + 2a \mathbb{E}(X_1^2) \mathbb{E}(X_1 X_2) + \mathbb{E}(X_1^2) \mathbb{E}(X_2^2) \\ &= (a \mathbb{E}(X_1^2) + \mathbb{E}(X_1 X_2))^2 + \mathbb{E}(X_1^2) \mathbb{E}(X_2^2) - (\mathbb{E}(X_1 X_2))^2. \end{aligned}$$

- Hieraus folgt (26) für $a = -\mathbb{E}(X_1 X_2) / \mathbb{E}(X_1^2)$.

- Die Gültigkeit von (27) ergibt sich unmittelbar aus (26), wenn in (26) die Zufallsvariablen X_1 bzw. X_2 durch $X_1 - \mathbb{E} X_1$ bzw. $X_2 - \mathbb{E} X_2$ ersetzt werden.

Korollar 4.11

1. Falls X_1 und X_2 unabhängig sind, dann gilt

$$\text{Cov}(X_1, X_2) = 0. \quad (28)$$

d.h., X_1 und X_2 sind unkorreliert.

2. Falls $\text{Var} X_1 > 0$ und $\text{Var} X_2 > 0$, dann gilt

$$-1 \leq \varrho(X_1, X_2) \leq 1. \quad (29)$$

Beweis

- Aus (19) und (24) ergibt sich unmittelbar die Gültigkeit von (28).
- Aus (27) und aus der Definitionsgleichung (21) des Korrelationskoeffizienten ergeben sich die Ungleichungen in (29).

Beachte Die Aussage 1 in Korollar 4.11 läßt sich *nicht* umkehren, denn aus der Unkorreliertheit zweier Zufallsvariablen X_1 und X_2 folgt im allgemeinen nicht, daß X_1 und X_2 unabhängig sind.

Beispiel (zweimaliger Münzwurf)

- Seien $Y_1, Y_2 : \Omega \rightarrow \{0, 1\}$ zwei unabhängige (und identisch verteilte) Zufallsvariablen, die nur die beiden Werte 0 oder 1 annehmen können, mit

$$P(Y_i = j) = \begin{cases} \frac{1}{2}, & \text{falls } j = 1, \\ \frac{1}{2}, & \text{falls } j = 0, \end{cases}$$

für $i = 1, 2$.

- Man kann sich leicht überlegen, daß dann die Zufallsvariablen $X_1 = Y_1 + Y_2$ und $X_2 = Y_1 - Y_2$ zwar unkorreliert, jedoch nicht unabhängig sind.
- Denn es gilt $\mathbb{E} X_1 = 1$, $\mathbb{E} X_2 = 0$ und $\mathbb{E}(X_1 X_2) = 0$. Andererseits gilt

$$P(X_1 = 0, X_2 = 0) = \frac{1}{4} \neq \frac{1}{4} \frac{1}{2} = P(X_1 = 0)P(X_2 = 0).$$

4.2.2 Linearer Zusammenhang von Zufallsvariablen

Die Korrelation $\varrho(X_1, X_2)$ zweier Zufallsvariablen X_1 und X_2 mit $0 < \text{Var} X_1, \text{Var} X_2 < \infty$ kann man als Grad ihres linearen (stochastischen) „Zusammenhanges“ auffassen.

Eine genauere Formulierung dieses Phänomens liefert

Theorem 4.12 Seien $a, b \in \mathbb{R}$ beliebige Zahlen, und X_1, X_2 seien Zufallsvariable mit $0 < \text{Var} X_1, \text{Var} X_2 < \infty$.

1. Die erwartete quadratische Abweichung $\mathbb{E}((X_2 - (a X_1 + b))^2)$ zwischen den Zufallsvariablen X_2 und $a X_1 + b$ ist minimal, wenn a und b wie folgt gewählt werden:

$$a = \varrho(X_1, X_2) \frac{\sqrt{\text{Var} X_2}}{\sqrt{\text{Var} X_1}}, \quad b = \mathbb{E} X_2 - a \mathbb{E} X_1. \quad (30)$$

2. Insbesondere gilt $\mathbb{E}((X_2 - (aX_1 + b))^2) = 0$, d.h. $P(X_2 = aX_1 + b) = 1$ genau dann, wenn

$$|\varrho(X_1, X_2)| = 1.$$

Beweis

- Betrachten die Zufallsvariable $Y = X_2 - (aX_1 + b)$.
- Wegen (17) hängt dann $\text{Var } Y$ nicht von b ab.
- Deshalb kann man bei der Minimierung von

$$\mathbb{E}(Y^2) = (\mathbb{E}Y)^2 + \text{Var } Y$$

zunächst $(\mathbb{E}Y)^2$ für jedes feste a durch die entsprechende Wahl von b minimieren.

- Es ist klar, daß $(\mathbb{E}Y)^2$ für $b = \mathbb{E}X_2 - a\mathbb{E}X_1$ minimal ist, weil dann $\mathbb{E}Y = 0$ gilt.
- Es ist nun noch a so zu bestimmen, daß

$$\begin{aligned} \text{Var } Y &= \mathbb{E}\left(\left((X_2 - \mathbb{E}X_2) - a(X_1 - \mathbb{E}X_1)\right)^2\right) \\ &= \text{Var } X_2 - 2a \text{Cov}(X_1, X_2) + a^2 \text{Var } X_1 \\ &= \text{Var } X_1 \left(a - \frac{\text{Cov}(X_1, X_2)}{\text{Var } X_1}\right)^2 + \text{Var } X_2 \left(1 - \frac{(\text{Cov}(X_1, X_2))^2}{\text{Var } X_1 \text{Var } X_2}\right) \end{aligned}$$

minimal wird, wobei sich die letzte Gleichung durch quadratische Ergänzung ergibt.

- Hieraus folgt, daß $\text{Var } Y$ minimal ist, falls

$$a = \frac{\text{Cov}(X_1, X_2)}{\text{Var } X_1} = \varrho(X_1, X_2) \frac{\sqrt{\text{Var } X_2}}{\sqrt{\text{Var } X_1}}.$$

- Außerdem folgt hieraus, daß $\mathbb{E}(Y^2) = 0$ genau dann, wenn $|\varrho(X_1, X_2)| = 1$.

4.2.3 Erwartungswertvektor und Kovarianzmatrix

Wir zeigen zunächst, wie der Begriff der Kovarianz genutzt werden kann, um das in Korollar 4.8 angegebene Additionstheorem (20) für Varianzen zu verallgemeinern.

Theorem 4.13 Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariable mit $\mathbb{E}(X_i^2) < \infty$ für jedes $i \in \{1, \dots, n\}$. Dann gilt

$$\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var } X_i + 2 \sum_{1 \leq j < k \leq n} \text{Cov}(X_j, X_k). \quad (31)$$

Falls die Zufallsvariablen X_1, \dots, X_n paarweise unkorreliert sind, dann gilt insbesondere

$$\text{Var}(X_1 + \dots + X_n) = \text{Var } X_1 + \dots + \text{Var } X_n. \quad (32)$$

Beweis

- Wir betrachten zunächst den Fall $n = 2$.
- Dann ergibt sich sofort aus dem Beweis von Korollar 4.8, daß

$$\text{Var}(X_1 + X_2) = \text{Var } X_1 + \text{Var } X_2 + 2 \text{Cov}(X_1, X_2).$$

- Für beliebiges $n \in \mathbb{N}$ ergibt sich die Gültigkeit von (31) nun mittels vollständiger Induktion.

- Die Gleichung (32) ergibt sich unmittelbar aus (28) und (31).

Beachte Neben den Erwartungswerten $\mathbb{E} X_1, \dots, \mathbb{E} X_n$ und den Varianzen $\text{Var} X_1, \dots, \text{Var} X_n$ sind die in (31) auftretenden Kovarianzen $\{\text{Cov}(X_i, X_j), 1 \leq i < j \leq n\}$ wichtige Charakteristiken des Zufallsvektors $X = (X_1, \dots, X_n)$.

Dies führt zu den folgenden Begriffsbildungen.

Definition 4.14 Seien $X_1, \dots, X_n : \Omega \rightarrow \mathbb{R}$ beliebige Zufallsvariable mit $\mathbb{E}(X_i^2) < \infty$ für jedes $i \in \{1, \dots, n\}$.

1. Der Vektor $(\mathbb{E} X_1, \dots, \mathbb{E} X_n)$ heißt der *Erwartungswertvektor* des Zufallsvektors $X = (X_1, \dots, X_n)$.
Schreibweise: $\mathbb{E} X = (\mathbb{E} X_1, \dots, \mathbb{E} X_n)$
2. Die $n \times n$ -Matrix

$$\text{Cov } X = \begin{pmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix} = (\text{Cov}(X_i, X_j))_{ij}$$

heißt die *Kovarianzmatrix* von $X = (X_1, \dots, X_n)$.

Theorem 4.15 Die Kovarianzmatrix $\text{Cov } X$ ist

1. *symmetrisch*, d.h., für beliebige $i, j \in \{1, \dots, n\}$ gilt

$$\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i). \quad (33)$$

2. *nichtnegativ definit*, d.h., für jedes $x \in \mathbb{R}^n$ gilt

$$x^\top \text{Cov } X x \geq 0, \quad (34)$$

wobei x^\top ist der zu x transponierte Vektor ist.

Beweis

- Die Symmetrieeigenschaft (33) ergibt sich unmittelbar aus der Definition 4.9 der Kovarianz.
- Die Ungleichung (34) ergibt sich durch eine einfache Rechnung aus (15) und aus der Definition 4.9 der Kovarianz.

Beachte Die Matrix $\text{Cov } X$ heißt *positiv definit*, falls

$$x^\top \text{Cov } X x > 0 \quad (35)$$

für jedes $x \in \mathbb{R}^n$ mit $x \neq 0$.

Beispiel (zweidimensionale Normalverteilung)

- Betrachten eine (weitere) Verallgemeinerung der zweidimensionalen Normalverteilung, die in Abschnitt 3.5.3 eingeführt wurde.
- Seien $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ und $\varrho \in [0, 1)$ beliebige, jedoch fest vorgegebene Zahlen.
- Sei $X = (X_1, X_2)$ ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)} \left(\left(\frac{x_1-\mu_1}{2\sigma_1}\right)^2 - 2\varrho\left(\frac{x_1-\mu_1}{2\sigma_1}\right)\left(\frac{x_2-\mu_2}{2\sigma_2}\right) + \left(\frac{x_2-\mu_2}{2\sigma_2}\right)^2 \right)\right). \quad (36)$$

- Dann heißt X *normalverteilt* mit dem Erwartungswertvektor $\mathbb{E} X = (\mu_1, \mu_2)$ und der Kovarianzmatrix

$$\text{Cov } X = \begin{pmatrix} \sigma_1^2 & \sigma_1 \sigma_2 \varrho \\ \sigma_1 \sigma_2 \varrho & \sigma_2^2 \end{pmatrix}. \quad (37)$$

Beachte

- Die Verteilung eines normalverteilten Zufallsvektors $X = (X_1, X_2)$ wird *eindeutig* durch den Erwartungswertvektor $\mathbb{E} X = (\mu_1, \mu_2)$ und die Kovarianzmatrix $\text{Cov } X$ bestimmt. Für beliebige (nicht normalverteilte) Zufallsvektoren gilt diese Eindeutigkeitsaussage jedoch im allgemeinen nicht.
- Es gilt $\varrho = \varrho(X_1, X_2)$. Außerdem ergibt sich aus (36), daß $f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ genau dann, wenn $\varrho = 0$. Die Komponenten X_1, X_2 eines normalverteilten Zufallsvektors $X = (X_1, X_2)$ sind also genau dann unabhängig, wenn sie unkorreliert sind.
- Weil $\varrho < 1$ vorausgesetzt wird, ist die Determinante der Kovarianzmatrix $\text{Cov } X$ nicht Null, d.h., die Matrix $\text{Cov } X$ ist positiv definit und invertierbar. Man kann sich deshalb leicht überlegen, daß die in (36) gegebenen Dichte f_X von X auch wie folgt dargestellt werden kann: Es gilt

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^2 \frac{1}{\sqrt{\det K}} \exp\left(-\frac{1}{2}(x - \mu)^\top K^{-1}(x - \mu)\right) \quad (38)$$

für jedes $x = (x_1, x_2) \in \mathbb{R}^2$, wobei $\mu = (\mu_1, \mu_2)$ und K^{-1} die inverse Matrix zu der in (37) gegebenen Kovarianzmatrix $K = \text{Cov } X$ bezeichnet.

Schreibweise: $X \sim N(\mu, K)$.

- Man kann sich leicht überlegen, daß die Randverteilungen von $X \sim N(\mu, K)$ (eindimensionale) Normalverteilungen sind mit $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1, 2$.
- Manchmal betrachtet man auch Zufallsvektoren $X = (X_1, X_2)$, deren Komponenten X_1, X_2 normalverteilt sind mit $|\varrho(X_1, X_2)| = 1$. Aus Theorem 4.12 folgt dann, daß $P(X_2 = aX_1 + b) = 1$ für ein Zahlenpaar $a, b \in \mathbb{R}$. In diesem Fall ist der Zufallsvektor $X = (X_1, X_2)$ *nicht absolutstetig*, obwohl seine Komponenten diese Eigenschaft besitzen.

Für Zufallsvektoren mit einer beliebigen Dimension $n \in \mathbb{N}$ kann man den Begriff der n -dimensionalen Normalverteilung einführen, indem man eine zu (38) analoge Dichte-Formel betrachtet.

Definition 4.16

- Sei $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ ein beliebiger Vektor, und sei K eine symmetrische und positiv definite $n \times n$ -Matrix.
- Sei $X = (X_1, \dots, X_n)$ ein absolutstetiger Zufallsvektor mit der gemeinsamen Dichte

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{1}{\sqrt{\det K}} \exp\left(-\frac{1}{2}(x - \mu)^\top K^{-1}(x - \mu)\right) \quad (39)$$

für jedes $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

- Man sagt dann, daß $X = (X_1, \dots, X_n)$ *normalverteilt* ist mit dem Erwartungswertvektor $\mu = (\mu_1, \dots, \mu_n)$ und der Kovarianzmatrix K .

4.3 Abschätzungen und Grenzwertsätze

In vielen Fällen läßt sich die Wahrscheinlichkeit $P(A)$ von interessierenden Ereignissen $A \in \mathcal{F}$ nicht in geschlossenen Formeln ausdrücken.

Manchmal ist es jedoch möglich, Ungleichungen herzuleiten, um (obere) Schranken, d.h. Abschätzungen für $P(A)$ zu erhalten. Oft ist es auch nützlich, das asymptotische (Grenz-) Verhalten der Wahrscheinlichkeit $P(A)$ zu kennen, wenn bestimmte Modellparameter unendlich groß bzw. klein werden.

4.3.1 Tschebyschewsche Ungleichung

In diesem Abschnitt wird die sogenannte *Tschebyschewsche Ungleichung* diskutiert.

Sie liefert eine obere Schranke für die Wahrscheinlichkeit, daß die Abweichungen $|X(\omega) - \mathbb{E} X|$ der Werte $X(\omega)$ einer Zufallsvariablen X von ihrem Erwartungswert $\mathbb{E} X$ einen vorgegebenen Schwellenwert $\varepsilon > 0$ überschreiten.

Theorem 4.17 Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann gilt für jedes $\varepsilon > 0$

$$P(|X - \mathbb{E} X| > \varepsilon) \leq \frac{\text{Var } X}{\varepsilon^2}. \quad (40)$$

Beweis

- Wir zeigen die Gültigkeit von (40) nur für die beiden (grundlegenden) Fälle, daß X diskret bzw. absolutstetig ist.
- Falls X diskret ist mit $P(X \in C) = 1$ für eine abzählbare Menge $C \subset \mathbb{R}$, dann ergibt sich aus (9) für $\varphi(x) = (x - \mathbb{E} X)^2$, daß für jedes $\varepsilon > 0$

$$\begin{aligned} \text{Var } X &= \sum_{x \in C} (x - \mathbb{E} X)^2 P(X = x) \\ &\geq \sum_{x \in C: |x - \mathbb{E} X| > \varepsilon} (x - \mathbb{E} X)^2 P(X = x) \\ &\geq \sum_{x \in C: |x - \mathbb{E} X| > \varepsilon} \varepsilon^2 P(X = x) \\ &= \varepsilon^2 \sum_{x \in C: |x - \mathbb{E} X| > \varepsilon} P(X = x) \\ &= \varepsilon^2 P\left(\bigcup_{x \in C} \{X = x, |x - \mathbb{E} X| > \varepsilon\}\right) \\ &= \varepsilon^2 P(|X - \mathbb{E} X| > \varepsilon). \end{aligned}$$

- Falls X absolutstetig ist mit der Dichte f_X , dann ergibt sich auf analoge Weise aus (10), daß für jedes $\varepsilon > 0$

$$\begin{aligned} \text{Var } X &= \int_{-\infty}^{\infty} (x - \mathbb{E} X)^2 f_X(x) dx \\ &\geq \int_{x \in \mathbb{R}: |x - \mathbb{E} X| > \varepsilon} (x - \mathbb{E} X)^2 f_X(x) dx \\ &\geq \int_{x \in C: |x - \mathbb{E} X| > \varepsilon} \varepsilon^2 f_X(x) dx \\ &= \varepsilon^2 \int_{x \in C: |x - \mathbb{E} X| > \varepsilon} f_X(x) dx \\ &= \varepsilon^2 P(|X - \mathbb{E} X| > \varepsilon). \end{aligned}$$

Korollar 4.18 Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}(X^2) < \infty$. Dann gilt $\text{Var } X = 0$ genau dann, wenn

$$P(X = \mathbb{E} X) = 1. \quad (41)$$

Beweis

- Falls $\text{Var } X = 0$, dann ergibt sich aus (40), daß

$$P(|X - \mathbb{E} X| > \varepsilon) = 0$$

für jedes $\varepsilon > 0$.

- Außerdem kann man sich leicht überlegen, daß dann

$$1 - P(X = \mathbb{E} X) = P(|X - \mathbb{E} X| > 0) = \lim_{\varepsilon \rightarrow 0} P(|X - \mathbb{E} X| > \varepsilon) = 0.$$

- Dies impliziert (41).
- Andererseits impliziert (41) die Gültigkeit von $P(X^2 = (\mathbb{E} X)^2) = 1$, d.h., X^2 ist eine diskrete Zufallsvariable.
- Aus der Definitionsgleichung (3) für den Erwartungswert diskreter Zufallsvariablen ergibt sich dann, daß $\mathbb{E}(X^2) = (\mathbb{E} X)^2$.
- Wegen (16) folgt hieraus, daß $\text{Var } X = 0$.

Beachte

- Die Tschebyschewsche Ungleichung (40) ist *nicht* an spezielle Annahmen über die Form der Verteilung der Zufallsvariablen X gebunden.
- Der „Preis“ hierfür ist, daß (40) in vielen Fällen zu relativ groben Abschätzungen führt.
- Wenn zusätzliche Annahmen über die Verteilung von X gemacht werden, dann lassen sich genauere Abschätzungen herleiten bzw. die Wahrscheinlichkeit $P(|X - \mathbb{E} X| > \varepsilon)$ läßt sich explizit bestimmen.

Beispiele

1. fehlerbehaftete Messungen

- Von einem Meßgerät sei bekannt, daß die Meßergebnisse fehlerbehaftet sind.
- Die n -te Messung einer (unbekannten) Größe $\mu \in \mathbb{R}$ liefere den Wert $\mu + X_n(\omega)$ für $\omega \in \Omega$.
- Die Meßfehler X_1, X_2, \dots seien unabhängige und identisch verteilte Zufallsvariable.
- Über die Verteilung von X_n sei lediglich bekannt, daß

$$\mathbb{E} X_n = 0 \quad \text{und} \quad \text{Var } X_n = 1. \quad (42)$$

- Es soll nun die Frage diskutiert werden, wieviele Messungen erforderlich sind, um mit Hilfe der Tschebyschewschen Ungleichung (40) schlußfolgern zu können, daß das arithmetische Mittel

$$Y_n = \frac{1}{n} \sum_{i=1}^n (\mu + X_i)$$

der zufälligen Meßwerte $\mu + X_i$ höchstens mit Wahrscheinlichkeit 0.1 um mehr als 1 vom „wahren“, jedoch unbekanntem Wert μ abweicht.

- Aus den Korollaren 4.6. und 4.8 ergibt sich, daß

$$\mathbb{E} Y_n = \mu \quad \text{und} \quad \text{Var } Y_n = n^{-1}.$$

- Hieraus und aus der Tschebyschewschen Ungleichung (40) ergibt sich, daß

$$P(|Y_n - \mu| > 1) \leq \frac{1}{n}.$$

- Es gilt also $P(|Y_n - \mu| > 1) \leq 0.1$, falls $n^{-1} \leq 0.1$.

- Aus diesen Überlegungen folgt, daß die obengenannten Genauigkeitsvorgaben erfüllt sind, falls $n \geq 10$ Messungen durchgeführt werden.

2. normalverteilte Meßfehler

- Es wird nun zusätzlich angenommen, daß die Meßfehler normalverteilt sind, d.h. $X_n \sim N(0, 1)$.
- Man kann zeigen, daß dann $Y_n \sim N(\mu, 1/n)$ bzw. $\sqrt{n}(Y_n - \mu) \sim N(0, 1)$, vgl. das Beispiel in Abschnitt 3.6.2 bzw. den Kommentar nach Theorem 3.22.
- Es gilt also

$$\begin{aligned} P(|Y_n - \mu| > 1) &= P(\sqrt{n}|Y_n - \mu| > \sqrt{n}) \\ &= P(\sqrt{n}(Y_n - \mu) < -\sqrt{n}) + P(\sqrt{n}(Y_n - \mu) > \sqrt{n}) \\ &= \Phi(-\sqrt{n}) + (1 - \Phi(\sqrt{n})) \\ &= 2(1 - \Phi(\sqrt{n})), \end{aligned}$$

wobei

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \quad (43)$$

die Verteilungsfunktion der Standardnormalverteilung ist.

- Somit ist $P(|Y_n - \mu| > 1) \leq 0.1$ genau dann erfüllt, wenn $\Phi(\sqrt{n}) \geq 0.95$.
- Dies gilt dann, wenn $\sqrt{n} \geq 1.645$ bzw. $n \geq 3$.

Beachte

- Es gibt keine geschlossene Formel für die Stammfunktion des Integrals in (43).
- Die Werte $\Phi(x)$ der Verteilungsfunktion der Standardnormalverteilung müssen deshalb *numerisch* berechnet werden.
- Für $x \geq 0$ sind die Werte $\Phi(x)$ in Tabelle 1 gegeben.
- Für $x < 0$ erhält man $\Phi(x)$ dann aus der folgenden Symmetrieeigenschaft, die sich unmittelbar aus der Definitionsgleichung (43) ergibt.

Lemma 4.19 Für jedes $x \in \mathbb{R}$ gilt

$$\Phi(x) = 1 - \Phi(-x). \quad (44)$$

4.3.2 Gesetz der großen Zahlen

Wir diskutieren nun zwei allgemeinere Varianten des Gesetzes der großen Zahlen, das bereits in Abschnitt 4.1.1 im Zusammenhang mit dem Beispiel des wiederholten Würfels erwähnt wurde. Dabei

- betrachten wir eine beliebige Folge von Zufallsvariablen $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ mit dem gleichen Erwartungswert $\mu = \mathbb{E} X_i$ für alle $i = 1, 2, \dots$ und
- untersuchen die Frage, unter welchen Bedingungen und in welchem Sinne das arithmetische Mittel

$$Y_n = \frac{1}{n} \sum_{i=1}^n X_i$$

gegen μ strebt, falls n unendlich groß wird.

Aus der Tschebyschewschen Ungleichung (40) ergibt sich

Theorem 4.20 Sei $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ eine Folge von Zufallsvariablen mit dem gleichen Erwartungswert $\mu = \mathbb{E} X_i$ und mit $\mathbb{E}(X_i^2) < \infty$ für alle $i = 1, 2, \dots$.

1. Falls

$$\lim_{n \rightarrow \infty} \text{Var } Y_n = 0, \quad (45)$$

dann gilt für jedes $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(|Y_n - \mu| > \varepsilon) = 0. \quad (46)$$

2. Die Bedingung (45) ist insbesondere dann erfüllt, wenn die Zufallsvariablen X_1, X_2, \dots unabhängig sind mit der gleichen Varianz $\sigma^2 = \text{Var } X_i$ für alle $i = 1, 2, \dots$.

Beweis

- Mit Hilfe der Cauchy-Schwarzschen Ungleichung (26) kann man sich leicht überlegen, daß $\mathbb{E}(Y_n^2) < \infty$ für jedes $n = 1, 2, \dots$ gilt.
- Weil außerdem $\mathbb{E} Y_n = \mu$ für jedes $n = 1, 2, \dots$, ergibt sich (46) unmittelbar aus (40) und (45).
- Die Zufallsvariablen X_1, X_2, \dots seien nun unabhängig mit der gleichen Varianz $\sigma^2 = \text{Var } X_i$ für alle $i = 1, 2, \dots$.
- Dann ergibt sich aus (17) und (20), daß

$$\begin{aligned} \text{Var } Y_n &= \frac{1}{n^2} \sum_{i=1}^n \text{Var } X_i \\ &= \frac{1}{n^2} n \sigma^2 \\ &= \frac{\sigma^2}{n} \rightarrow 0 \end{aligned}$$

für $n \rightarrow \infty$.

Definition 4.21 Man sagt, daß die Zufallsvariablen Y_1, Y_2, \dots *stochastisch* gegen die Zahl $\mu \in \mathbb{R}$ konvergieren, falls (46) für jedes $\varepsilon > 0$ gilt. Schreibweise: $Y_n \xrightarrow{\text{st}} \mu$

Beachte

- Die stochastische Konvergenz wird auch *Konvergenz in Wahrscheinlichkeit* genannt.
- Die Konvergenz (46) des arithmetischen Mittels Y_n gegen den Erwartungswert μ heißt das *schwache Gesetz der großen Zahlen*.

Neben der stochastischen Konvergenz gibt es noch weitere Konvergenzarten von Zufallsvariablen. Insbesondere gilt neben Theorem 4.20 der folgende Grenzwertsatz.

Theorem 4.22 Sei $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E} |X_i| < \infty$ für alle $i = 1, 2, \dots$; $\mu = \mathbb{E} X_i$. Für das arithmetische Mittel $Y_n = n^{-1} \sum_{i=1}^n X_i$ gilt dann

$$P\left(\lim_{n \rightarrow \infty} Y_n = \mu\right) = 1. \quad (47)$$

Beweis Der Beweis von Theorem 4.22 ist tieflegend und geht über den Rahmen dieser einführenden Vorlesung hinaus.

Definition 4.23 Man sagt, daß die Zufallsvariablen Y_1, Y_2, \dots *fast sicher* gegen die Zahl $\mu \in \mathbb{R}$ konvergieren, falls (47) gilt. Schreibweise: $Y_n \xrightarrow{\text{f.s.}} \mu$

Beachte

- Die fast sichere Konvergenz wird auch *Konvergenz mit Wahrscheinlichkeit 1* genannt.
- Man kann zeigen, daß die fast sichere Konvergenz $Y_n \xrightarrow{\text{f.s.}} \mu$ einer Folge von Zufallsvariablen Y_1, Y_2, \dots stets die stochastische Konvergenz $Y_n \xrightarrow{\text{st}} \mu$ impliziert.
- Die Konvergenz (47) des arithmetischen Mittels Y_n gegen den Erwartungswert μ heißt deshalb das *starke Gesetz der großen Zahlen*.
- Schließlich sei noch erwähnt, daß man in Theorem 4.22 (analog zur Situation in Theorem 4.20) die Bedingungen der Unabhängigkeit bzw. der identischen Verteiltheit der Zufallsvariablen X_1, X_2, \dots durch schwächere Bedingungen ersetzen kann, ohne daß dabei die Gültigkeit von (47) verloren geht.

4.3.3 Beispiel (Buffonsches Nadelexperiment)

- Das Buffonsche Nadelexperiment ist ein Beispiel, bei dem das starke Gesetz der großen Zahlen angewendet wird.
- Es ist eine der ersten numerischen Methoden, die auf stochastischen Gesetzmäßigkeiten beruht.
- Der „Erfinder“ ist Georges Louis Leclerc Comte de Buffon (1707–1788).
- Heute sind solche Verfahren unter der Bezeichnung „Monte-Carlo-Simulation“ bekannt.

- Betrachten das System

$$K = \{(x, y) : (x, y) \in \{\dots, -1, 0, 1, \dots\} \times \mathbb{R}\} \subset \mathbb{R}^2$$

von parallelen und äquidistanten (vertikalen) Geraden in der euklidischen Ebene \mathbb{R}^2 .

- Werfen eine Nadel mit der Länge 1 „willkürlich“ in die Ebene \mathbb{R}^2 , wobei mit „willkürlich“ das folgende stochastische Modell gemeint ist.
- Betrachten zwei Zufallsvariable S und T , die die zufällige Lage der Nadel beschreiben, wobei
 - S der (orthogonale) Abstand des Nadelmittelpunktes zur nächsten linksliegenden Nachbargeraden von K ist,
 - T der Winkel ist, den die Nadel zum Lot auf die Geraden von K bildet, und
 - die Zufallsvariablen S und T unabhängig und gleichverteilt seien auf den Intervallen $[0, 1]$ bzw. $[-\pi/2, \pi/2]$.
- Bestimmen die Wahrscheinlichkeit des Ereignisses

$$A = \left\{0 < S < \frac{1}{2} \cos T\right\} \cup \left\{1 - \frac{1}{2} \cos T < S < 1\right\},$$

daß die willkürlich geworfene Nadel eine der Geraden von K schneidet.

- Es gilt

$$\begin{aligned}
 P(A) &= P\left(0 < S < \frac{1}{2} \cos T\right) + P\left(1 - \frac{1}{2} \cos T < S < 1\right) \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} P\left(0 < S < \frac{1}{2} \cos t\right) dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} P\left(1 - \frac{1}{2} \cos t < S < 1\right) dt \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt + \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1}{2} \cos t dt \\
 &= \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \cos t dt = \frac{2}{\pi}.
 \end{aligned}$$

- Aus der Gleichung $P(A) = 2/\pi$ ergibt sich nun eine Methode zur experimentellen Bestimmung der Zahl π , die auf dem Gesetz der großen Zahlen beruht.
- Seien $(S_1, T_1), \dots, (S_n, T_n)$ unabhängige und identisch verteilte Zufallsvektoren (mit der gleichen Verteilung wie (S, T)), die wir als das Ergebnis von n (unabhängig durchgeführten) Nadelexperimenten auffassen.
- Dann sind X_1, X_2, \dots, X_n mit

$$X_i = \begin{cases} 1, & \text{falls } S_i < \frac{1}{2} \cos T_i \text{ oder } 1 - \frac{1}{2} \cos T_i < S_i \\ 0 & \text{sonst} \end{cases}$$

unabhängige und identisch verteilte Zufallsvariable mit dem Erwartungswert $\mathbb{E} X_i = 2/\pi$.

- Aus Theorem 4.22 ergibt sich also, daß das arithmetische Mittel $Y_n = n^{-1} \sum_{i=1}^n X_i$ fast sicher gegen die Zahl $2/\pi$ strebt.
- D.h., für große n ist $2/Y_n$ mit hoher Wahrscheinlichkeit eine gute Näherung der Zahl π .

Beachte Im *Internet* gibt es zahlreiche Seiten, wo dieses Verfahren implementiert worden ist und mittels *JAVA-Applets* auch selbst durchgeführt werden kann, vgl. beispielsweise

- <http://www.mste.uiuc.edu/reese/buffon/buffon.html>

Ein *anderer Algorithmus* zur experimentellen Bestimmung der Zahl π hängt ebenfalls mit einem einfachen geometrischen Sachverhalt zusammen.

- Betrachten das Quadrat

$$B = [-1, 1] \times [-1, 1] \subset \mathbb{R}^2$$

und den Kreis

$$C = \{(x, y) : (x, y) \in B, x^2 + y^2 < 1\}.$$

- Werfen einen Punkt willkürlich in die Menge B .
- D.h., wir betrachten zwei unabhängige Zufallsvariable S und T , die jeweils gleichverteilt auf dem Intervall $[-1, 1]$ sind.

- Bestimmen die Wahrscheinlichkeit des Ereignisses

$$A = \{(S, T) \in C\} = \{S^2 + T^2 < 1\},$$

daß der „zufällige Punkt“ (S, T) in $C \subset B$ liegt.

- Es gilt

$$P(A) = P(S^2 + T^2 < 1) = \dots = \frac{|C|}{|B|} = \frac{\pi}{4},$$

wobei $|B|, |C|$ den Flächeninhalt von B bzw. C bezeichnet.

- Ähnlich wie beim Buffonschen Nadelexperiment ergibt sich nun aus der Gleichung $P(A) = \pi/4$ eine weitere Methode zur experimentellen Bestimmung der Zahl π , die auf dem Gesetz der großen Zahlen beruht und die sich leicht implementieren läßt.
- Seien $(S_1, T_1), \dots, (S_n, T_n)$ unabhängige und identisch verteilte Zufallsvektoren (mit der gleichen Verteilung wie (S, T)), die wir als das Ergebnis von n (unabhängig durchgeführten) Experimenten auffassen.
- Dann sind X_1, X_2, \dots, X_n mit

$$X_i = \begin{cases} 1, & \text{falls } S_i^2 + T_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

unabhängige und identisch verteilte Zufallsvariable mit dem Erwartungswert $\mathbb{E} X_i = \pi/4$.

- Aus Theorem 4.22 ergibt sich also, daß das arithmetische Mittel $Y_n = n^{-1} \sum_{i=1}^n X_i$ fast sicher gegen die Zahl $\pi/4$ strebt.
- D.h., für große n ist $4Y_n$ mit hoher Wahrscheinlichkeit eine gute Näherung der Zahl π .

Beachte

1. Bei der Implementierung dieser Monte-Carlo-Simulation kann man wie folgt vorgehen.
 - Erzeuge $2n$ auf $[0, 1]$ gleichverteilte Pseudozufallszahlen z_1, \dots, z_{2n} mit einem Zufallszahlengenerator.
 - Setze $s_i = 2z_i - 1$ und $t_i = 2z_{n+i} - 1$ für $i = 1, \dots, n$.
 - Setze

$$x_i = \begin{cases} 1, & \text{falls } s_i^2 + t_i^2 < 1 \\ 0 & \text{sonst} \end{cases}$$

- Berechne $4(x_1 + \dots + x_n)/n$.
2. Ein *JAVA-Applet*, mit dem dieses Simulationsverfahren selbst durchgeführt werden kann, findet man beispielsweise auf der Internet-Seite:
 - <http://www.daimi.aau.dk/~u951581/pi/MonteCarlo/pimc.html>

4.3.4 Zentraler Grenzwertsatz

Neben der stochastischen Konvergenz und der fast sicheren Konvergenz, die in Abschnitt 4.3.2 eingeführt wurden, gibt es noch weitere Konvergenzarten von Zufallsvariablen.

Wir diskutieren nun

- den Begriff *Konvergenz in Verteilung* und in diesem Zusammenhang

- eine weitere Kategorie von Grenzwertsätzen, den sogenannten *zentralen Grenzwertsatz*.
- Dabei wird eine andere Normierung der Summe $X_1 + \dots + X_n$ als beim Gesetz der großen Zahlen betrachtet.
- Während die Normierung $1/n$ beim Gesetz der großen Zahlen zu dem *deterministischen* Grenzwert μ führt, wird nun die (kleinere) Normierung $1/\sqrt{n}$ betrachtet, die zu einem nichtdeterministischen, d.h. zufälligen Grenzwert führt.

In Verallgemeinerung des zentralen Grenzwertsatzes von DeMoivre-Laplace, der bereits in Abschnitt 3.2.3 erwähnt wurde, gilt

Theorem 4.24 Sei $X_1, X_2, \dots : \Omega \rightarrow \mathbb{R}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit $\mathbb{E}(X_i^2) < \infty$ und $\text{Var } X_i > 0$ für alle $i = 1, 2, \dots$; $\mu = \mathbb{E} X_i$, $\sigma^2 = \text{Var } X_i$. Dann gilt für jedes $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad (48)$$

wobei $\Phi(x)$ die Verteilungsfunktion der Standardnormalverteilung ist.

Beweis Der Beweis von Theorem 4.24 ist tieflegend und geht über den Rahmen dieser einführenden Vorlesung hinaus.

Korollar 4.25 Unter den Voraussetzungen von Theorem 4.24 gilt

$$\lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x) \quad (49)$$

für jedes $x \in \mathbb{R}$, und

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b\right) = \Phi(b) - \Phi(a) \quad (50)$$

für beliebige $a, b \in \mathbb{R}$ mit $a \leq b$.

Beweis Die Behauptung (49) ergibt sich aus (48), weil

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < x\right) &= \lim_{n \rightarrow \infty} \lim_{h \downarrow 0} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x - h\right) \\ &= \lim_{h \downarrow 0} \lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq x - h\right) \\ &\stackrel{(48)}{=} \lim_{h \downarrow 0} \Phi(x - h) \\ &= \Phi(x). \end{aligned}$$

Die Behauptung (50) ergibt sich nun aus (48) und (49), denn es gilt

$$\begin{aligned} &\lim_{n \rightarrow \infty} P\left(a \leq \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b\right) \\ &= \lim_{n \rightarrow \infty} \left(P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b\right) - P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < a\right)\right) \\ &= \lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} \leq b\right) - \lim_{n \rightarrow \infty} P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < a\right) \\ &= \Phi(b) - \Phi(a) \end{aligned}$$

Beachte Es ist die folgende Sprechweise üblich: Falls (48) für jedes $x \in \mathbb{R}$ gilt, dann sagt man, daß

$$Y_n = \frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}}$$

in Verteilung gegen die Standardnormalverteilung (bzw. gegen eine $N(0,1)$ -verteilte Zufallsvariable Y) strebt.

Allgemein definiert man den Begriff der Verteilungskonvergenz von Zufallsvariablen wie folgt.

Definition 4.26 Sei $Y, Y_1, Y_2, \dots : \Omega \rightarrow \mathbb{R}$ eine beliebige Folge von Zufallsvariablen. Man sagt, daß die Zufallsvariablen Y_1, Y_2, \dots in Verteilung gegen die Zufallsvariable Y konvergieren, falls

$$\lim_{n \rightarrow \infty} P(Y_n \leq y) = P(Y \leq y) \quad (51)$$

für jedes $y \in \mathbb{R}$ gilt, das ein Stetigkeitspunkt der Verteilungsfunktion F_Y ist, d.h. $\lim_{h \downarrow 0} F_Y(y-h) = F_Y(y)$.
Schreibweise: $Y_n \xrightarrow{d} Y$

Beachte

- Die Formel (51) ist offenbar gleichbedeutend mit

$$\lim_{n \rightarrow \infty} F_{Y_n}(y) = F_Y(y)$$

für jeden Stetigkeitspunkt $y \in \mathbb{R}$ der Verteilungsfunktion F_Y .

- Man kann darüber hinaus zeigen, daß dies wiederum gleichbedeutend ist mit

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g(y) dF_{Y_n}(y) = \int_{-\infty}^{\infty} g(y) dF_Y(y) \quad (52)$$

für jede beschränkte und stetige Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$, wobei die Integrale in (52) sogenannte *Stieltjes-Integrale* sind.

- In diesem Zusammenhang spricht man auch von der *schwachen Konvergenz* der Verteilung von Y_n gegen die Verteilung von Y .

Beispiel (fehlerbehaftete Messungen)

- So wie in dem Beispiel, das bereits in Abschnitt 4.3.1 betrachtet wurde, nehmen wir an, daß die n -te Messung einer (unbekannten) Größe $\mu \in \mathbb{R}$ den Wert $\mu + X_n(\omega)$ liefert für $\omega \in \Omega$.
- Die Meßfehler X_1, X_2, \dots seien unabhängige und identisch verteilte Zufallsvariable.
- Über die Verteilung von X_n sei lediglich bekannt, daß

$$\mathbb{E} X_n = 0 \quad \text{und} \quad \text{Var} X_n = \sigma^2. \quad (53)$$

- Ein Ansatz zur „Schätzung“ der unbekanntes Größe μ ist durch das arithmetische Mittel

$$Y_n = \frac{1}{n} \sum_{i=1}^n (\mu + X_i)$$

der zufälligen Meßwerte $\mu + X_i$ gegeben.

- Mit Hilfe von Korollar 4.25 läßt sich die Wahrscheinlichkeit $P(|Y_n - \mu| > \varepsilon)$, daß der Schätzfehler $|Y_n - \mu|$ größer als ε ist, näherungsweise bestimmen.

- Und zwar gilt für große n

$$\begin{aligned}P(|Y_n - \mu| > \varepsilon) &= P\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| > \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\&= 1 - P\left(\left|\frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}\right| \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\&= 1 - P\left(-\frac{\varepsilon\sqrt{n}}{\sigma} \leq \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}} \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) \\&\stackrel{(50)}{\approx} 1 - \left(\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right) \\&= 2\left(1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right).\end{aligned}$$