Stochastik I

Henning Läuter

Universität Potsdam

Inhaltsverzeichnis

Teil I

Wahrscheinlichkeitstheorie

1 Warum Wahrscheinlichkeitstheorie und Statistik?

- wichtige Rolle in der Entwicklung des Denkens
- nützlich im täglichen Leben
- wichtig für mathematische Erziehung

Wir behandeln

- den praktischen Hintergrund der Wahrscheinlichkeitstheorie (statistische Gesetzmäßigkeiten)
- mathematische Theorie der Wahrscheinlichkeit
- Anwendung der Wahrscheinlichkeitstheorie zur Beschreibung zufälliger Massenerscheinungen
- Auswertung von Experimenten (Statistik)

Experimente:

$$\left.\begin{array}{l}\bullet\ \text{deterministische Ausgänge}\\[1ex] \bullet\ \text{zufällige Ausgänge}\end{array}\right\}\text{gemischte Ausgänge}$$

Experiment hängt häufig von der Genauigkeit des Beobachtens ab:

mechanisches Experiment: schiefe Ebene

$$\begin{array}{lcl}\text{grobe Messung} & \Rightarrow & \text{deterministischer Ausgang}\\ \text{feine Meßmethode} & \Rightarrow & \text{zufällige Schwankungen werden registriert}\end{array}$$

Ursprung der Wahrscheinlichkeitstheorie:

Erklärung von Zufälligkeiten bei Glücksspielen z. B. Kartenmischen (Beziehung zu moderner Technologie: Mischvorgänge in der Chemie u. a.)

$$\left.\begin{array}{ll}\text{Münze werfen:} & \text{Wappen oder Zahl}\\ \text{Urne mit Kugeln:} & \text{Ziehen von weißen Kugeln}\end{array}\right\}\begin{array}{l}\text{Wahrschein-}\\ \text{lichkeiten}\end{array}$$

2 Mathematische Beschreibung von Experimenten mit zufälligem Ausgang

Ω = Raum der Elementarereignisse ω

Münze	:	Ω besteht aus 2 Elementen
Würfel	:	Ω besteht aus 6 Elementen
Lebensdauer von Produkten	:	$\Omega = [0, \infty)$

Ω
- endlich
- unendlich
  - abzählbar
  - überabzählbar

Wahl von Ω wird durch Zweckmäßigkeit bestimmt

Teilmengen von Ω = (zufällige) Ereignisse

$$\left.\begin{array}{l}\text{Vereinigung}\\ \text{Durchschnitt}\\ \text{Differenz}\end{array}\right\}\ \text{von Ereignissen}\quad \begin{array}{ccc}A & \cup & B\\ A & \cap & B\\ A & \setminus & B\end{array}$$

Ω	sicheres Ereignis
$\emptyset$	unmögliches Ereignis
$\Omega \setminus A$	komplementäres Ereignis zu A
$AB = \emptyset$	disjunkte Ereignisse.

2.1 Diskrete Räume der Elementarereignisse

Definition 2.1 *Ω heißt diskret, falls Ω höchstens abzählbar unendlich viele Elemente hat.*

Es sei P eine Funktion : $\Omega \to [0,1]$ mit $\sum_{\omega\in\Omega} \mathsf{P}(\omega) = 1$. Weiter sei $A \in \mathfrak{P}(\Omega)$. Dann definieren wir

$$\mathsf{P}(A) = \sum_{\omega\in A} \mathsf{P}(\omega).$$

Da $\sum\limits_{\omega\in A} \mathsf{P}(\omega)$ absolut konvergiert, ist $\mathsf{P}(A)$ unabhängig von der Reihenfolge der Summanden. $\Rightarrow$ P ist definiert auf den Teilmengen A von Ω.

Definition 2.2 *Eine Funktion $\mathsf{P} : \Omega \Rightarrow [0,1]$ mit $\sum_{\omega\in\Omega} \mathsf{P}(\omega) = 1$ heißt Wahrscheinlichkeit oder Wahrscheinlichkeitsverteilung auf $\mathfrak{P}(\Omega)$.*

$\mathsf{P}(A)$ heißt Wahrscheinlichkeit des Ereignisses A. P ist bestimmt durch die Werte $\mathsf{P}(\omega), \omega \in \Omega$.

Es gilt:

$$
\begin{aligned}
\mathsf{P}(\emptyset) &= 0 \\
\mathsf{P}(\Omega) &= 1 \\
\mathsf{P}(A+B) &= \sum_{\omega \in A \cup B} \mathsf{P}(\omega) \\
&= \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cap B) \\
\mathsf{P}(\bar{A}) &= 1 - \mathsf{P}(A)\,, \qquad \bar{A} \text{ Komplement von } A
\end{aligned}
$$

A, B disjunkt $\Rightarrow \mathsf{P}(A \cap B) = 0$

$$
\begin{aligned}
&A_1, \ldots, A_n \text{ paarweise disjunkt:} \quad & \mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) &= \sum_{k=1}^{n} \mathsf{P}(A_k)\,, \qquad n \le \infty \\
&\text{allgemeine } A_1, \ldots, A_n\text{:} & \mathsf{P}\left(\bigcup A_k\right) &\le \sum \mathsf{P}(A_k)
\end{aligned}
$$

2.2 Das Bernoulli-Schema

Bei einem Vorgang wird beobachtet, ob ein Ereignis A eintritt oder nicht. Einmaliges Beobachten kann beschrieben werden durch

$$\Omega_1 = \{0, 1\}\,,$$

wobei $\mathsf{P}(1) = p$ die Wahrscheinlichkeit ist, daß A eingetreten ist. Aus der Eigenschaft der Wahrscheinlichkeitsverteilung ergibt sich sofort $\mathsf{P}(0) = 1 - p$. Wird dieser Vorgang r-mal beobachtet, so kann dies beschrieben werden durch den Raum

$$\Omega_r = \{(a_1, \ldots, a_r) \,:\, a_i \in \{0,1\}, \quad i = 1, \ldots, r\}\,,$$

d.h. die Elemente ω des Raumes der Elementarereignisse sind r-Tupel, deren Komponenten Nullen oder Einsen sind.
Es wird sich später (Abschnitt 4.3) zeigen, daß es sinnvoll ist, die Wahrscheinlichkeit eines r-Tupels, das aus k Einsen und $r-k$ Nullen besteht (unabhängig davon, an welcher Stelle diese Einsen und Nullen stehen), folgendermaßen zu definieren:

$$\mathsf{P}(\omega) = p^k (1-p)^{r-k}\,, \quad \text{falls } \omega \text{ genau } k \text{ Einsen enthält.}$$

Um zu zeigen, daß das so definierte P eine Wahrscheinlichkeit ist, zeigen wir

(1) $0 \le \mathsf{P}(\omega) \le 1 \quad$ für alle $\omega \in \Omega_r \quad$ und

(2) $\sum_{\omega \in \Omega_r} \mathsf{P}(\omega) = 1\,.$

Die erste Aussage ist trivial.
Die zweite Aussage weisen wir folgendermaßen nach

$$\begin{aligned}
\sum_{\omega\in\Omega_r} \mathsf{P}(\omega) &= \sum_{k=0}^{r} \sum_{\substack{\omega: \\ \omega \text{ enthält genau} \\ k \text{ Einsen}}} \mathsf{P}(\omega) \\
&= \sum_{k=0}^{r} \binom{r}{k} p^k (1-p)^{r-k} \\
&= (p+1-p)^r = 1
\end{aligned}$$

Hierbei wurde benutzt, daß die Anzahl der r-Tupel mit k Einsen (unabhängig von der Stelle, an der die Einsen stehen) $\binom{r}{k}$ ist.

Aus diesen Überlegungen ergibt sich, daß die Wahrscheinlichkeit, bei r-maligem Beobachten des Vorganges genau k-mal A zu beobachten, durch

$$\mathsf{P}(k,r) = \binom{r}{k} p^k (1-p)^{r-k}$$

gegeben ist. Denn dieses Ereignis kann beschrieben werden durch die Vereinigung aller r-Tupel mit genau k Einsen. Alle diese r-Tupel sind unvereinbar (disjunkt), so daß sich nach den oben beschriebenen Gesetzmäßigkeiten ergibt:

$$\mathsf{P}(k,r) = \sum_{\substack{\{\omega:\, \omega \text{ enthält} \\ k \text{ Einsen}\}}} p^k (1-p)^{r-k} = \binom{r}{k} p^k (1-p)^{r-k}$$

Definition 2.3 *Die Funktion $\tilde{\mathsf{P}}$, definiert über $\Omega = \{0, 1, \dots, r\}$ mit*

$$\tilde{\mathsf{P}}(k) = \binom{r}{k} p^k (1-p)^{r-k}, \qquad k \in \Omega$$

heißt Binomialverteilung $B(r,p)$.

Die Binomialverteilung $B(r,p)$ tritt auf, wenn r unabhängige Versuche durchgeführt werden, bei denen nur zwei mögliche Ausgänge eintreten können:

$1 =$ Erfolg
$0 =$ Mißerfolg
p Wahrscheinlichkeit für Erfolg.

3 Wahrscheinlichkeitsräume

Bisher betrachteten wir diskrete Ω, definierten P für beliebige Teilmengen von Ω.
$\Rightarrow$ P war definiert auf $\mathfrak{P}(\Omega)$. Es galt:

$$\mathsf{P}(\Omega) = 1\,, \quad \mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A)$$

$$A_1, \dots, A_n \text{ paarweise disjunkt } \Rightarrow \mathsf{P}\left(\bigcup_1^n A_i\right) = \sum_1^n \mathsf{P}(A_i)$$

Jetzt sei Ω nicht mehr notwendig diskret. Es sei Ω ein beliebiger Raum von Elementarereignissen. $\mathfrak{A}$ heißt Algebra, wenn $\mathfrak{A}$ Menge von Teilmengen von Ω ist, und folgende Eigenschaften erfüllt sind:

$$\begin{aligned}
&\text{A1}: && \Omega \in \mathfrak{A} \\
&\text{A2}: && A, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}, A \cap B \in \mathfrak{A} \\
&\text{A3}: && A \in \mathfrak{A} \Rightarrow \bar{A} \in \mathfrak{A}
\end{aligned}$$

In A 2 braucht nur $A \cup B \in \mathfrak{A}$ gefordert zu werden, denn

$$A, B \in \mathfrak{A} \Rightarrow \bar{A}, \bar{B} \in \mathfrak{A} \Rightarrow \bar{A} \cup \bar{B} \in \mathfrak{A} \Rightarrow \underset{\substack{\parallel \\ A \cap B}}{\overline{\bar{A} \cup \bar{B}}} \in \mathfrak{A}$$

$\mathfrak{A}$ heißt σ-Algebra, falls $\mathfrak{A}$ eine Algebra ist und

A 2': Ist $\{A_n\}$ Folge von Mengen aus $\mathfrak{A}$, so

$$\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}\,, \qquad \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A}\,.$$

gilt. Auch hier ist nur eine der Relationen notwendig. $\mathfrak{A}$ ist abgeschlossen bezüglich der Bildung abzählbarer Durchschnitte und Vereinigungen und des Komplements.

Definition 3.1 $(\Omega, \mathfrak{A})$ *heißt meßbarer Raum, falls* $\mathfrak{A}$ *σ-Algebra über* Ω.

Elemente von $\mathfrak{A}$ heißen (meßbare) Ereignisse.

Beispiel 3.1 Borelsche σ-Algebra $\mathfrak{B}^n$. $\mathfrak{B}^n$ wird erzeugt durch die offenen oder halboffenen oder abgeschlossenen Intervalle des $\mathbb{R}^n$. λ_n bezeichnet das Lebesgue-Maß auf $\mathfrak{B}^n$.

Beispiel 3.2 Ω diskret.
Beh.: $\mathfrak{P}(\Omega)$ ist die von den Elementen aus Ω erzeugte σ-Algebra.

1. $\Omega \in \mathfrak{P}(\Omega)$
2. $\bigcup\limits_{n=1}^{\infty} A_n \in \mathfrak{P}(\Omega) \qquad \text{bei } A_n \in \mathfrak{P}(\Omega)$
3. $A \in \mathfrak{P}(\Omega) \Rightarrow \bar{A} \in \mathfrak{P}(\Omega)$
4. $A \in \mathfrak{P}(\Omega) \Leftrightarrow A = \bigcup \{a_n\}, \qquad a_n \in \Omega$

Daher ist $\mathfrak{P}(\Omega)$ kleinste σ-Algebra.

Beispiel 3.3
$\Omega = [0,1]$. Betrachte die σ-Algebra $\mathfrak{A}$, die aufgespannt wird von einer diskreten Menge von Punkten aus Ω. $\Rightarrow$ $\mathfrak{A}$ enthält nur Mengen vom Lebesgue-Maß 0 oder 1 (magere oder dichte Mengen).

Die mathematische Beschreibung von Experimenten erfolgt durch die Zuordnung:
Experiment $\rightarrow (\Omega, \mathfrak{A})$ = meßbarer Raum;
diese Abbildung bedeutet die mathematische Modellierung des Experiments.

Definition 3.2 *Wahrscheinlichkeitsverteilung auf* $\mathfrak{A}$
Eine Wahrscheinlichkeitsverteilung auf der σ*-Algebra* $\mathfrak{A}$ *eines meßbaren Raumes* $(\Omega, \mathfrak{A})$ *ist eine reellwertige Funktion* P *auf* $\mathfrak{A}$ *mit den Eigenschaften:*

$$\left.\begin{array}{ll} \textit{P1: } \mathsf{P}(A) \geq 0 & \forall\, A \in \mathfrak{A} \\ \textit{P2: } \mathsf{P}(\Omega) = 1 & \end{array}\right\} \begin{array}{l}\textit{natürliche}\\ \textit{Forderungen}\end{array}$$

P3: Ist $\{A_n\}$ *Folge paarweise unvereinbarer Ereignisse, d. h.* $A_i A_j = 0$*,* $i \neq j$*, so gilt:*

$$\mathsf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathsf{P}(A_n)$$

(mathematisch motivierte Forderung)

(σ*-Additivität von* P*)*

Zur Erinnerung: P ist eine nichtnegative, σ-additive Mengenfunktion auf $\mathfrak{A}$ und damit ist P ein Maß. Ein Maß μ heißt endlich, falls $\mu(\Omega) < \infty$. $\Rightarrow$ P ist ein auf 1 normiertes Maß auf $\mathfrak{A}$.
$\Rightarrow$ Bezeichnung: P = Wahrscheinlichkeitsmaß

Beispiel 3.4

1. Sei $\Omega = [0,1]$, $a \in (0,1)$ fest. Die durch $(0,a)$ erzeugte σ-Algebra $\mathfrak{A}$ ist gegeben durch $\emptyset, \Omega$, $(0,a)$, $\{0\} \cup [a,1]$.
   Wählt man $\mathsf{P} = \lambda_1$, so ist P eine Wahrscheinlichkeitsverteilung über $(\Omega, \mathfrak{A})$.

2. Ist $\Omega = \mathbb{R}^1, \mathfrak{A} = \mathfrak{B}^1$, $\mathsf{P} = \lambda_1$, so ist λ_1 keine Wahrscheinlichkeitsverteilung.

3. Es bezeichne X die Anzahl der Wappen beim Wurf von zwei Münzen. Dann ist $\Omega = \{0,1,2\}$, $\mathfrak{A} = \mathfrak{P}(\Omega)$. P ist bestimmt durch $\mathsf{P}(0)$, $\mathsf{P}(1)$, $\mathsf{P}(2)$. Es ist $\mathsf{P}(0) = \mathsf{P}(2) = \frac{1}{4}$, $\mathsf{P}(1) = \frac{1}{2}$.

Definition 3.3 *Eine endliche, nichtnegative, additive Mengenfunktion μ heißt stetig, falls für eine monoton fallende Folge von Ereignissen $\{A_n\}$, $A_{n+1} \subset A_n$ gilt:*

$$\begin{array}{rrcll} & \mu(A_n) & \to & \mu\left(\bigcap\limits_{j=1}^{\infty} A_j\right) & \quad \textit{für } n \to \infty, \\ \textit{d. h.} & \lim \mu(A_n) & = & \mu\left(\lim\limits_{j} A_j\right) & \end{array}$$

Satz 3.1 *Eine endliche, nichtnegative additive Mengenfunktion μ über $\mathfrak{A}$ ist σ-additiv genau dann, wenn μ stetig ist.*

Beweis:

$(\Rightarrow)$ $B_n \in \mathfrak{A}$, $\quad B_n \supset B_{n+1} \supset \dots$, $\quad B_n \to B = \bigcap\limits_{j=1}^{\infty} B_j$

$$C_k = B_k \bar{B}_{k+1}, \qquad k = 1, 2, \dots$$

$$\begin{array}{rrcl} \Rightarrow & C_k C_{k+1} & = & B_k \bar{B}_{k+1} B_{k+1} \bar{B}_{k+2} = \emptyset, \\ & \emptyset = C_k C_{k+2} & = & B_k \bar{B}_{k+1} B_{k+2} \bar{B}_{k+3} \end{array}$$

Außerdem gilt $BC_k = \emptyset$.

$\Rightarrow$ Die Mengen sind paarweise unvereinbar.

$$B_n = B \cup \bigcup_{k=n}^{\infty} C_k \qquad \forall n$$

$$\textit{P3} \;\Rightarrow\; \mu(B_1) = \sum_{k=1}^{\infty} \mu(C_k) + \mu(B), \quad \text{d. h. die Reihe konvergiert.}$$

$$\Rightarrow\; \mu(B_n) = \sum_{k=n}^{\infty} \mu(C_k) + \mu(B) \qquad \Rightarrow \text{ Stetigkeit}$$

$$\downarrow$$

$$0 \text{ für } n \to \infty, \qquad \text{d. h. } \mu(B_n) \to \mu\left(\bigcap_{j=1}^{\infty} B_j\right).$$

$(\Leftarrow)$ $A_n \in \mathfrak{A}\,, \quad A_iA_j = \emptyset\,, \quad i \neq j\,.$ Es gilt:

$$\begin{aligned}
\mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mu\left(\bigcup_{k=1}^{n} A_k\right) + \mu\left(\bigcup_{k=n+1}^{\infty} A_k\right)\,, \\
\sum_{k=1}^{n} \mu(A_k) &= \mu\left(\bigcup_{k=1}^{n} A_k\right) \\
\Rightarrow \quad \lim_{n\to\infty} \sum_{k=1}^{n} \mu(A_k) &= \lim_{n} \mu\left(\bigcup_{k=1}^{n} A_k\right) \\
&= \lim_{n\to\infty}\left[\mu\left(\bigcup_{k=1}^{\infty} A_k\right) - \mu\left(\bigcup_{k=n+1}^{\infty} A_k\right)\right] \\
&= \mu\left(\bigcup_{k=1}^{\infty} A_k\right) - \lim_{n\to\infty} \underbrace{\mu\left(\bigcup_{k=n+1}^{\infty} A_k\right)}_{\substack{\text{monoton fallend} \\ \to 0 \text{ wegen Stetigkeit}}}
\end{aligned}$$

Auf beiden Seiten existiert Limes.

$$\Rightarrow \quad \mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k)$$

$\square$

Definition 3.4 $(\Omega, \mathfrak{A}, \mathsf{P})$ *für ein Wahrscheinlichkeitsmaß* P *heißt Wahrscheinlichkeitsraum.*

Diese Definition geht auf Kolmogorov zurück.

Beispiel 3.5 Experiment = zufälliges Werfen eines Punktes in das Einheitsquadrat $[0,1] \times [0,1]$.

$\Rightarrow \Omega = [0,1] \times [0,1]$
$\Rightarrow$ Durchführung des Experiments = Auswahl $\omega \in \Omega$

Mit der Borel-σ-Algebra $\mathfrak{B}^2$ ist $\mathfrak{A} = \mathfrak{B}^2 \cap \Omega$ eine σ-Algebra und für $B \in \mathfrak{A}$ definieren wir $\mathsf{P}(B) = \lambda_2(B)$. Dann ist $(\Omega, \mathfrak{A}, \mathsf{P})$ ein Wahrscheinlichkeitsraum.

4 Wahrscheinlichkeitsverteilungen

4.1 Eigenschaften von Wahrscheinlichkeitsverteilungen

- $\mathsf{P}(\emptyset) = 0$
- $\mathsf{P}(\bar{A}) = 1 - \mathsf{P}(A)$
- $A, B \in \mathfrak{A}, \quad A \subset B \Rightarrow \mathsf{P}(A) \leq \mathsf{P}(B)$
- $\mathsf{P}(A) \leq 1$
- $\mathsf{P}(A \cup B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$
- $$\mathsf{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathsf{P}(A_k) - \sum_{k<\ell} \mathsf{P}(A_k A_\ell) + \sum_{k<\ell<m} \mathsf{P}(A_k A_\ell A_m) - \cdots + (-1)^{n-1}\mathsf{P}(A_1 \cdots A_n)$$
  (Beweis durch vollständige Induktion)
- $\mathsf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathsf{P}(A_k)$, denn: $B_n = \Omega \setminus \bigcup_{k=1}^{n-1} A_k \left(= \overline{\bigcup_{k=1}^{n-1} A_k} \right)$
  $\Rightarrow (A_n B_n)(A_m B_m) = \emptyset$ und $\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} B_k A_k$
  $$\Rightarrow \mathsf{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathsf{P}(B_k A_k) \leq \sum_{k=1}^{\infty} \mathsf{P}(A_k)$$
- $\{A_k\}$ monoton wachsende Folge: $A_k \subset A_{k+1}$, $\quad A := \bigcup_{k=1}^{\infty} A_k$
  Dann gilt: $\quad \mathsf{P}(A) = \lim_{k\to\infty} \mathsf{P}(A_k)$, $\quad$ denn:
  definiere $\quad B_n = A \setminus A_n = \bar{A}_n \cap A \quad \Rightarrow \quad B_{n+1} \subset B_n$, $\quad \bigcap_{n=1}^{\infty} B_n = \emptyset$
  Stetigkeitsaxiom $\Rightarrow$ $\quad \mathsf{P}(A \setminus A_n) \to 0 \qquad n \to \infty$
  und $\quad \mathsf{P}(A \setminus A_n) = \mathsf{P}(A) - \mathsf{P}(A_n)$.

4.2 Produktexperimente

Es seien $(\Omega_i, \mathfrak{A}_i, \mathsf{P}_i)$ Wahrscheinlichkeitsräume. $\qquad i = 1, \ldots, n$
$(\Omega_i, \mathfrak{A}_i, \mathsf{P}_i)$ beschreibe das i-te Experiment.

Ziel: Beschreibung des Experiments, das durch Hintereinanderausführung dieser Teilexperimente entsteht.

Ereignisse $A = (A_1, A_2, \ldots, A_n)$, $\qquad A_i \in \mathfrak{A}_i$

$$\mathsf{P}(A) := \prod_{i=1}^{n} \mathsf{P}_i(A_i)$$

Setze $\quad \Omega = \mathop{X}_{i=1}^{n} \Omega_i, \quad \mathfrak{A} = \mathop{X}_{i=1}^{n} \mathfrak{A}_i.$

$(\Omega, \mathfrak{A}, \mathsf{P})$ heißt Produkt der Wahrscheinlichkeitsräume $(\Omega_i, \mathfrak{A}_i, \mathsf{P}_i)$.

Das so definierte P ist ein Wahrscheinlichkeitsmaß über $(\Omega, \mathfrak{A})$.

4.3 Beispiele für Wahrscheinlichkeitsverteilungen für Produktexperimente

4.3.1 Binomialverteilung

Ein Experiment (Versuch) besteht aus der r-fachen Wiederholung eines Einzelexperiments. Jedes Einzelexperiment habe genau zwei verschiedene Ergebnisse:

Erfolg	–	Mißerfolg
gut	–	schlecht
1	–	0

Bezeichnet man die Ausgänge mit $0, 1$, so ist $\omega_i = \begin{cases} 0 \\ 1 \end{cases}$, $\Omega_i = \{0, 1\}$ und

$\Omega = \{\omega \,|\, \omega = (\omega_1, \ldots, \omega_r)\} = \{0, 1\}^r$.

$\vartheta = \mathsf{P}(1)$ unabhängig von i (Erfolgswahrscheinlichkeit).

$\mathsf{P}(\omega) = \vartheta^k (1 - \vartheta)^{r-k}$, falls genau k-mal 1 in ω ist. Beim Bernoulli-Schema wurde gezeigt, daß für

$$\mathcal{E}_k = \{\omega : \sum \omega_i = k\} \quad \text{gilt:}$$

$\mathsf{P}(\mathcal{E}_k) = \binom{r}{k} \vartheta^k (1 - \vartheta)^{r-k} =: b_{r,\vartheta}(k)$

$\mathcal{E}_k$ bilden disjunkte Zerlegung von $\{\omega : \omega \in \{0, 1\}^r\}$

$\Rightarrow \sum_{k=0}^{r} b_{r,\vartheta}(k) = 1$

$\Rightarrow$ die $b_{r,\vartheta}(k)$ bilden Wahrscheinlichkeitsmaß auf $\{0, \ldots, r\}$

$=$ Binomialverteilung $B(r, \vartheta)$.

Beispiel 4.1 10 mal Würfeln
Ein Würfelexperiment hat nicht zwei, sondern sechs mögliche Ausgänge. Fragt man aber z. B. nach der Wahrscheinlichkeit bei 10 Würfen 3 Sechsen zu erhalten, so können wir die geworfenen Sechsen als Erfolge betrachten und die übrigen Würfe als Mißerfolge. Die Erfolgswahrscheinlichkeit ϑ ist $1/6$. Also ist die gesuchte Wahrscheinlichkeit $\binom{10}{3} (\frac{1}{6})^3 (\frac{5}{6})^7$.

Beispiel 4.2 Geburt eines Jungen
Ist $\vartheta = 0,51$ die Wahrscheinlichkeit für die Geburt eines Jungen, und sind aufeinanderfolgende Geburten unabhängig, so ist die Wahrscheinlichkeit dafür, daß in einem Krankenhaus unter 4 neugeborenen Kindern 2 Jungen und 2 Mädchen vorkommen, $\binom{4}{2} \vartheta^2 (1 - \vartheta)^2 \approx 0,374$.

4.3.2 Multinomialverteilung

Ein Experiment bestehe aus der n-fachen Wiederholung eines Einzelexperiments, das genau r mögliche Ausgänge $1, \dots, r$ habe. Es bezeichne $\vartheta_j = \mathsf{P}\{j\}$, $\Omega = \{1, \dots, r\}^n$. Falls $\vartheta_1, \dots, \vartheta_r$ gegeben sind, so ist die Wahrscheinlichkeit, in den n Einzelexperimenten insgesamt k_1-mal die 1, k_2-mal die 2, ... und k_r-mal das r zu erhalten

$$\underbrace{\frac{n!}{k_1!k_2!\cdots k_r!}}_{\text{Multinomialkoeffizient}} \vartheta_1^{k_1}\vartheta_2^{k_2}\cdots\vartheta_r^{k_r}$$

falls $k_i \geq 0$, $\sum k_i = n$ und sonst $= 0$.

Derartige Experimente kommen bei r möglichen unterschiedlichen Ergebnissen vor. Im Prinzip sind das alle praktisch relevanten Fälle, nur, daß r und damit die Anzahl der verschiedenen ϑ_i groß ist. Hier haben wir $\vartheta = (\vartheta_1, \dots, \vartheta_r)$ als unbekannten Parameter mit $\mathsf{P}_\vartheta(\omega) = \vartheta_1^{k_1}\cdots\vartheta_r^{k_r}$ für k_1-mal 1, ..., k_r-mal r in ω.

4.3.3 Geometrische Verteilung

In einem Bernoulli-Experiment mit der Erfolgswahrscheinlichkeit ϑ wird gezählt, in welchem Einzelversuch das erste Mal ein Erfolg auftritt. Die Wahrscheinlichkeit, den ersten Erfolg im k-ten Einzelversuch zu haben ist

$$\vartheta(1-\vartheta)^{k-1} \qquad k = 1, 2, \dots$$

Es werden eine unendliche Folge von Einzelversuchen zugelassen. Die Ergebnismenge ist hierbei $\{1, 2, \dots\}$. Bezeichnen wir mit Y die Variable, die den Wert k in einem Versuch annimmt, wenn im k-ten Einzelversuch das erste Mal Erfolg eintritt, so gilt

$$\mathsf{P}_\vartheta(k) = \vartheta(1-\vartheta)^{k-1}.$$

Wir haben $\displaystyle\sum_{k=1}^{\infty} \mathsf{P}_\vartheta(k) = 1.$

Diese Verteilung heißt *geometrische Verteilung.*

5 Die bedingte Wahrscheinlichkeit

Beispiel 5.1 Das Experiment bestehe im dreimaligen Werfen einer Münze. Zur Beschreibung wählen wir $\Omega = \{(\omega_1, \omega_2, \omega_3), \omega_i \in \{z, w\}\}$. Die Wahrscheinlichkeit dafür, daß Wappen genau einmal fällt, ist $3/8$. A sei das Ereignis $\{wzz,\ zwz,\ zzw\}$. Wenn zusätzlich bekannt ist, daß das Ereignis $B = \{$Anzahl der gefallenen Wappen ist ungerade$\}$ eintritt, so ändert sich die Wahrscheinlichkeit für A unter B. B enthält die

Elementarereignisse wzz, zwz, zzw, www. 3 davon sind in A. $\Rightarrow$ Es ist sinnvoll, die Wahrscheinlichkeit von A dann gleich 3/4 zu setzen.

Beispiel 5.2 Etwas allgemeiner: Es seien n gleichwahrscheinliche Elementarereignisse. A bestehe aus r, B aus m und AB aus k Elementarereignissen. Es ist natürlich (vgl. voriges Beispiel), die Wahrscheinlichkeit für A unter der Bedingung, daß B eingetreten ist, in folgender Weise zu bestimmen:

$$\mathsf{P}(A|B) = \frac{k}{m} = \frac{k/n}{m/n} = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}$$

Definition 5.1 *Sei* $(\Omega, \mathfrak{A}, \mathsf{P})$ *Wahrscheinlichkeitsraum,* $A, B \in \mathfrak{A}$ *mit* $\mathsf{P}(B) > 0$. *Die bedingte Wahrscheinlichkeit von* A *unter der Bedingung* B *wird definiert durch*

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}\,.$$

Zu zeigen: Es ist Wahrscheinlichkeitsmaß.

Definition 5.2 *Sei* $(\Omega, \mathfrak{A}, \mathsf{P})$ *Wahrscheinlichkeitsraum,* $A, B \in \mathfrak{A}$. *Die Ereignisse* A, B *heißen unabhängig, falls* $\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B)$. *(symmetrischer Begriff)*

5.1 Eigenschaften unabhängiger Ereignisse

- Unabhängigkeit von A, B mit $\mathsf{P}(B) > 0$ genau dann, wenn $\mathsf{P}(A|B) = \mathsf{P}(A)$.
- A, B unabhängig $\Rightarrow$ $\bar{A}$, B unabhängig
  $\mathsf{P}(\bar{A}B) = \mathsf{P}(B \setminus AB) = \mathsf{P}(B) - \mathsf{P}(AB) = \mathsf{P}(B)(1 - \mathsf{P}(A)) = \mathsf{P}(\bar{A})\mathsf{P}(B)$.
- $A, B_1, B_2 \in \mathfrak{A}$; A, B_1 sowie A, B_2 seien unabhängig und sei $B_1B_2 = \emptyset$
  $\Rightarrow$ A und $B_1 \cup B_2$ unabhängig
  denn: $$\begin{aligned}\mathsf{P}(A(B_1 \cup B_2)) &= \mathsf{P}(AB_1 \cup AB_2) = \mathsf{P}(AB_1) + \mathsf{P}(AB_2) \\ &= \mathsf{P}(A)(\mathsf{P}(B_1) + \mathsf{P}(B_2)) = \mathsf{P}(A)\mathsf{P}(B_1 \cup B_2)\,.\end{aligned}$$

Beispiel 5.3 Werfen zweier Münzen

A = „Erste Münze zeigt Wappen"
B = „Zweite Münze zeigt Zahl"

Wahrscheinlichkeit jedes Ereignisses ist 1/2 und

$$\mathsf{P}(AB) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = \mathsf{P}(A) \cdot \mathsf{P}(B)\,.$$

Definition 5.3 *Die Ereignisse $B_1, \ldots, B_n$ heißen vollständig unabhängig, falls für beliebige $1 \leq i_1 < i_2 < \ldots < i_r \leq n\,, \quad r = 2, 3, \ldots, n$*

$$\mathsf{P}\left(\bigcap_{k=1}^{r} B_{i_k}\right) = \prod_{k=1}^{r} \mathsf{P}(B_{i_k})$$

gilt.

Satz 5.1 *Aus der paarweisen Unabhängigkeit von n Ereignissen folgt nicht die vollständige Unabhängigkeit dieser Ereignisse.*

Beweis durch Angabe eines Beispiels:

Ein Tetraeder wird geworfen; er hat Begrenzungsflächen mit den Farben rot, blau, grün und rot/blau/grün.
Ereignis R (=rot): beim Wurf des Tetraeders liegt eine Fläche unten, die rot enthält.
Entsprechend für die Ereignisse B, G.

$$\Rightarrow \quad \mathsf{P}(R) = \mathsf{P}(B) = \mathsf{P}(G) = \frac{1}{2}\,.$$

$\mathsf{P}(RBG) = \frac{1}{4}$
R, B, G paarweise unabhängig $\quad \mathsf{P}(RB) = \mathsf{P}(R)\mathsf{P}(B) \quad$ usw.
$\mathsf{P}(R)\mathsf{P}(B)\mathsf{P}(G) = \frac{1}{8} \neq \frac{1}{4} \quad \Rightarrow$ nicht vollständig unabhängig

Definition 5.4 *Eine Familie von Ereignissen $\{B_i, i \in I\}$ heißt unabhängig, wenn für jede endliche Teilfamilie die Produktformel*

$$\mathsf{P}\left(\bigcap_{k=1}^{r} B_k\right) = \prod_{k=1}^{r} \mathsf{P}(B_k)$$

gilt. Hier kann I insbesondere auch überabzählbar sein.

5.2 Totale Wahrscheinlichkeit und Bayessche Formel

Es sei A Ereignis in $\mathfrak{A}$, $B_1, B_2, \ldots, B_n$, $n \leq \infty$ seien paarweise unvereinbare Ereignisse in $\mathfrak{A}$ mit $\mathsf{P}(B_i) > 0$ und $A \subset \bigcup_{j=1}^{n} B_j$. Dann gilt die

Formel der totalen Wahrscheinlichkeit

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

Beweis: Nach Definition der bedingten Wahrscheinlichkeit gilt für jedes j:

$$\mathsf{P}(AB_j) = \mathsf{P}(B_j) \cdot \mathsf{P}(A|B_j)$$

Außerdem gilt $A = \bigcup_{j=1}^{n} B_j A$ und $AB_1, AB_2, \ldots$ sind paarweise unvereinbar.

$$\Rightarrow \mathsf{P}(A) = \sum_{j=1}^{n} \mathsf{P}(AB_j) = \sum_{j=1}^{n} \mathsf{P}(B_j)\mathsf{P}(A|B_j)\,.$$

□

Folgerung: Unter der zusätzlichen Voraussetzungen $\mathsf{P}(A) > 0$ gilt die **Bayessche Formel**

$$\underline{\mathsf{P}(B_j|A)} = \frac{\mathsf{P}(AB_j)}{\mathsf{P}(A)} = \underline{\frac{\mathsf{P}(B_j)\mathsf{P}(A|B_j)}{\sum_{k=1}^{n} \mathsf{P}(B_k)\mathsf{P}(A|B_k)}}\,.$$

Denn: Zunächst erhält man aus der Definition der bedingten Wahrscheinlichkeit für beliebiges B :

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

Mit der Formel der totalen Wahrscheinlichkeit erhält man mit $B = B_j$ die Bayessche Formel.

Beispiel 5.4 Ausschußautos
Eine Firma produziert Autos in verschiedenen Standorten S_1, S_2. In S_2 werden k-mal mehr Autos hergestellt als in S_1. Die Ausschußquote in S_i sei p_i. Die Autos kommen alle in den Verkauf (gemischt). Wie groß ist die Wahrscheinlichkeit dafür, daß ein gekauftes Auto aus dem Standort S_2 stammt, wenn man weiß, daß es Ausschuß ist.

Lösung:

$\left.\begin{matrix} B_1 \\ B_2 \end{matrix}\right\}$ Ereignis, daß Auto aus $\left\{\begin{matrix} S_1 \\ S_2\,. \end{matrix}\right.$

A Ereignis, daß gekauftes Auto Ausschuß ist.

Gesucht $\mathsf{P}(B_2|A)$. Es gilt, wenn b_i die Anzahl der am Standort S_i hergestellten Autos ist,

$$\mathsf{P}(B_1) = \frac{b_1}{N}\,, \qquad \mathsf{P}(B_2) = \frac{b_2}{N}$$

$$\Rightarrow b_1 + b_2 = N\,, \qquad \frac{b_2}{b_1} = k$$

$$\Rightarrow (1+k)b_1 = N \quad \Rightarrow \quad \frac{b_1}{N} = \frac{1}{1+k} \quad \Rightarrow \quad \frac{b_2}{N} = \frac{k}{1+k}$$

$$\Rightarrow \mathsf{P}(B_1) = \frac{1}{1+k}\,, \qquad \mathsf{P}(B_2) = \frac{k}{1+k}\,.$$

Es gilt $\mathsf{P}(A|B_1) = p_1$, $\mathsf{P}(A|B_2) = p_2$.

Bayessche Formel: $A \subseteq B_1 \cup B_2$

$$\begin{aligned}
\mathsf{P}(B_2|A) &= \frac{\mathsf{P}(B_2)\mathsf{P}(A|B_2)}{\mathsf{P}(B_1)\mathsf{P}(A|B_1) + \mathsf{P}(B_2)\mathsf{P}(A|B_2)} \\
&= \frac{\frac{k}{1+k}p_2}{\frac{1}{1+k}p_1 + \frac{k}{1+k}p_2} = \underline{\underline{\frac{kp_2}{p_1 + kp_2}}}
\end{aligned}$$

Bei $\quad p_1 = p_2 \quad$ folgt: $\quad \mathsf{P}(B_2|A) = \frac{k}{1+k} = \mathsf{P}(B_2)$

$\quad p_1 = \frac{1}{2}p_2 \quad \Rightarrow \quad \mathsf{P}(B_2|A) = \frac{2k}{1+2k} > \frac{k}{1+k}$

Bezeichnung:

$\mathsf{P}(B_i)$	a-priori Wahrscheinlichkeiten
$\mathsf{P}(B_i|A)$	a-posteriori Wahrscheinlichkeiten für B_1, B_2.

6 Zufallsgrößen und ihre Momente

Grundvoraussetzung ist ein Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, \mathsf{P})$.
Wir konnten bisher Wahrscheinlichkeiten von Ereignissen berechnen.
Es ist zweckmäßig, diesen Begriff zu verallgemeinern.

Es sei Y eine Abbildung

$$\Omega \xrightarrow{Y} \mathcal{Y}.$$

Sei $\mathfrak{C}$ eine σ-Algebra über $\mathcal{Y}$. $\Rightarrow (\mathcal{Y}, \mathfrak{C})$ ist meßbarer Raum.

Wir wollen zeigen, daß durch Y über $(\mathcal{Y}, \mathfrak{C})$ ein Wahrscheinlichkeitsmaß in natürlicher Weise gegeben ist.

Definition 6.1 *Zu zwei meßbaren Räumen $(\Omega, \mathfrak{A})$, $(\mathcal{Y}, \mathfrak{C})$ heißt eine Abbildung $Y : \Omega \to \mathcal{Y}$ meßbar (genauer $\mathfrak{A}$ -$\mathfrak{C}$-meßbar), falls für alle $C \in \mathfrak{C}$ gilt: $\quad Y^{-1}(C) \in \mathfrak{A}$. (auch: Y meßbare Funktion von Ω in $\mathcal{Y}$. Wir sagen statt $\mathfrak{A}$-$\mathfrak{C}$-meßbar nur meßbar, wenn es nicht zu Mißverständnissen kommen kann.)*

Definition 6.2 *Eine meßbare Funktion Y von Ω in $\mathcal{Y}$ heißt Zufallsgröße (Zufallsvariable).*

Durch eine Zufallsgröße Y wird über $(\mathcal{Y}, \mathfrak{C})$ ein Wahrscheinlichkeitsmaß P^Y induziert:

$$\mathsf{P}^Y(C) := \mathsf{P}\left(Y^{-1}(C)\right) = \mathsf{P}(\omega : Y(\omega) \in C)$$

6.1 Nachweis der Meßbarkeit

1. Bei diskretem Ω, $\mathfrak{A} = \mathfrak{P}(\Omega)$ ist jede Abbildung Y meßbar.

2. Sei $(\mathcal{Y}, \mathfrak{C})$ allgemein. Dann hat man den folgenden Satz.

Satz 6.1 *Gilt für ein Erzeugendensystem $\mathfrak{C}'$ von $\mathfrak{C}$:* $\quad Y^{-1}(C') \in \mathfrak{A} \quad \forall C' \in \mathfrak{C}'$, *so ist $Y : (\Omega, \mathfrak{A}) \to (\mathcal{Y}, \mathfrak{C})$ meßbar.*

Beweis: Sei $\mathfrak{C}'$ ein Erzeugendensystem für $\mathfrak{C}$. Dann ist jedes $C \in \mathfrak{C}$ darstellbar durch abzählbare Anwendung der Vereinigung und Komplementbildung von Mengen aus $\mathfrak{C}'$. Deshalb genügt es z.z.:

$$Y^{-1}\left(\bigcup_{i=1}^{\infty} C_i'\right) = \bigcup_{i=1}^{\infty} Y^{-1}(C_i') \in \mathfrak{A},$$

$$Y^{-1}(\overline{C'}) = \overline{Y^{-1}(C')} \in \mathfrak{A}.$$

Zunächst gilt

$$Y^{-1}(\mathcal{Y}) = \Omega.$$

Weiter gilt beispielsweise $Y^{-1}(C_1' \cup C_2') = Y^{-1}(C_1') \cup Y^{-1}(C_2') \in \mathfrak{A}$.

$$\begin{aligned} Y^{-1}(\overline{C'}) = \{\omega : Y(\omega) \in \overline{C'}\} &= \Omega \setminus \{\omega : Y(\omega) \in C'\} \\ &= \overline{Y^{-1}(C')} \in \mathfrak{A} \end{aligned}$$

$\Rightarrow$ Behauptung

$\square$

Gilt speziell $(\mathcal{Y}, \mathfrak{C}) = (\mathbb{R}^n, \mathfrak{B}^n)$. Dann wissen wir

$$\left.\begin{array}{l} \text{Menge der offenen Intervalle} \\ \text{Menge der halboffenen Intervalle} \\ \text{Menge der geschlossenen Intervalle} \end{array}\right\} \text{sind Erzeugendensysteme von } \mathfrak{B}^n.$$

$\Rightarrow$ Y mit Werten in $\mathbb{R}^n$ ist meßbar genau dann, wenn $\{\omega : Y(\omega) \leq y\} \in \mathfrak{A}$.

Haben wir $(\mathbb{R}^m, \mathfrak{B}^m) \xrightarrow{Y} (\mathbb{R}^n, \mathfrak{B}^n)$

$\Rightarrow$ Jede stetige Funktion Y ist meßbar.
Jede monotone Funktion Y ist meßbar.

Bei $(\Omega, \mathfrak{A}) \xrightarrow{Y} (\mathbb{R}^1, \mathfrak{B}^1)$, $Y(\omega) = I_A(\omega)$ für $A \in \mathfrak{A}$ ist
Y meßbar.

Satz 6.2 *Gilt $(\Omega, \mathfrak{A}) \xrightarrow{Y} (\mathcal{Y}, \mathfrak{C}) \xrightarrow{T} (\mathcal{T}, \mathfrak{D})$ und sind Y, T meßbar, so ist auch $T \circ Y$ meßbar.*

Satz 6.3 *Sind $Y_1, \dots, Y_n$ reellwertige meßbare Funktionen auf Ω, so ist durch $Y(\omega) = (Y_1(\omega), \dots, Y_n(\omega))'$ eine $\mathbb{R}^n$-wertige meßbare Funktion Y definiert.*

Beweis: Für $a, b \in \mathbb{R}^n$ ist

$$Y^{-1}((a,b]) = \bigcap_{i=1}^{n} Y_i^{-1}((a_i, b_i])$$

Diese Intervalle $(a, b]$ erzeugen die Borelsche σ-Algebra $\mathfrak{B}^n$

□

Satz 6.4 *Sind* $Y_1, Y_2, \dots$ *reellwertige meßbare Funktionen,*
$\alpha_1, \alpha_2, \dots \in \mathbb{R}$,

so sind $\alpha_1 Y_1 + \dots + \alpha_n Y_n$

$Y_1 \cdots Y_n$

$\sup_i(Y_i), \inf_i(Y_i)$

$\limsup Y_i, \liminf Y_i$

meßbar (mit Wertebereich $\mathbb{R}^C = [-\infty, \infty]$).

Beweis: Setzt man $Y = (Y_1, \dots, Y_n)$ und $g(y_1, \dots, y_n) = \alpha_1 y_1, \dots, \alpha_n y_n$, so ist $\alpha_1 Y_1 + \dots + \alpha_n Y_n = g \circ Y$. g ist stetig $\Rightarrow$ meßbar. Nach vorhergehenden Sätzen folgt hierfür die Behauptung, ebenso für das Produkt.

Meßbarkeit abzählbarer Suprema folgt aus

$$\{\omega : \sup_i Y_i \leq x\} = \bigcap_i \{\omega : Y_i \leq x\},$$

abzählbarer Infima aus

$$\{\omega : \inf_i Y_i < x\} = \bigcup_i \{\omega : Y_i < x\}.$$

ebenso: $\limsup Y_i = \inf_k \left(\sup_{i \geq k} Y_i \right), \dots$

□

6.2 Verteilungsfunktionen für Zufallsgrößen

Sei $(\mathcal{Y}, \mathfrak{C}) = (\mathbb{R}^n, \mathfrak{B}^n)$. Die Verteilungsfunktion einer Zufallsgröße $Y : \Omega \to \mathbb{R}^n$ ist definiert durch:

$$F(y) = F^Y(y) = \mathsf{P}(Y \leq y).$$

Mit dem induzierten Maß gilt dann auch

$$F^Y(y) = \mathsf{P}^Y\left((-\infty, y]\right) = \mathsf{P}\left(Y^{-1}\left((-\infty, y]\right)\right).$$

Dann gilt

Satz 6.5 *Jede Verteilungsfunktion F ist*

1. *monoton nicht fallend,*
2. *rechtsseitig stetig,* und es gilt
3. $\lim\limits_{y\to-\infty} F(y) = 0\,, \quad \lim\limits_{y\to\infty} F(y) = 1$

Beweis:

1. Sei $y_1 < y_2 \Rightarrow \{\omega \,|\, Y(\omega) \le y_1\} \subseteq \{\omega \,|\, Y(\omega) \le y_2\}$. Wegen der Monotonie von P gilt $F(y_1) \le F(y_2)$.
2. Sei $\{y_n\}$ Folge, $\ge y_0$ und $y_n \to y_0$, o. B. d. A. sei $\{y_n\}$ monoton nicht steigend. Betrachte $B_n = \{\omega : Y(\omega) \le y_n\}$. $\Rightarrow B_n$ fallend.
   Wegen σ-Additivität, d. h. Stetigkeit von P, folgt
   $$F(y_n) = \mathsf{P}(B_n) \to \mathsf{P}\left(\underbrace{\bigcap_{j=1}^{\infty} B_j}_{B_0}\right) = \mathsf{P}(B_0) = F(y_0)$$
3. Sei $y_n \to -\infty$, monoton. $\Rightarrow B_n = \{\omega : Y(\omega) \le y_n\}$ ist monoton fallend. Es gilt $\bigcap\limits_{j=1}^{n} B_j \to \emptyset$. Aus der Stetigkeit von P folgt: $F(y_n) \to 0$. Außerdem folgt genauso $F(y_n) \to 1$ für $y_n \to \infty$.

□

Satz 6.6 *(Umkehrung) Es sei* $(\mathfrak{Y}, \mathfrak{C}) = (\mathbb{R}^1, \mathfrak{B}^1)$. *Zu jeder Funktion F mit 1., 2., 3. existiert ein Maß Q über* $(\mathbb{R}^1, \mathfrak{B}^1)$ *mit*

$$Q\left((-\infty, y]\right) = F(y)\,.$$

Beweis: Die halboffenen Intervalle erzeugen $\mathfrak{B}^1$. Es folgt zunächst aus der Definition $Q\left((-\infty, y]\right) \ge 0$. $\Rightarrow Q(B) \ge 0$ für $B \in \mathfrak{B}^n$; $Q(\mathbb{R}^1) = 1$ folgt aus $F(y) \to 1$ für $y \to \infty$. Da ein nichtnegatives, additives Maß genau dann σ-additiv ist, wenn es stetig ist, folgt aus der rechtsseitigen Stetigkeit die σ-Additivität von Q.

□

Die Umkehrung für $n > 1$ gilt unter einer zusätzlichen Bedingung, die die Nichtnegativität des Maßes sichert. Für $n = 2$ ist es hinreichend zu beschreiben.

Satz 6.7 *Für $n = 2$ gelte zusätzlich zu 1., 2., 3. aus Satz 6.5 die Bedingung*

$$F(y_1, y_2) - F(z_1, y_2) - F(y_1, z_2) + F(z_1, z_2) \geq 0 \quad \textit{(Nichtnegativitätsbedingung)}$$

für alle $z \leq y$, d. h. $z_1 \leq y_1$, $z_2 \leq y_2$. Dann existiert ein Maß Q über $(\mathbb{R}^2, \mathfrak{B}^2)$ mit

$$Q((-\infty, y]) = F(y)$$

Allgemein kann man zeigen, daß auch für $n > 2$ zu jeder Funktion mit 1.,2. und 3. und einer zusätzlichen Nichtnegativitätsbedingung ein Maß über $(\mathbb{R}^n, \mathfrak{B}^n)$ gehört. In diesem Sinne kann man sagen, daß zu jeder Verteilungsfunktion ein Maß gehört und umgekehrt.

Definition 6.3 *Falls Q ein Maß über $(\mathbb{R}^n, \mathfrak{B}^n)$ ist, so heißt Q in $y \in \mathbb{R}^n$ stetig, falls $Q(y) = 0$. Falls $Q = \mathsf{P}^Y$, so heißt Y stetig, falls $\mathsf{P}^Y(y) = 0 \quad \forall y \in \mathbb{R}^n$ gilt.*

Die Folgerung für die Verteilungsfunktion ist:
Wenn $Q(y_0) = 0$, so ist F mit $F(y) = Q((-\infty, y])$ in y_0 stetig. Denn wir haben $Q(y_0) = F(y_0) - F(y_0 - 0)$.

6.3 Beispiele für Verteilungsfunktionen

Diskrete Verteilungen
Eine Verteilung heißt diskret, falls höchstens abzählbar unendlich viele Punkte $\{y_i\}$ aus $\mathcal{Y}$ existieren mit

$$Q(y_i) > 0\,, \qquad \sum_{i=1}^{\infty} Q(y_i) = 1\,.$$

Dann heißt auch eine Zufallsgröße, die eine diskrete Verteilung hat, diskret. Damit ist eine diskrete Zufallsgröße gekennzeichnet durch die Werte y_i, die sie annehmen kann, und die dazugehörigen Wahrscheinlichkeiten p_i. Für $\mathcal{Y} = \mathbb{R}^n$ haben wir: Mit $F(y) = Q((-\infty, y])$ ist F eine Sprungfunktion, stückweise konstant. Die Sprunghöhe ist gleich der Wahrscheinlichkeit des jeweiligen Punktes.

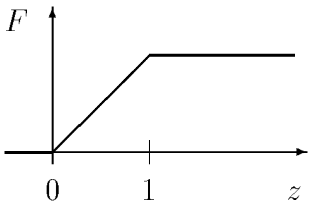
Gleichverteilung

Definition 6.4 *Z heißt gleichverteilt über $(\mathcal{Y}, \mathfrak{C})$, $\quad \mathcal{Y} \subseteq \mathbb{R}^p$, $\mathcal{Y} \in \mathfrak{B}^p$, $\mathfrak{C} = \mathfrak{B}^p \cap \mathcal{Y}$, falls Z stetig ist, $0 < \lambda_p(\mathcal{Y}) < \infty$ und*

$$\mathsf{P}^Z(A) = \frac{\lambda_p(A \cap \mathcal{Y})}{\lambda_p(\mathcal{Y})}\,.$$

Falls Z diskret ist mit $\sum_{i=1}^{m} \mathsf{P}^Z(z_i) = 1$, so heißt Z gleichverteilt, wenn $\mathsf{P}^Z(z_i) = \dfrac{1}{m} \quad \forall i$.

Beispiel 6.1 Gleichverteilung über dem Intervall $[0,1]$.
Sei $\mathcal{Y} = [0,1]$, $A = (0,z] \Rightarrow \mathsf{P}^Z((0,z]) = z$, d. h. $\mathsf{P}(Z \leq z) = z$.



Satz 6.8 *Y sei eine reellwertige Zufallsgröße, $Y : \Omega \to \mathbb{R}^1$ mit stetiger Verteilungsfunktion im $\mathbb{R}^1$. Dann ist die Zufallsgröße $F(Y) := F \circ Y$ gleichverteilt auf* $[0,1]$.

Beweis: Zu zeigen: $\mathsf{P}\left(F(Y) \leq z\right) = z$ für $0 \leq z \leq 1$.
Es ist stets F^{-1} streng monoton.

$$\begin{aligned}
\Rightarrow \quad \mathsf{P}\left(F(Y) \leq z\right) &= \mathsf{P}\left(F^{-1}\left(F(Y)\right) \leq F^{-1}(z)\right) \\
&\quad (\text{bei } F^{-1}(z) = \min\{t : F(t) \geq z\}) \\
&= \mathsf{P}\left(Y \leq F^{-1}(z)\right) = F\left(F^{-1}(z)\right) = z
\end{aligned}$$

□

Satz 6.9 *Sei U eine gleichverteilte Variable über* $[0,1]$. *Sei F eine streng monoton wachsende Verteilungsfunktion. Dann besitzt die Zufallsgröße $F^{-1}(U)$ die Verteilungsfunktion F.*

Beweis: $z = \mathsf{P}(U \leq z) = \mathsf{P}\left(F^{-1}(U) \leq F^{-1}(z)\right)$
Ist G die Verteilungsfunktion von $F^{-1}(U)$, so gilt

$$G\left(F^{-1}(z)\right) = z \qquad \forall z \in [0,1]\,.$$

$\Rightarrow F^{-1}\left(G\left(F^{-1}(z)\right)\right) = F^{-1}(z)$. F^{-1} nimmt alle Werte aus $(0,1)$ an.
$\Rightarrow G \circ F^{-1} = I\,, \quad F^{-1} \circ G = I \quad \Rightarrow \quad G = F.$

□

Satz 6.10 *Eine Verteilungsfunktion F über $\mathbb{R}^1$ besitzt höchstens abzählbar unendlich viele Unstetigkeitsstellen.*

Beweis: Sei A_k die Menge derjenigen Punkte, an denen die Verteilungsfunktion einen Sprung hat, der $\geq \frac{1}{k}$ ist. Wegen $0 \leq F(y) \leq 1$ folgt $\#A_k \leq k$. Die Menge aller Unstetigkeitsstellen ist $A = \bigcup\limits_{k=1}^{\infty} A_k$ und damit ist A höchstens abzählbar.

□

6.4 Verteilungsfunktion transformierter Variabler

Sei φ eine streng monoton wachsende Funktion: $\mathcal{Y} \to \mathbb{R}^1$. Dann ist $Z := \varphi(Y)$ wieder eine Zufallsgröße. Gesucht ist der Zusammenhang zwischen F^Z und F^Y.

$$\begin{aligned} F^Z(z) &= \mathsf{P}(Z \leq z) = \mathsf{P}\left(\varphi(Y) \leq z\right) = \mathsf{P}\left(\varphi(Y) \leq \varphi\left(\varphi^{-1}(z)\right)\right) \\ &= \mathsf{P}\left(Y \leq \varphi^{-1}(z)\right) = F^Y\left(\varphi^{-1}(z)\right). \end{aligned}$$

z. B. sei $Y \geq 0$, $Y \in \mathbb{R}^1$ und $Z = Y^2$.

$$\Rightarrow F^Z(z) = F^Y(\sqrt{z})$$

Wenn jetzt Y beliebig, $Y \in \mathbb{R}^1$, $Z = Y^2$. (Hier ist $\varphi(y) = y^2$ nicht monoton). Dann erhält man:

$$\begin{aligned} F^Z(z) &= \mathsf{P}(Y^2 \leq z) = \mathsf{P}(-\sqrt{z} \leq Y \leq \sqrt{z}) = \mathsf{P}(Y \leq \sqrt{z}) - \mathsf{P}(Y < -\sqrt{z}) \\ \Rightarrow \quad F^Z(z) &= F^Y(\sqrt{z}) - \mathsf{P}(Y \leq -\sqrt{z}) + \mathsf{P}(Y = -\sqrt{z}) \\ &= F^Y(\sqrt{z}) - F^Y(-\sqrt{z}) + \mathsf{P}(Y = -\sqrt{z}). \end{aligned}$$

6.5 Unabhängigkeit von Zufallsgrößen

Es seien $Y_1, Y_2, \ldots$ Zufallsgrößen über $(\Omega, \mathfrak{A}, \mathsf{P})$ und

$$Y_i : (\Omega, \mathfrak{A}) \to (\mathcal{Y}_i, \mathfrak{C}_i).$$

Wir haben bisher den Begriff „unabhängig“ mehr heuristisch benutzt. Wir kannten eine Familie von unabhängigen Ereignissen.

Definition 6.5 *Eine Familie $\{Y_i\}_{i \in I}$ von Zufallsgrößen heißt unabhängig, falls für alle $C_i \in \mathfrak{C}_i$ die Ereignisse $A_i := \{\omega : Y_i(\omega) \in C_i\}$ für $i \in I$ unabhängig sind, d. h.*

$$\mathsf{P}\left(\bigcap_{j_i=1}^{k} A_{j_i}\right) = \prod_{j_i=1}^{k} \mathsf{P}(A_{j_i}) \qquad \forall k, \forall \{j_i\}.$$

Satz 6.11 *Seien $Y_1, \ldots, Y_n$ unabhängig. Dann gilt für beliebige $C_1, \ldots, C_n$, $\quad C_i \in \mathfrak{C}_i$:*

$$\mathsf{P}(Y_1 \in C_1, \ldots, Y_n \in C_n) = \prod_{i=1}^{n} \mathsf{P}(Y_i \in C_i)$$

BEWEIS:

$$\begin{aligned}
\mathsf{P}\Big(\omega\,:\,Y_1(\omega)\in C_1,\dots,Y_n(\omega)\in C_n\Big) &= \mathsf{P}\left(\bigcap_{i=1}^{n}\{\omega\,:\,Y_i(\omega)\in C_i\}\right)\\
&= \mathsf{P}\left(\bigcap_{i=1}^{n} A_i\right)\\
&= \prod_{i=1}^{n}\mathsf{P}(A_i)=\prod_{i=1}^{n}\mathsf{P}(Y_i\in C_i)\,.
\end{aligned}$$

□

Die Umkehrung gilt nach Definition, denn durch die Wahl $C_i=\mathcal{Y}_i$ erreicht man die für die Unabhängigkeit notwendigen Bedingungen.

Satz 6.12 *$\{Y_i\}_{i\in I}$ sind unabhängig genau dann, wenn $\mathfrak{C}'_i$ ein Erzeugendensystem für $\mathfrak{C}_i$, $i\in I$, und für alle $C'_i\in\mathfrak{C}'_i$ die Ereignisse $A'_i:=\{\omega\,:\,Y_i(\omega)\in C'_i\}$, $i\in I$, unabhängig sind.*

o. B.

Satz 6.13 *Sind $Y_1,\dots,Y_n$ unabhängig und sind $g_1,\dots,g_n$ beliebige meßbare Funktionen über $\mathcal{Y}$, so sind $g_1(Y_1),\dots,g_n(Y_n)$ unabhängig.*

BEWEIS: Sei $g_i\,:\,\mathcal{Y}_i\to\mathcal{G}_i$. Dann folgt

$$\begin{aligned}
\mathsf{P}(g_1(Y_1)\in G_1,\dots,g_n(Y_n)\in G_n) &= \mathsf{P}\left(\bigcap_{i=1}^{n}\{\omega\,:\,g_i(Y_i(\omega))\in G_i\}\right)\\
&= \mathsf{P}\left(\bigcap_{i}\{\omega\,:\,Y_i(\omega)\in g_i^{-1}(G_i)\right)\\
&= \prod_{i}\mathsf{P}(g_i(Y_i)\in G_i)\,.
\end{aligned}$$

□

6.6 Unabhängige Versuche

Gegeben ist ein Wahrscheinlichkeitsraum $(\Omega,\mathfrak{A},\mathsf{P})$. Die Modellierung unabhängiger Versuche kann in verschiedener Weise erfolgen. Als Beispiel diene die folgende Aufgabe:

Es sollen die beiden Versuchsanstellungen modelliert werden

a) n Würfel werden mit einem Mal geworfen,

b) 1 Würfel wird n mal hintereinander geworfen.

Nach Vollendung dieser Würfe ist die Wahrscheinlichkeit des Ereignisses $C = \{$Auftreten der 6 bei einem Würfel$\}$ zu bestimmen.

Derartige Versuchsanstellungen a) und b) müßten in der Modellierung und Bestimmung von $\mathsf{P}_\vartheta(C)$ zum selben Ergebnis führen.

A) Es seien n Zufallsgrößen $Y_1, \ldots, Y_n$ gegeben:

$$(\Omega, \mathfrak{A}, \mathsf{P}) \xrightarrow{Y_i} (\mathcal{Y}, \mathcal{C})$$

Y_i beschreibt die Ergebnisse des i-ten Würfels. Es wird vorausgesetzt, daß $Y_1, \ldots, Y_n$ unabhängige, identisch verteilte Zufallsgrößen sind. Unter a) wird eine Realisierung $y_1, \ldots, y_n$ von $Y_1, \ldots, Y_n$ mit $y_i = Y_i(\omega)$ bestimmt. Als Schätzung für $\mathsf{P}_\vartheta(C)$ dient

$$\hat{\vartheta} = \frac{\#\{y_i : y_i = 6\}}{n} .$$

B) Es ist eine Zufallsgröße Y_1 gegeben:

$$(\Omega, \mathfrak{A}, \mathsf{P}) \xrightarrow{Y_1} (\mathcal{Y}, \mathcal{C})$$

Y_1 beschreibt das Ergebnis des einen Würfels. Die Unabhängigkeit der n nacheinander durchgeführten Würfe beschreibt man durch n unabhängige, identisch verteilte Zufallsgrößen $V_1, \ldots, V_n$

$$(\Omega, \mathfrak{A}, \mathsf{P}) \xrightarrow{V_i} (\Omega, \mathfrak{A})$$

mit $\mathsf{P} = \mathsf{P}^{V_i}$ für alle i. Dann sind $Y_1(V_1), \ldots, Y_1(V_n)$ unabhängige und wegen

$$\mathsf{P}^{Y_1 \circ V_i}(C) = \mathsf{P}^{V_i}(Y_1^{-1}(C)) = \mathsf{P}(Y_1^{-1}(C)) = \mathsf{P}^{Y_1}(C)$$

auch identisch verteilte Zufallsgrößen und es gilt

$$\mathsf{P}^{Y_1 \circ V_i} = \mathsf{P}^{Y_1} .$$

Die Ergebnisse der n hintereinander durchgeführten Würfe seien $y_1, \ldots, y_n$. Diese werden mit dem einem Würfel bei unabhängigen Versuchen erzielt, so wird

$$y_i = Y_1(V_i(\omega_i)) , \quad i = 1, \ldots, n$$

gesetzt. Mit $V_i(\omega_i) = \tilde{\omega}_i$ kann man dann

$$y_i = Y_1(\tilde{\omega}_i)$$

schreiben. Wegen $\mathsf{P}^{Y_1 \circ V_i} = \mathsf{P}^{Y_1}$ dient wiederum als Schätzung für $\mathsf{P}_\vartheta(C)$ der Wert $\hat{\vartheta}$.

Ergebnis: Unabhängige Versuche mit identisch verteilten Ergebnissen können beschrieben werden durch

$$Y_1, \ldots, Y_n \quad \text{und} \quad y_i = Y_i(\omega)$$

oder durch Y_1, $V_1, \ldots, V_n$ mit $\mathsf{P}^{V_i} = \mathsf{P}$ und

$$y_i = Y_1(\tilde{\omega}_i)\,, \quad \tilde{\omega}_i = V_i(\omega_i)\,.$$

6.7 Unabhängigkeit reellwertiger Zufallsgrößen

Seien $Y_1, Y_2, \ldots$ reellwertige Zufallsgrößen, $Y_i : \Omega \to \mathbb{R}^1$.
$\Rightarrow Y = (Y_1, \ldots, Y_n)'$ ist Zufallsgröße mit Werten in $\mathbb{R}^n$.
$\Rightarrow \mathsf{P}(Y \in B)$ ist für beliebige $B \in \mathfrak{B}^n$ erklärt.

Die halboffenen Intervalle $(-\infty, y_1]$ erzeugen $\mathfrak{B}^1$. Deshalb folgt aus Satz 6.12:
$Y_1, \ldots, Y_n$ sind unabhängig genau dann, wenn

$$\mathsf{P}(Y_1 \le y_1, \ldots, Y_n \le y_n) = \mathsf{P}(Y_1 \le y_1) \cdots \mathsf{P}(Y_n \le y_n) \qquad \forall y_1, \ldots, y_n\,,$$

d. h. $$F^Y(y) = \prod_{j=1}^{n} F^{Y_i}(y_i)\,. \tag{$*$}$$

$\Rightarrow$ Folgerung: $Y_1, \ldots, Y_n$ sind unabhängig genau dann, wenn $(*)\,\forall y_1, \ldots, y_n$ gilt.

Es ist $F^{Y_j}(y_j) = F^Y(\infty, \ldots, \infty, y_j, \infty, \ldots, \infty)$, d. h. $y_i = \infty$ für $i \neq j$, und diese Funktionen F^{Y_j} heißen Randverteilungsfunktionen. So bedeutet $(*)$, daß im Falle der Unabhängigkeit F^Y das Produkt der Randverteilungsfunktionen ist.

Ein weiteres Kriterium kann über die Dichten formuliert werden.

6.8 Dichten von Wahrscheinlichkeitsverteilungen

Sei λ_p das Lebesgue-Maß über $(\mathbb{R}^p, \mathfrak{B}^p)$. Sei F absolut stetig, d. h. es existiert eine $\mathfrak{B}^p$-meßbare Funktion g mit

$$F(y) = \int_{-\infty}^{y} g(\xi)\lambda_p(\,\mathrm{d}\xi)\,.$$

Wenn ein solches $g \ge 0$ existiert, so heißt g Dichte von F; dann ist $\dfrac{\mathrm{d}}{\mathrm{d}y}F(y) = g(y)$.

Allgemeiner spricht man von einer Dichte bezüglich einem σ-endlichen Maß μ unter folgenden Bedingungen. Es sei Q ein Wahrscheinlichkeitsmaß über $(\mathcal{Y}, \mathfrak{C})$ und μ sei ein σ-endliches Maß. Q heißt μ-stetig, falls für jedes $\mathcal{N} \in \mathfrak{C}$ mit $\mu(\mathcal{N}) = 0$ auch $Q(\mathcal{N}) = 0$ folgt.

Satz 6.14 *Satz von Radon-Nikodym*
Sei μ σ-endliches Maß, Q sei μ-stetig. Dann gibt es eine endliche, $\mathfrak{C}$-meßbare Funktion q auf $\mathcal{Y}$ mit

$$Q(C) = \int_C q(y)\mu(\,\mathrm{d}y) \qquad \forall C \in \mathfrak{C}\,.$$

Bezeichnung:
q heißt Radon-Nikodym Ableitung von Q bezüglich μ. Man schreibt auch $q = \dfrac{\mathrm{d}Q}{\mathrm{d}\mu}$.
q ist bis auf μ-Nullmengen aus $\mathfrak{C}$ eindeutig bestimmt. (Krickeberg S. 126–127)
q heißt auch Dichte von Q. Gilt $Q = \mathsf{P}^Y$, so ist q die Dichte von Y.

6.8.1 Beispiele (diskrete Verteilungen, Gleichverteilung, Normalverteilung)

Beispiel 6.2 $(\mathcal{Y}, \mathfrak{C}) = (\mathbb{R}^n, \mathfrak{B}^n)$. Sei Q ein diskretes Wahrscheinlichkeitsmaß über $\mathcal{Y}$, d.h. $Q(C)$ ist definiert und es existieren höchstens abzählbar unendlich viele y_i mit $Q(y_i) > 0$, $\sum_{i=1}^{\infty} Q(y_i) = 1$, μ sei Abzählmaß über $\mathcal{Y}$ mit

$$\mu(C) = \#\{y_i \in C\}\,.$$

$\Rightarrow$ μ σ-endliches Maß, nicht normiert.
Q ist μ-stetig, denn sei $\mu(\mathcal{N}) = 0 \quad \Rightarrow \quad Q(\mathcal{N}) = 0$.
$\Rightarrow Q(C) = \int_C q(y)\mu(\,\mathrm{d}y)$ hat eine $\mathfrak{C}$-meßbare Lösung.

$$q(y) = \begin{cases} 0 & \text{für} \quad y \neq y_i \quad \forall i \\ Q(y_i) & \text{für} \quad y = y_i \end{cases}$$

$\Rightarrow$ bei diskreten Verteilungen sind die Wahrscheinlichkeiten der y_i die von 0 verschiedenen Werte der Dichte q.

Beispiel 6.3 Gleichverteilung über B
Sei Q Gleichverteilung über $B \in \mathfrak{B}^n$, d.h. $\lambda_n(B) > 0$ und

$$Q(A) = \frac{\lambda_n(AB)}{\lambda_n(B)}, \qquad A \in \mathfrak{B}^n\,.$$

Das bedeutet $Q(A) = \int_A I_B(\xi)\lambda_n(\,\mathrm{d}\xi) \cdot \dfrac{1}{\lambda_n(B)}$, so daß

$$q(\xi) = \begin{cases} \dfrac{1}{\lambda_n(B)} & \text{für } \xi \in B \\ 0 & \text{sonst} \end{cases}$$

die Dichte von Q ist. Sie ist über B konstant (> 0) und sonst $= 0$.

Beispiel 6.4 Normalverteilung
Es gilt

$$\int\limits_{\mathbb{R}^n} e^{-\frac{1}{2}(\xi_1^2+\cdots+\xi_n^2)}\,\mathrm{d}\xi_1\cdots\mathrm{d}\xi_n = \left(\sqrt{2\pi}\right)^n .$$

Damit ist

$$q(\xi) = \frac{1}{\left(\sqrt{2\pi}\right)^n}\, e^{-\frac{1}{2}|\xi|^2}, \qquad \xi \in \mathbb{R}^n$$

für $|\xi|^2 = \sum\limits_{i=1}^{n} \xi_i^2$ eine Dichte über $\mathbb{R}^n$. Eine Zufallsgröße mit dieser Dichte heißt n-dimensionale standardisierte normalverteilte Zufallsgröße (Dichte hat die Gestalt einer Gaußkurve).

6.8.2 Dichten unabhängiger Zufallsgrößen

Die Unabhängigkeit von Zufallsgrößen kann auch durch die Dichten geprüft werden.

Satz 6.15 *Sind $Y_1, \ldots, Y_n$ unabhängige, reellwertige Zufallsgrößen und hat Y_i die Dichte f_i bezüglich μ_i, so hat $Y = (Y_1, \ldots, Y_n)'$ die Dichte*

$$f(y_1, \ldots, y_n) = \prod_{i=1}^{n} f_i(y_i)$$

bezüglich $\mu = \mu_1 \times \cdots \times \mu_n$.

Beweis: Nach Voraussetzung sind f_i Dichten bezüglich μ_i. Dann ist f eine Dichte bezüglich μ. Zu zeigen ist nur, daß Y dieses f als Dichte hat. Sei Q das Wahrscheinlichkeitsmaß, das f als Dichte hat. $\Rightarrow$ Für beliebige $a, b \in \mathbb{R}^n$ gilt bei $a \leq b$

$$\begin{aligned}
Q\left((a,b]\right) &= \int\limits_a^b f(y_1,\ldots,y_n)\mu_1(\mathrm{d}y_1)\cdots\mu_n(\mathrm{d}y_n) \\
&= \prod_{i=1}^{n} \int\limits_{a_i}^{b_i} f_i(y_i)\mu_i(\mathrm{d}y_i) = \prod_{i=1}^{n} \mathsf{P}\left(Y_i \in (a_i, b_i]\right) \\
&= \mathsf{P}\left(Y_i \in (a_i, b_i] \quad \forall i\right) = \mathsf{P}^Y\left((a,b]\right)
\end{aligned}$$

Da die halboffenen Intervalle $\mathfrak{B}^n$ erzeugen, gilt $Q = \mathsf{P}^Y$.

$\square$

Bemerkung: In den Integralen muß die untere Grenze ausgeschlossen werden, da $\mathsf{P}^{Y_i}((a_i, b_i]) \neq \mathsf{P}^{Y_i}([a_i, b_i])$ im allgemeinen gilt, man aber in beiden Fällen formal

$$\int_{a_i}^{b_i} f_i(y_i)\mu_i(\,\mathrm{d}y_i)$$

als Darstellung wählen würde.

Satz 6.16 *(Umkehrung) Hat* $Y = (Y_1, \dots, Y_n)$ *die Dichte* $f(y) = \prod_{i=1}^{n} f_i(y_i)$ *bezüglich* $\mu = \mu_1 \times \cdots \times \mu_n$ *und gilt* $\int f_i(y_i)\mu_i(\,\mathrm{d}y_i) = 1$, $f_i \geq 0$, *so sind die* Y_i *unabhängig und diese haben die Dichte* f_i.

Beweis: Wir haben

$$\begin{aligned}
\mathsf{P}^{Y}\left((a,b]\right) &= \mathsf{P}\left(Y_i \in (a_i, b_i] \quad \forall i\right) = \int_a^b f(y)\mu(\,\mathrm{d}y) \\
&= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(y_1, \dots, y_n)\mu_1(\,\mathrm{d}y_1)\cdots\mu_n(\,\mathrm{d}y_n) = (1)
\end{aligned}$$

Setzt man für festes j für alle $i \neq j$: $a_i = -\infty$, $b_i = +\infty$, so folgt hieraus

$$\mathsf{P}\left(Y_j \in (a_j, b_j]\right) = \int_{a_j}^{b_j} f_j(y_j)\mu_j(\,\mathrm{d}y_j) \cdot \prod_{i \neq j} \int_{-\infty}^{\infty} f_i(y_i)\mu_i(\,\mathrm{d}y_i) \qquad (*)$$

Wegen

$$\int_{-\infty}^{\infty} f_i(y_i)\mu_i(\,\mathrm{d}y_i) = 1 \qquad \forall i$$

folgt daraus, daß Y_j die Dichte f_j hat. Außerdem folgt nun

$$\begin{aligned}
(1) &= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_1(y_1)\cdots f_n(y_n)\mu_1(\,\mathrm{d}y_1)\cdots\mu_n(\,\mathrm{d}y_n) \\
&= \prod_i \int_{a_i}^{b_i} f_i(y_i)\mu(\,\mathrm{d}y_i) = \prod_i \mathsf{P}^{Y_i}\left((a_i, b_i]\right)
\end{aligned}$$

$\Rightarrow$ Unabhängigkeit

$\square$

Mit diesem Satz ist die Unabhängigkeit von Zufallsgrößen, die eine Dichte besitzen, beschrieben, d.h. insbesondere, daß die Unabhängigkeit diskreter Zufallsgrößen charakterisiert wurde. Denn diese haben in jedem Fall eine Dichte, nämlich bzgl. des Abzählmaßes. Man kann für diskrete Zufallsgrößen die Bedingung der Unabhängigkeit explizit auch direkt beschreiben. Wir demonstrieren dies für $n = 2$.
Es sei $Y = (Y_1, Y_2)$ eine diskrete Zufallsgröße mit Werten in $\mathbb{R}^2$. Dann nimmt Y höchstens abzählbar unendlich viele Werte $y_1, y_2, ...$ an und es gilt $\mathsf{P}(y_i) = p_i$. Da Y eine diskrete Zufallsgröße ist, sind auch Y_1 und Y_2 diskret verteilt. Nimmt $Y_j, j = 1, 2$ die Werte $y_{jt}, t = 1, 2, ...$ mit den Wahrscheinlichkeiten p_{jt} an, so folgt

$$\{y_i, i = 1, 2, ...\} \subseteq \{(y_{1,t}, y_{2,s}, t, s = 1, 2, ...\},$$

beide Mengen sind abzählbar. Man kann von vornherein davon ausgehen, daß beide Mengen übereinstimmen, denn andernfalls würde man die in der Menge der linken Seite fehlenden Punkte mit der zugehörigen Wahrscheinlichkeit 0 hinzufügen. Damit gilt bei geeigneter Numerierung

$$y_i = (y_{1t}, y_{2s}).$$

Mit dieser Numerierung bezeichnen wir die Wahrscheinlichkeiten $p_i = q_{ts}$. Diese Wahrscheinlichkeiten q_{ts} sind die Werte der Dichte der zweidimensionalen diskreten Zufallsgröße Y. Der vorhergehende Satz besagt dann, daß die diskreten Zufallsgrößen Y_1, Y_2 genau dann unabhängig sind, wenn

$$q_{ts} = p_{1t} p_{2s} \quad \text{für alle } t, s = 1, 2, ...$$

gilt.

6.8.3 Dichten transformierter Zufallsgrößen

$Y : \Omega \to \mathbb{R}^n$
Y habe die Dichte f bezüglich des Lebesgue-Maßes λ_n.

$$\Rightarrow \qquad \mathsf{P}(Y \leq y) = \int_{-\infty}^{y} f(\xi)\lambda_n(\,\mathrm{d}\xi)$$

Satz 6.17 *Sei $Z = h(Y)$ und sei h eine stetig differenzierbare, eineindeutige Funktion. Dann ist die Dichte von Z*

$$g(\eta) = f(h^{-1}(\eta)) \left| \frac{\partial h^{-1}(\eta)}{\partial \eta} \right|,$$

wobei $\left|\frac{\partial h^{-1}(\eta)}{\partial \eta}\right|$ die Funktionaldeterminante der Transformation h bezeichnet.

BEWEIS:

$$
\begin{aligned}
\mathsf{P}(h(Y) \in B) &= \mathsf{P}(Y \in h^{-1}(B)) \\
&= \int\limits_{h^{-1}(B)} f(\xi)\,\mathrm{d}\xi \\
&= \int\limits_{\underbrace{h(h^{-1}(B))}_{B}} f(h^{-1}(\eta)) \left| \frac{\partial h^{-1}(\eta)}{\partial \eta} \right| \,\mathrm{d}\eta \\
&= \int\limits_{B} f(h^{-1}(\eta)) \left| \frac{\partial h^{-1}(\eta)}{\partial \eta} \right| \,\mathrm{d}\eta
\end{aligned}
$$

Hier benutzt man die Substitution $\eta = h(\xi)$.

□

Folgerung: Für $n = 1$ ist die Dichte

$$f(h^{-1}(\eta)) \frac{1}{|h'(h^{-1}(\eta))|} \, .$$

BEWEIS: Hier benutzen wir

$$\left| \frac{\partial h^{-1}(\eta)}{\partial \eta} \right| = \frac{1}{|h'(h^{-1}(\eta))|} \, .$$

□

Bemerkung: Durch die Berechnung der Dichten transformierter Zufallsgrößen erhält man die explizite Beschreibung induzierter Maße. Wir haben

$$\mathsf{P}^{h\circ Y}(B) = \mathsf{P}^{Y}(h^{-1}(B)) \, ,$$

d. h.

$$\int\limits_{B} \mathsf{P}^{h\circ Y}(\,\mathrm{d}z) = \int\limits_{h^{-1}(B)} \mathsf{P}^{Y}(\,\mathrm{d}y) \, .$$

Existieren für beide Maße die Dichten, dann folgt aus dem vorigen Satz

$$\int\limits_{B} g(\eta)\,\mathrm{d}\eta = \int\limits_{h^{-1}(B)} f(\xi)\,\mathrm{d}\xi \, .$$

Beispiel 6.5

1. Gilt $Z = Y + a$, so gilt

$$\begin{aligned} & h(Y) = Y + a \quad \text{und} \quad \eta = h(\xi) \\ \Rightarrow \quad & h^{-1}(\eta) = \xi = \eta - a\,. \end{aligned}$$

   $\Rightarrow Y + a$ hat die Dichte $g(\eta) = f(\eta - a)$

2. Gilt $Z = cY$, $c \neq 0$, so haben wir

$$g(\eta) = \frac{1}{|c|} f\left(\frac{\eta}{c}\right)$$

Beispiel 6.6 Normalverteilung
Y sei standardisiert normalverteilt im $\mathbb{R}^n$ und $Z = CY + a$ für eine nichtsinguläre $n \times n$ Matrix C und $a \in \mathbb{R}^n$. Dann gilt mit $\Sigma = CC'$

$$g(\eta) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\eta-a)'\Sigma^{-1}(\eta-a)}\,.$$

Bezeichnung der Verteilung: $Z \sim \mathsf{N}(a, \Sigma)$.

6.8.4 Dichte der Summe von Zufallsgrößen

Satz 6.18 *Ist $Y = (Y_1, Y_2)$ Zufallsgröße mit der Dichte $f(\xi_1, \xi_2)$ bzgl. λ_2. Dann hat $Y_1 + Y_2$ die Dichte*

$$f_{Y_1+Y_2}(\xi_2) = \int_{-\infty}^{\infty} f(\xi_1, \xi_2 - \xi_1)\,\mathrm{d}\xi_1\,.$$

Beweis: Wir definieren $\eta = h(\xi) = C\xi$ mit $C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

$\Rightarrow \quad C^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad \Rightarrow$ Dichte von CY ist

$$f(\eta_1, \eta_2 - \eta_1)\,.$$

$$\Rightarrow \quad \mathsf{P}(Z \leq z) = \int_{-\infty}^{z} f(\eta_1, \eta_2 - \eta_1)\,\mathrm{d}\eta_1\,\mathrm{d}\eta_2 \quad \text{für } Z = \begin{pmatrix} Y_1 \\ Y_1 + Y_2 \end{pmatrix}.$$

$$\Rightarrow \quad \mathsf{P}(Y_1 + Y_2 \leq x) = \mathsf{P}(Y_1 \leq \infty, Y_1 + Y_2 \leq x) = \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(\eta_1, \eta_2 - \eta_1)\,\mathrm{d}\eta_1\,\mathrm{d}\eta_2$$

$$= \int_{-\infty}^{x} f_{Y_1+Y_2}(\eta_2)\,\mathrm{d}\eta_2.$$

□

Folgerung: Sind Y_1, Y_2 unabhängige Zufallsgrößen mit Dichten f_1, f_2, so hat $Y_1 + Y_2$ die Dichte

$$\underset{\uparrow}{f_1 * f_2}(v) = \int_{-\infty}^{\infty} f_1(u) f_2(v-u)\,\mathrm{d}u$$

(Faltung von f_1, f_2).

Als Beispiel kann der folgende Satz bewiesen werden.

Satz 6.19 *Sind Y_1, Y_2 unabhängig und ist $Y_i \sim \mathsf{N}(\mu_i, \sigma_i^2)$, so ist $Y_1 + Y_2 \sim \mathsf{N}(\mu, \sigma^2)$ mit $\mu = \mu_1 + \mu_2$, $\sigma^2 = \sigma_1^2 + \sigma_2^2$.*

6.9 Momente von Zufallsgrößen

Zu gegebenen Zufallsgrößen sind die Verteilungen berechenbar, jedoch möchte man häufig nur einige Charakterisierungsgrößen der Verteilungen berechnen oder für den Vergleich von Verteilungen nur spezielle Werte nutzen.

Sei speziell Y eine reellwertige, diskrete Zufallsgröße mit Werten $y_1, y_2, \ldots$ und $\mathsf{P}^Y(y_i) = p_i$. Sucht man einen mittleren Wert der Zufallsgröße Y, so scheint die gewichtete Summe

$$\sum_{i=1}^{\infty} y_i p_i = \sum_{i=1}^{\infty} y_i \mathsf{P}^Y(y_i)$$

geeignet. Bei Gleichverteilung ist das das arithmetische Mittel.

Notwendig ist die absolute Konvergenz, d. h. $\sum\limits_{i=1}^{\infty} |y_i| p_i < \infty$. Man möchte derartige Funktionale nicht nur für diskrete Zufallsgrößen, sondern auch für beliebige Zufallsgrößen definieren. Dazu nutzen wir die äquivalente Schreibweise

$$\sum_{i=1}^{\infty} y_i \mathsf{P}^Y(y_i) = \int_{\mathcal{Y}} y \mathsf{P}^Y(\,\mathrm{d}y)$$

Sei Y eine $\mathbb{R}^1$-wertige Zufallsgröße über $(\Omega, \mathfrak{A}, \mathsf{P})$.

Definition 6.6 *Falls $\int |y|^k \mathsf{P}^Y(\,\mathrm{d}y) < \infty$ für ein ganzes $k > 0$, so heißt*

$$\int y^k \mathsf{P}^Y(\,\mathrm{d}y)$$

k-tes Moment von Y (oder von P^Y).

Hier haben wir ein Integral über $(\mathcal{Y}, \mathfrak{C})$. Es gilt die allgemeine Transformationsformel für Integrale:

$$(\Omega, \mathfrak{A}) \xrightarrow{Y} (\mathcal{Y}, \mathfrak{C}) \xrightarrow{T} (\mathcal{T}, \mathfrak{D})$$

$$\int_D t \mathsf{P}^{T\circ Y}(\,\mathrm{d}t) = \int_{T^{-1}(D)} T(y)\mathsf{P}^Y(\,\mathrm{d}y) = \int_{Y^{-1}(\underbrace{T^{-1}(D)}_{C})} T(Y(\omega))\mathsf{P}(\,\mathrm{d}\omega)$$

Dies läßt sich auch mit Dichten = Radon-Nikodym-Ableitungen schreiben.
Oder allgemeiner:

$$\int_B h(t)\mathsf{P}^{T\circ Y}(\,\mathrm{d}t) = \int_{T^{-1}(D)} h(T(y))\mathsf{P}^Y(\,\mathrm{d}y) = \int_{Y^{-1}(T^{-1}(D))} h(T(Y(\omega)))\mathsf{P}(\,\mathrm{d}\omega)\,.$$

Bezeichnung: Das erste Moment
$\int y\mathsf{P}^Y(\,\mathrm{d}y)$ heißt Erwartungswert von Y. Aus der Transformationsformel folgt:

$$\mathsf{E}Y = \int y\mathsf{P}^Y(\,\mathrm{d}y) = \int Y(\omega)\mathsf{P}(\,\mathrm{d}\omega)\,.$$

$\Rightarrow$ $Y - \mathsf{E}Y = Z$ hat Erwartungswert 0, denn

$$\int |y - \mathsf{E}Y|\mathsf{P}^Y(\,\mathrm{d}y) \le \int |y|\mathsf{P}^Y(\,\mathrm{d}y) + |\mathsf{E}Y| < \infty \quad \Rightarrow \quad \mathsf{E}Z \text{ existiert.}$$

$$\mathsf{E}Z = \int Z(\omega)\mathsf{P}(\,\mathrm{d}\omega) = \int Y(\omega)\mathsf{P}(\,\mathrm{d}\omega) - \mathsf{E}Y = 0\,.$$

Berechnung des Erwartungswertes von $B(n, \vartheta)$:

$$\mathsf{P}(Y = k) = \binom{n}{k}\vartheta^k(1-\vartheta)^{n-k}$$

$$\sum_{k=0}^{n} k\binom{n}{k}\vartheta^k(1-\vartheta)^{n-k} = (*)$$

$$k\binom{n}{k} = k\frac{n!}{k!(n-k)!} = n\binom{n-1}{k-1}$$

$$\begin{aligned}
(*) &= \sum_{k=1}^{n} n\binom{n-1}{k-1}\vartheta^k(1-\vartheta)^{n-k} \\
&= n\sum_{k=0}^{n-1}\binom{n-1}{k}\vartheta^{k+1}(1-\vartheta)^{n-(k+1)} \\
&= n\sum_{k=0}^{n-1}\binom{n-1}{k}\vartheta^{k}(1-\vartheta)^{n-1-k}\vartheta \\
&= n\vartheta
\end{aligned}$$

Definition 6.7 *Falls* $\int |y - \mathsf{E}Y|^k \mathsf{P}^Y(\,\mathrm{d}y) < \infty$, *so heißt*

$$\mathsf{E}(Y - \mathsf{E}Y)^k = \int (y - \mathsf{E}Y)^k \mathsf{P}^Y(\,\mathrm{d}y)$$

k-tes zentrales Moment.

Speziell: $k = 2$

Wenn $\int (y - \mathsf{E}Y)^2 \mathsf{P}^Y(\,\mathrm{d}y) < \infty$, so heißt $\int (y - \mathsf{E}Y)^2 \mathsf{P}^Y(\,\mathrm{d}y) = \mathsf{Var}\, Y$ Varianz von Y.

$\sqrt{\mathsf{Var}\, Y}$ heißt Standardabweichung.

6.9.1 Erwartungswerte

Satz 6.20 *Ist φ eine meßbare Abbildung $\mathcal{Y} \to \mathbb{R}^1$, so ist*

$$\int \varphi(y)\mathsf{P}^Y(\,\mathrm{d}y)$$

der Erwartungswert $\mathsf{E}\varphi(Y)$ von $\varphi(Y)$ bei $\int |\varphi(y)|\mathsf{P}^Y(\,\mathrm{d}y)$.

Beweis: Es gilt $\varphi \circ Y : (\Omega, \mathfrak{A}) \to (\mathbb{R}^1, \mathfrak{B}^1)$ und so ist $\varphi \circ Y$ eine Zufallsgröße und es gilt

$$\int \varphi(y)\mathsf{P}^Y(\,\mathrm{d}y) = \int \varphi(Y(\omega))\mathsf{P}(\,\mathrm{d}\omega) = \mathsf{E}\varphi \circ Y = \mathsf{E}\varphi(Y)\,.$$

□

Beispiel 6.7 $\varphi(Y) = Y_1 + Y_2 + \cdots + Y_n, \quad Y = (Y_1, Y_2, \dots, Y_n)$

$$
\begin{aligned}
\Rightarrow \int \varphi(y) \mathsf{P}^Y(\,\mathrm{d}y) &= \int \varphi(Y(\omega)) \mathsf{P}(\,\mathrm{d}\omega) \\
&= \int Y_1(\omega) \mathsf{P}(\,\mathrm{d}\omega) + \cdots + \int Y_n(\omega) \mathsf{P}(\,\mathrm{d}\omega) \\
&= \mathsf{E}Y_1 + \mathsf{E}Y_2 + \cdots + \mathsf{E}Y_n
\end{aligned}
$$

Der Erwartungswert einer Summe von Zufallsgrößen ist die Summe der Erwartungswerte.

Beispiel 6.8 Binomialverteilung
Wir hatten speziell ausgerechnet $\mathsf{E}S = n\vartheta$ für $S = \sum Y_i$, $Y_i \sim B(1, \vartheta)$. Das folgt nun aus vorhergehendem Beispiel.

Das, was für die Summe gilt, gilt nicht für Produkte.
So ist im allgemeinen $\mathsf{E}Y_1 \cdots Y_n \neq (\mathsf{E}Y_1) \cdots (\mathsf{E}Y_n)$. Besonders wichtig sind die Erwartungswerte des Produkts zweier Zufallsgrößen (siehe unter Kovarianz und Korrelation).

Lemma 1 *Falls $\mathsf{E}Y^k$ existiert, so existiert $\mathsf{E}Y^\ell$ für $\ell < k$.*

Beweis: Sei $Z(\omega) = \begin{cases} 1 & \text{für } |Y(\omega)| \leq 1 \\ Y(\omega) & \text{für } |Y(\omega)| > 1 \end{cases}$

$\Rightarrow \quad |Y(\omega)| \leq |Z(\omega)| \quad \Rightarrow \quad |Y|^\ell \leq |Z|^\ell$

Es genügt zu zeigen $\mathsf{E}|Z|^\ell < \infty$.

$$
\mathsf{E}|Z|^\ell = \int |Z(\omega)|^\ell \mathsf{P}(\,\mathrm{d}\omega) = \int\limits_{|Y(\omega)| \leq 1} 1 \mathsf{P}(\,\mathrm{d}\omega) + \int\limits_{|Y(\omega)| > 1} |Y(\omega)|^\ell \mathsf{P}(\,\mathrm{d}\omega) < \infty
$$

□

Lemma 2 *Falls $\mathsf{E}Y^k$ existiert, so existiert $\mathsf{E}(Y-a)^k$ für beliebige $a \in \mathbb{R}^1$.*

Beweis: Es gilt der Binomische Satz

$$
\begin{aligned}
(Y-a)^k &= \sum_{i=0}^{k} \binom{k}{i} Y^i (-a)^{k-i} \\
|Y-a|^k &\leq \sum_{i=0}^{k} \binom{k}{i} |Y|^i |a|^{k-i} \\
\mathsf{E}|Y-a|^k &< \infty \qquad \text{wegen vorigem Lemma.}
\end{aligned}
$$

□

Folgerung: Wenn für eine Zufallsgröße V für irgendeine Zahl κ das k-te Moment $\mathsf{E}(V-\kappa)^k$ existiert, so existieren alle Momente niedrigerer Ordnung von V mit beliebigem κ_1, d. h. aus $\mathsf{E}(V-\kappa)^k < \infty$ folgt die Existenz von $\mathsf{E}(V-\kappa_1)^\ell$ für $\ell \leq k$.

Speziell folgt, daß für beschränkte Zufallsgrößen Y alle Momente beliebiger Ordnung, unabhängig davon, ob zentral oder nichtzentral, existieren.

6.9.2 Rechenregeln für Erwartungswerte

Seien Z_1, Z_2 reellwertige Zufallsgrößen mit existierenden Erwartungswerten. Dann gilt:

1. $\mathsf{E}(\lambda Z_1) = \lambda \mathsf{E} Z_1$
2. $\mathsf{E}(\alpha Z_1 + \beta Z_2) = \alpha \mathsf{E} Z_1 + \beta \mathsf{E} Z_2$
3. Sind Z_1, Z_2 unabhängig, so gilt $\mathsf{E} Z_1 Z_2 = \mathsf{E} Z_1 \mathsf{E} Z_2$.
4. $Z_1 \leq Z_2 \quad \Rightarrow \quad \mathsf{E} Z_1 \leq \mathsf{E} Z_2$
5. $Z_2 \geq 0 \quad \Rightarrow \quad \mathsf{E} Z_2 \geq 0$
6. Gilt für $Z_2 \geq 0$ aber $\mathsf{E} Z_2 = 0$, so $Z_2 = 0$ P-fast-sicher.

Beweis: 1. + 2. klar.
zu 3. Sei $Z = (Z_1, Z_2)$.
P^Z ist induziertes Maß, d. h. für $C \in \mathfrak{B}^2$ gilt $\mathsf{P}^Z(C) = \mathsf{P}(\omega : Z(\omega) \in C) = \mathsf{P}(Z^{-1}(C))$.
Speziell für $C = C_1 \times C_2$ für $C_i \in \mathfrak{B}^1$ gilt

$$\begin{aligned}
\mathsf{P}^Z(C_1 \times C_2) &= \mathsf{P}(\omega : Z_1(\omega) \in C_1, Z_2(\omega) \in C_2) \\
&\overset{\substack{\text{unabh.}\\ \downarrow}}{=} \mathsf{P}(\omega : Z_1(\omega) \in C_1)\mathsf{P}(\omega : Z_2(\omega) \in C_2) \\
&= \mathsf{P}^{Z_1}(C_1)\mathsf{P}^{Z_2}(C_2) \\
\Rightarrow \qquad \mathsf{P}^Z(\mathrm{d}z) &= \mathsf{P}^{Z_1}(\mathrm{d}z_1)\mathsf{P}^{Z_2}(\mathrm{d}z_2) \\
\Rightarrow \qquad \mathsf{E}|Z_1 Z_2| &= \int |Z_1(\omega)||Z_2(\omega)|\mathsf{P}(\mathrm{d}\omega) = \int\limits_{\mathbb{R}^2} |Z_1||Z_2|\mathsf{P}^Z(\mathrm{d}z) \\
&= \int\limits_{\mathbb{R}^1}\int\limits_{\mathbb{R}^1} |Z_1||Z_2|\mathsf{P}^{Z_1}(\mathrm{d}z_1)\mathsf{P}^{Z_2}(\mathrm{d}z_2) \\
&= \int\limits_{\mathbb{R}^1} |Z_1|\mathsf{P}^{Z_1}(\mathrm{d}z_1) \int\limits_{\mathbb{R}^1} |Z_2|\mathsf{P}^{Z_2}(\mathrm{d}z_2) < \infty
\end{aligned}$$

$$\Rightarrow \qquad \mathsf{E} Z_1 Z_2 = \int\limits_{\mathbb{R}^1} Z_1 \mathsf{P}^{Z_1}(\mathrm{d}z_1) \int\limits_{\mathbb{R}^1} Z_2 \mathsf{P}^{Z_2}(\mathrm{d}z_2) = \mathsf{E} Z_1 \mathsf{E} Z_2\,.$$

4., 5. und 6. folgen unmittelbar.

□

6.9.3 Kovarianz und Korrelation

Es seien Z_1, Z_2 Zufallsgrößen: $Z_i : (\Omega, \mathfrak{A}) \to (\mathcal{Y}_i, \mathfrak{C}_i)$

Voraussetzung: Es sei $\mathsf{E}Z_i = 0$ und $\mathsf{E}(Z_i^2) < \infty$. (V)

Satz 6.21 *Cauchy-Schwarzsche Ungleichung*
Unter (V) gilt

$$(\mathsf{E}(Z_1 Z_2))^2 \leq \mathsf{E}(Z_1^2)\mathsf{E}(Z_2^2)\,. \qquad (*)$$

Gleichheit gilt genau dann, wenn reelle Zahlen a, b existieren mit

$$\mathsf{E}(Z_1 - (a + bZ_2))^2 = 0 \text{ oder } \mathsf{E}Z_i^2 = 0 \text{ für ein } i,$$

d. h. es gilt P*-fast-sicher*

$$Z_1 = a + bZ_2 \text{ oder } Z_i = 0 \text{ für ein } i.$$

Beweis:

a) Es existiert $\mathsf{E}(Z_1 Z_2)$, denn

$$\mathsf{E}|Z_1 Z_2| = \int |Z_1(\omega) Z_2(\omega)| \mathsf{P}(\,\mathrm{d}\omega) \leq \frac{1}{2} \int (|Z_1^2(\omega)| + |Z_2^2(\omega)|) \mathsf{P}(\,\mathrm{d}\omega) < \infty\,.$$

b) Setze $\alpha = \mathsf{E}Z_2^2$, $\beta = -\mathsf{E}Z_1 Z_2 \Rightarrow |\beta| \neq \infty$

1. $\alpha = 0$
$\Rightarrow \int Z_2^2(\omega)\mathsf{P}(\,\mathrm{d}\omega) = 0 \Rightarrow Z_2(\omega) = 0$ bis auf Nullmenge bezüglich P.
$\Rightarrow \mathsf{E}Z_1 Z_2 = 0 \Rightarrow 0 = 0$.
2. $\alpha > 0$
Setze $V = \alpha Z_1 + \beta Z_2$
Wegen $V^2 \leq \alpha^2 Z_1^2 + \beta^2 Z_2^2 + \underbrace{2\alpha|\beta||Z_1||Z_2|}_{\leq \alpha|\beta|(Z_1^2 + Z_2^2)} \leq (\alpha^2 + \alpha|\beta|)Z_1^2 + (\beta^2 + \alpha|\beta|)Z_2^2$
gilt $\mathsf{E}V^2 < \infty$, d. h. $\mathsf{E}V^2$ existiert, denn für $U_1(\omega) \leq U_2(\omega)$ gilt $\mathsf{E}U_1 \leq \mathsf{E}U_2$.

$$\begin{aligned} V^2 &= \alpha^2 Z_1^2 + \beta^2 Z_2^2 + 2\alpha\beta Z_1 Z_2 \\ \mathsf{E}V^2 &= \alpha^2 \mathsf{E}Z_1^2 + \beta^2 \mathsf{E}Z_2^2 + 2\alpha\beta \mathsf{E}Z_1 Z_2 \\ &= \alpha^2 \mathsf{E}Z_1^2 + \alpha\beta^2 - 2\alpha\beta^2 \\ &= \alpha\,[\mathsf{E}Z_2^2 \mathsf{E}Z_1^2 - (\mathsf{E}Z_1 Z_2)^2] \\ &\geq 0\,. \end{aligned}$$

Gilt in $(*)$ „=“, so ist $\mathsf{E}V^2 = 0$, d. h. $\mathsf{E}(Z_1 + \frac{\beta}{\alpha} Z_2)^2 = 0$
$\Rightarrow$ Behauptung

$\square$

Sei Z_1 Zufallsgröße.
$\Rightarrow \mathsf{E}Z_1^2 = 0$ bedeutet $Z_1(\omega) = 0$ bis auf P-Nullmengen. Das ist der triviale Fall, denn Z_1 ist im wesentlichen $= 0$. Wenn wir diesen Fall ausschließen, so gilt $\mathsf{E}Z_1^2 > 0$.

Definition 6.8 *Unter (V) heißt* $\mathsf{E}Z_1Z_2 =: \mathsf{Cov}\,(Z_1, Z_2)$ *die Kovarianz von* Z_1, Z_2. *Mit* $\sigma_i^2 = \mathsf{E}Z_i^2$ *und* $\sigma_i^2 \neq 0$ *heißt*

$$\varrho_{Z_1,Z_2} := \frac{\mathsf{Cov}\,(Z_1, Z_2)}{\sigma_1\sigma_2}$$

Korrelation zwischen Z_1, Z_2.

Rechtfertigung: Wenn $\mathsf{E}Z_i^2 < \infty$, d. h. die Varianz von Z_i existiert, so existiert wegen vorigem Satz die Kovarianz. Wegen des Satzes gilt $|\varrho_{Z_1,Z_2}| \leq 1$ und $|\varrho_{Z_1,Z_2}| = 1$ nur dann, wenn P-fast-sicher $Z_1 = 0$ oder $Z_2 = 0$ oder wenn $Z_1 = a + bZ_2$ für Konstanten a, b gilt.

Man kann die Kovarianz für beliebige Zufallsgrößen Y_i, für die das zweite Moment existiert, definieren. Denn existiert $\mathsf{E}Y_i^2$, so gilt für $Z_i = Y_i - \mathsf{E}Y_i$:
$\mathsf{E}Z_i = 0$ und $\mathsf{E}Z_i^2 < \infty$.

Wir haben $\mathsf{Var}\,Y_i = \mathsf{E}Z_i^2$.

$$\Rightarrow \qquad |\underbrace{\mathsf{E}(Y_1 - \mathsf{E}Y_1)(Y_2 - \mathsf{E}Y_2)}_{\mathsf{Cov}\,(Y_1 - \mathsf{E}Y_1, Y_2 - \mathsf{E}Y_2)}|^2 \leq \underbrace{\mathsf{E}(Y_1 - \mathsf{E}Y_1)^2}_{\mathsf{Var}\,Y_1}\underbrace{\mathsf{E}(Y_2 - \mathsf{E}Y_2)^2}_{\mathsf{Var}\,Y_2}$$

Da die Kovarianz immer für zentrierte Zufallsgrößen definiert ist, schreibt man auch

$$\mathsf{Cov}\,(Y_1, Y_2) := \mathsf{Cov}\,(Y_1 - \mathsf{E}Y_1, Y_2 - \mathsf{E}Y_2)$$

$\Rightarrow$ Äquivalente Definition:

Definition 6.9 *Sei* $\mathsf{E}Y_i^2 < \infty$. *Dann ist* $\mathsf{Cov}\,(Y_1, Y_2) = \mathsf{E}(Y_1 - \mathsf{E}Y_1)(Y_2 - \mathsf{E}Y_2)$ *die Kovarianz von* Y_1, Y_2 *und*

$$\varrho_{Y_1,Y_2} = \frac{\mathsf{Cov}\,(Y_1, Y_2)}{\sqrt{\mathsf{Var}\,Y_1}\sqrt{\mathsf{Var}\,Y_2}}$$

die Korrelation.

Y_1, Y_2 heißen unkorreliert, falls $\mathsf{Cov}\,(Y_1, Y_2) = 0$.

Positive Korrelation bedeutet, daß mehrheitlich $Y_1 - \mathsf{E}Y_1$ und $Y_2 - \mathsf{E}Y_2$ gleiche Vorzeichen haben. Gilt $|\varrho_{Y_1,Y_2}| = 1$, so liegen die Werte von $Y_1 - \mathsf{E}Y_1$ und $Y_2 - \mathsf{E}Y_2$ auf einer Geraden.

Satz 6.22 *Sind* Y_1, Y_2 *unabhängig, so sind* Y_1, Y_2 *unkorreliert.*

BEWEIS: Da Y_1, Y_2 unabhängig sind, gilt dies auch für $Y_1 - \mathsf{E}Y_1$ und $Y_2 - \mathsf{E}Y_2$. Damit gilt

$$\mathsf{E}(Y_1 - \mathsf{E}Y_1)(Y_2 - \mathsf{E}Y_2) = \mathsf{E}(Y_1 - \mathsf{E}Y_1) \cdot \mathsf{E}(Y_2 - \mathsf{E}Y_2) = 0\,.$$

□

Die Umkehrung dieses Satzes gilt nicht.

Gegenbeispiel: Sei $\Omega = \{1, 2, 3, 4\}$
$Y_1(1) = -2, Y_1(2) = -1, Y_1(3) = 1, Y_1(4) = 2, \quad Y_2(i) = Y_1^2(i), \quad \mathsf{P}(i) = \frac{1}{4}$.
$\Rightarrow \quad \mathsf{E}Y_1 = 0$ und $\varrho_{Y_1,Y_2} = 0$.
Aber Y_1 bestimmt eindeutig die Werte von Y_2.

6.9.4 Rechenregeln für Varianz und Kovarianz

Satz 6.23 *Seien a, b, c, d reelle Zahlen*
$Y, Y_1, Y_2, \ldots, Y_n$ Zufallsgrößen mit existierenden Varianzen $\Rightarrow$

1. $\mathsf{Var}\, Y = \mathsf{E}Y^2 - (\mathsf{E}Y)^2$
2. $\mathsf{Var}\,(aY + b) = a^2 \mathsf{Var}\, Y$
3. $\mathsf{Cov}\,(Y_1, Y_2) = \mathsf{E}Y_1Y_2 - \mathsf{E}Y_1\mathsf{E}Y_2$
4. $\mathsf{Cov}\,(aY_1 + b, cY_2 + d) = ac\,\mathsf{Cov}\,(Y_1, Y_2)$
5. $\mathsf{Var}\,(Y_1 + \cdots + Y_n) = \sum_{i=1}^{n} \mathsf{Var}\, Y_i + \sum_{i \neq j} \mathsf{Cov}\,(Y_i, Y_j)$

BEWEIS: direkte Folgerungen

□

Beispiel 6.9 Binomialverteilung
Sei $Y_i \sim B(1, \vartheta)$ und seien $Y_1, \ldots, Y_n$ unabhängig. Setze $S = \sum Y_i \quad \Rightarrow \quad S \sim B(n, \vartheta)$
$\Rightarrow \quad \mathsf{E}S = \sum \mathsf{E}Y_i = n\vartheta$.

$$\mathsf{Var}\, S = \sum \mathsf{Var}\, Y_i, \quad \begin{aligned} \mathsf{Var}\, Y_i &= \mathsf{E}(Y_i - \vartheta)^2 = \mathsf{E}Y_i^2 - \vartheta^2 \\ &= \vartheta - \vartheta^2 = \vartheta(1 - \vartheta) \end{aligned}$$

$\Rightarrow \quad \mathsf{Var}\, S = n\vartheta(1 - \vartheta)$.

6.9.5 Unabhängigkeit bei Normalverteilung

Beispiel 6.10 Die Dichte von Y sei

$$\mathsf{Pr}\vartheta(y) = \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right)$$

für ein $\mu \in \mathbb{R}^p$ und eine nichtsinguläre, positiv definite $\Sigma \,:\, p \times p$.

$$\vartheta = (\mu, \Sigma)$$

Frage: Sind die Komponenten von Y unabhängig?

- hängt ab von Σ, bei Diagonalgestalt ja.

<u>Bestimmung des Erwartungswertes von Y</u>

$$\int p_\vartheta(y) y \,\mathrm{d}y = \int \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right) \,\mathrm{d}y$$

y Wir hatten früher bei den Dichten:
Hat U die Dichte $f(\xi)$, d. h.

$$\mathsf{P}(U \leq z) = \int_{-\infty}^{z} f(\xi)\,\mathrm{d}\xi\,,$$

so hat $U + a$ die Dichte $f(z - a)$.
$Z = CU + a$ hat bei nichtsingulärer Matrix $C \,:\, p \times p$ die Gestalt:

$$\text{es ist} \quad Z = h(U) = CU + a$$

$$\Rightarrow \quad U = C^{-1}(Z - a) = h^{-1}(Z)$$

$\Rightarrow$ Dichte von Z ist

$$f(C^{-1}(\eta - a)) \left| \frac{\partial h^{-1}(\eta)}{\partial \eta} \right|$$

Es gilt

$$\frac{\partial h^{-1}(\eta)}{\partial \eta} = \left| \frac{\partial C^{-1}(\eta - a)}{\partial \eta} \right| = ||C^{-1}||$$

— $\Rightarrow$ Dichte von $Z = CU + a$ ist

$$f(C^{-1}(\eta - a))|C^{-1}|$$

Beispiel war: Normalverteilte Variable $U_1, \ldots, U_p$, unabhängig.

Hat u_i die Dichte $f_i(u_i) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}u_i^2}$, so hat U die Dichte

$$\begin{aligned}
\prod f_i(u_i) = f(u) &= \frac{1}{\left(\sqrt{2\pi}\right)^p} \mathrm{e}^{-\frac{1}{2}\sum\limits_i u_i^2} \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^p} \mathrm{e}^{-\frac{1}{2}|u|^2}
\end{aligned}$$

$Z = CU + a$ hat die Dichte

$$\begin{aligned}
f(C^{-1}(\eta - a))|C^{-1}| &= \frac{1}{\left(\sqrt{2\pi}\right)^p |C|} \mathrm{e}^{-\frac{1}{2}|C^{-1}(\eta-a)|^2} \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^p |\Sigma|^{\frac{1}{2}}} \mathrm{e}^{-\frac{1}{2}(\eta-a)'\Sigma^{-1}(\eta-a)} \qquad \text{für } \Sigma = CC' \qquad (*)
\end{aligned}$$

$\Rightarrow$ Unsere gegebene Variable Y hat bei $a = \mu$ dieselbe Verteilung wie Z.
$\Rightarrow$ Die Momente von Y und die von Z sind dieselben. Wir bestimmen zuerst $\mathsf{E}U$, $\mathsf{Var}\,U$.

Wir haben $\mathsf{E}U = 0$, $\mathsf{Var}\,U = I$ denn:

$$\int \cdots \int \mathrm{e}^{-\frac{1}{2}(u_1^2+\cdots+u_p^2)} \,\mathrm{d}u_1 \ldots \mathrm{d}u_p = \left(\sqrt{2\pi}\right)^p$$

und

$$\int \cdots \int u_1 \,\mathrm{e}^{-\frac{1}{2}(u_1^2+\cdots+u_p^2)} \,\mathrm{d}u_1 \ldots \mathrm{d}u_p = \left(\sqrt{2\pi}\right)^{p-1} \underbrace{\int u_1 \,\mathrm{e}^{-\frac{1}{2}u_1^2} \,\mathrm{d}u_1}_{=0\,.}$$

Analog die anderen Komponenten. $\Rightarrow$ $\mathsf{E}U = 0$.

$$\begin{aligned}
\mathsf{Var}\,U_i &= \frac{1}{\left(\sqrt{2\pi}\right)^p} \int \cdots \int u_1^2 \,\mathrm{e}^{-\frac{1}{2}|u|^2} \,\mathrm{d}u_1 \ldots \mathrm{d}u_p \\
&= \frac{1}{\sqrt{2\pi}} \int u_1^2 \,\mathrm{e}^{-\frac{1}{2}u_1^2} \,\mathrm{d}u_1 \\
&= \frac{1}{\sqrt{2\pi}} \left(-u_1 \,\mathrm{e}^{-\frac{1}{2}u_1^2} \Big|_{-\infty}^{\infty} + \int \mathrm{e}^{-\frac{1}{2}u_1^2} \,\mathrm{d}u_1 \right) \\
&= 1
\end{aligned}$$

$\mathsf{Cov}\,(U_i, U_j) = \mathsf{E}U_iU_j = \mathsf{E}U_i\mathsf{E}U_j = 0$, wegen Unabhängigkeit. q $\Rightarrow$ $\mathsf{Var}\,U = I_p$.

Nun $\mathsf{E}Y$, $\mathsf{Var}\,Y$. Es gilt $Y \sim Z$

$$\mathsf{E}Y = \mathsf{E}(CU + a) = C\mathsf{E}U + \mu = \mu$$

$$\mathsf{Var}\,Y = \mathsf{Var}\,(CU + a) = \mathsf{Var}\,(CU) = C(\mathsf{Var}\,U)C' = CC'$$

Dichte von Y ist $(*)$

$\Rightarrow \Sigma = \mathsf{Var}\, Y$.

Gesucht ist jetzt eine Matrix D derart, daß $DY = V$ unkorrelierte Komponenten besitzt, z. B.

$$D = \Sigma^{-\frac{1}{2}}\,.$$

Dann gilt $\mathsf{Var}\, V = I$. Bestimmung der Dichte ist

$$\frac{1}{\left(\sqrt{2\pi}\right)^p} \exp\left(-\frac{1}{2}(v-a)'(v-a)\right)\,, \qquad a = D\mu$$

Diese hat Produktform. $\Rightarrow$ Die unkorrelierten Komponenten sind unabhängig. Das gilt nur bei der Normalverteilung.

7 Tschebyscheffsche Ungleichung

Für Zufallsgrößen mit existierendem 2. Moment gilt eine universelle Ungleichung.

Satz 7.1 *Sei Y beliebige Zufallsgröße mit $\mathsf{E}Y = \mu$, $\mathsf{Var}\, Y = \sigma^2 < \infty$. Dann gilt*

$$\mathsf{P}(|Y-\mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}\,, \qquad \lambda > 0$$

(Tschebyscheffsche Ungleichung).

Beweis: Es gilt

$$\begin{aligned}\sigma^2 &= \int |Y(\omega)-\mu|^2 \mathsf{P}(\,\mathrm{d}\omega) \geq \int\limits_{\{\omega : |Y(\omega)-\mu| \geq \lambda\sigma\}} |Y(\omega)-\mu|^2 \mathsf{P}(\,\mathrm{d}\omega)\\ &\geq \lambda^2\sigma^2\mathsf{P}(|Y(\omega)-\mu| \geq \lambda\sigma)\,.\end{aligned}$$

□

8 Momente mehrdimensionaler Zufallsgrößen

Sei $Y = (Y_1, \dots, Y_p)'$ eine p-dimensionale Zufallsgröße, wobei die zweiten Momente der Y_i existieren sollen. Dann nutzen wir die Bezeichnungen:

$$\text{Erwartungswert von } Y = \mathsf{E}Y := \begin{pmatrix} \mathsf{E}Y_1 \\ \vdots \\ \mathsf{E}Y_n \end{pmatrix},$$

$$\text{Varianz von } Y = \mathsf{Var}\, Y = \begin{pmatrix} \mathsf{Var}\, Y_1 & \mathsf{Cov}\,(Y_1, Y_2) & \cdots & \mathsf{Cov}\,(Y_1, Y_p) \\ \vdots & \ddots & & \vdots \\ \mathsf{Cov}\,(Y_p, Y_1) & \cdots & & \mathsf{Var}\, Y_p \end{pmatrix}.$$

$\mathsf{Var}\, Y$ wird auch Kovarianzmatrix von Y genannt. Diese Definitionen sind gerechtfertigt wegen der Rechenregeln für Erwartungswerte und Varianzen, denn es gilt

$$\mathsf{E}\left(\sum \ell_i Y_i\right) = \ell' \mathsf{E}Y \qquad \forall \ell \in \mathbb{R}^p,$$
$$\mathsf{Var}\,(\ell' Y) = \ell' \mathsf{Var}\, Y \ell \qquad \forall \ell \in \mathbb{R}^p.$$

Sei $W := \mathsf{Var}\, Y$. Dann gelten die folgenden Aussagen.

Satz 8.1 $\mathsf{Var}\,(\ell' Y)$ *ist eine quadratische Form und es gilt für beliebige* $\ell \in \mathbb{R}^p$*:*

$$\mathsf{Var}\,(\ell' Y) = \ell' W \ell$$

Satz 8.2 *Wenn für beliebige* ℓ : $\mathsf{Var}\,(\ell' Y) = \ell' A \ell$ *gilt und wenn* A *symmetrisch ist, so ist* A *eindeutig bestimmt zu* $A = W$.

Folgerung: $W = \mathsf{E}\{(Y - \mathsf{E}Y)(Y - \mathsf{E}Y)'\}$.

Satz 8.3 *Die Kovarianzmatrix* W *ist positiv semidefinit.*

Satz 8.4 *Wenn die Komponenten von* Y *unkorreliert sind, so ist* W *diagonal.*

Folgerung: Sind die Y_i unabhängig und identisch verteilt mit $\mathsf{Var}\, Y_i = \sigma^2$, so gilt $W = \sigma^2 I$.

Satz 8.5 *Es gilt für beliebige* $\ell, m \in \mathbb{R}^p$*:*

$$\mathsf{Cov}\,(\ell' Y, m' Y) = \ell' W m$$

Definition 8.1 *Sind* Y *und* Z *Zufallsgrößen mit Werten in* $\mathbb{R}^p$ *bzw.* $\mathbb{R}^r$*, so heißt*

$$\mathsf{Cov}\,(Y, Z) = \mathsf{E}\{(Y - \mathsf{E}Y)(Z - \mathsf{E}Z)'\} =: U$$

Kovarianz von Y, Z.

$\Rightarrow \quad \mathsf{Cov}\,(Y,Y) = \mathsf{Var}\,Y$

Satz 8.6 *Für beliebige $\ell \in \mathbb{R}^p, m \in \mathbb{R}^r$ gilt*

$$\mathsf{Cov}\,(\ell'Y, m'Z) = \ell' U m\,.$$

Für beliebige $L\,:\,c \times n, M\,:\,t \times r$ gilt

$$\mathsf{Cov}\,(LY, MZ) = LUM'\,.$$

o. B.

Im allgemeinen benutzt man solche geschlossenen Darstellungen für die Momente nur für die ersten beiden Momente.

9 Charakteristische Funktionen

Bisher wurden reellwertige Zufallsgrößen ξ betrachtet. Funktionen heißen komplexwertig, wenn $\xi = \xi_1 + \mathrm{i}\,\xi_2$ für reellwertige Zufallsgrößen ξ_1, ξ_2 gilt. Es gibt dann die eineindeutige Zuordnung zwischen den komplexwertigen Größen und zufälligen Vektoren:

$$\xi(\omega) = \xi_1(\omega) + \mathrm{i}\,\xi_2(\omega) \quad \Leftrightarrow \quad (\xi_1, \xi_2) \quad \text{zufälliger Vektor.}$$

Damit übertragen sich die Rechenregeln für zweidimensionale Zufallsgrößen auf die komplexwertigen Zufallsgrößen, z. B.: $\quad \mathsf{E}\xi \;=\; \mathsf{E}\xi_1 + \mathrm{i}\,\mathsf{E}\xi_2$.

Komplexwertige Zufallsgrößen ξ, η heißen unabhängig, falls $(\xi_1, \xi_2, \eta_1, \eta_2)$ unabhängig sind. Aus der Unabhängigkeit folgt wie bei den reellwertigen Zufallsgrößen $\mathsf{E}\xi\eta = \mathsf{E}\xi\;\mathsf{E}\eta$.

Spezielle komplexwertige Zufallsgrößen sind $\mathrm{e}^{\mathrm{i}tX}$ für reellwertige Zufallsgrößen X und $t \in \mathbb{R}$. Durch Erwartungswertbildung erhält man die sogenannten *charakteristischen Funktionen.* Diese charakterisieren Verteilungen eindeutig und sind für die Bestimmung von Grenzverteilungen, die Untersuchung von Summen von Zufallsgrößen und die Herleitung von Momenten wichtig.

Gnedenko	Lehrbuch der Wahrscheinlichkeitsrechnung, Akademie Verlag
E. Lukacs	Charakteristische Funktionen, Wiley Verlag
Rao C.R.	Lineare statistische Methoden, Wiley Verlag.

Definition 9.1 *Als charakteristische Funktion (c. F.) einer reellen Zufallsgröße X bezeichnet man die komplexwertige Funktion*

$$\varphi_X(t) = \mathsf{E}\exp(\mathrm{i}\,tX) = \mathsf{E}\,\mathrm{e}^{\mathrm{i}\,tX} = \int\limits_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,tz}\,\mathrm{d}F(z)\,,$$

wobei $t \in \mathbb{R}^1$, $\mathrm{i} = \sqrt{-1}$, $\mathrm{e}^{\mathrm{i}\,tz} = \cos(tz) + \mathrm{i}\,\sin(tz)$.

In der Theorie der Integraltransformation ist φ_X die Fourier-Stieltjes-Transformation der Funktion F. Wenn X eine Dichte p bezüglich des Lebesgue-Maßes hat, so gilt

$$\varphi_X(t) \quad = \quad \int\limits_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\,tz} p(z)\,\mathrm{d}z \qquad \text{Fourier-Transformation.}$$

Rechtfertigung: φ_X existiert, da

$$\int\limits_{-\infty}^{\infty} |\,\mathrm{e}^{\mathrm{i}\,tz}|\,\mathrm{d}F(z) = 1$$

Beispiel 9.1 Normalverteilung: $X \sim \mathsf{N}(a, \sigma^2)$

$$\begin{aligned}
\mathsf{E}\,\mathrm{e}^{\mathrm{i}\,tX} &= \frac{1}{\sigma\sqrt{2\pi}} \int \mathrm{e}^{\mathrm{i}\,tz - \frac{(z-a)^2}{2\sigma^2}}\,\mathrm{d}z = \varphi_X(t) \\
& \quad \text{Substitution:} \quad v = \frac{z-a}{\sigma} - \mathrm{i}\,t\sigma\,, \quad \mathrm{d}v = \frac{1}{\sigma}\,\mathrm{d}z \\
v^2 &= \frac{(z-a)^2}{\sigma^2} - t^2\sigma^2 - 2\,\mathrm{i}\,t(z-a) \\
v^2 &= \frac{(z-a)^2}{\sigma^2} - 2\,\mathrm{i}\,tz - t^2\sigma^2 + 2\,\mathrm{i}\,ta \\
-\frac{1}{2}v^2 &= -\frac{(z-a)^2}{2\sigma^2} + \mathrm{i}\,tz + \frac{t^2\sigma^2}{2} - \mathrm{i}\,ta \\
\Rightarrow \varphi_X(t) &= \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{\mathrm{i}\,at - \frac{1}{2}\sigma^2 t^2} \underbrace{\int\limits_{-\infty - \mathrm{i}\,t\sigma}^{\infty - \mathrm{i}\,t\sigma} \mathrm{e}^{-\frac{v^2}{2}}\,\mathrm{d}v}_{\sqrt{2\pi}} \\
&= \mathrm{e}^{\mathrm{i}\,at - \frac{1}{2}\sigma^2 t^2}
\end{aligned}$$

9.1 Eigenschaften

Satz 9.1 *Eine c. F. ist auf der ganzen reellen Achse gleichmäßig stetig und es ist*

$$\varphi_X(0) = 1\,, \quad |\varphi_X(t)| \le 1\,.$$

Beweis: Gleichmäßige Stetigkeit:

$$\begin{aligned}
|\varphi_X(t+h) - \varphi_X(t)| &= \left| \int \mathrm{e}^{\mathrm{i}\,tz}\left(\mathrm{e}^{\mathrm{i}\,hz} - 1\right)\,\mathrm{d}F(z) \right| \\
&\le \int \underbrace{|\,\mathrm{e}^{\mathrm{i}\,tz}|}_{=1}\,|\,\mathrm{e}^{\mathrm{i}\,hz} - 1|\,\mathrm{d}F(z)\,.
\end{aligned}$$

Wegen $|\mathrm{e}^{\mathrm{i}\,hz}-1| \le 2$ folgt bei geeigneter Wahl von a:

$$\le \; 2\underbrace{\int\limits_{-\infty}^{-a} \mathrm{d}F(z)}_{\le\frac{\varepsilon}{5}} + 2\underbrace{\int\limits_{a}^{\infty} \mathrm{d}F(z)}_{\le\frac{\varepsilon}{5}} + \int\limits_{-a}^{a} |\mathrm{e}^{\mathrm{i}\,hz}-1|\,\mathrm{d}F(z)\,.$$

Wir wählen dann δ so, daß $|\mathrm{e}^{\mathrm{i}\,\delta z}-1| \le \frac{\varepsilon}{5}$ für $z \in [-a,a]$ gilt.
$\Rightarrow$ $|\mathrm{e}^{\mathrm{i}\,hz}-1| \le \frac{\varepsilon}{5}$ für $z \in [-a,a]$
$\Rightarrow$ Behauptung.

□

Satz 9.2 *Gilt $Y = aX + b$, so ist*

$$\varphi_Y(t) = \mathrm{e}^{\mathrm{i}\,tb}\varphi_X(ta)\,.$$

Satz 9.3 *Wenn $X_1, \dots, X_n$ unabhängige Zufallsgrößen sind, so gilt für die c. F. von $S_n = \sum\limits_{i=1}^{n} X_i$*

$$\varphi_{S_n}(t) = \prod_{i=1}^{n} \varphi_{X_i}(t)\,.$$

Satz 9.4 *Sei $\mathsf{E}|X|^k < \infty$, $k \ge 1$. Dann ist φ_X k-mal stetig differenzierbar, und es gilt*

$$\varphi_X^{(k)}(0) = \mathrm{i}^{\,k}\mathsf{E}X^k\,.$$

Beweis: Es sei $\varphi_X(t) = \int \mathrm{e}^{\mathrm{i}\,tz}\,\mathrm{d}F(z)$. Die l-te Ableitung von $\mathrm{e}^{\mathrm{i}\,tz}$ ist $\mathrm{i}^{\,l}z^l\mathrm{e}^{\mathrm{i}\,tz}, l \le k$. Wir haben $|\mathrm{i}^{\,l}z^l\mathrm{e}^{\mathrm{i}\,tz}| \le |z|^l$ und $|z|^l$ ist integrierbar, $\int |z|^l\,\mathrm{d}F(z) \le \infty$.
$\Rightarrow$ $\frac{\mathrm{d}^l}{\mathrm{d}t^l}\varphi_X(t) = \int \frac{\mathrm{d}^l}{\mathrm{d}t^l}\mathrm{e}^{\mathrm{i}\,tz}\,\mathrm{d}F(z)$.
Außerdem gilt

$$|\varphi_X^{(k)}(t+h) - \varphi_X^{(k)}(t)| = |\int z^k\,\mathrm{e}^{\mathrm{i}\,tz}(\mathrm{e}^{\mathrm{i}\,hz}-1)\,\mathrm{d}F(z)|$$

$$\le \int |z|^k|\mathrm{e}^{\mathrm{i}\,hz}-1|\,\mathrm{d}F(z)$$

$$= \int_{-\infty}^{-a} \; + \int_{a}^{\infty} \; + \int_{-a}^{a} \; = (*).$$

Sei $\varepsilon > 0$. Wir wählen $a > 0$ so, daß

$$\int_{-\infty}^{-a} |z|^k\,\mathrm{d}F \le \varepsilon, \quad \int_{a}^{\infty} |z|^k\,\mathrm{d}F \le \varepsilon$$

Es gilt $|\mathrm{e}^{\mathrm{i}\,hz} - 1| \leq 2$. Wir wählen δ_0 derart, daß

$$|\mathrm{e}^{\mathrm{i}\,\delta z} - 1| \leq \varepsilon \quad \forall \delta \leq \delta_0, z \in [-a, a].$$

Daraus folgt

$$(*) \leq 4\varepsilon + \varepsilon \int_{-a}^{a} |z|^k \,\mathrm{d}F(z).$$

Daraus folgt die Behauptung.

□

Folgerung:

$$\mathsf{E}|X|^k < \infty \quad \Rightarrow \quad \varphi_X(t) = 1 + \sum_{j=1}^{k} \frac{t^j\,\mathrm{i}^{\,j}}{j!}\mathsf{E}X^j + o(t^k) \qquad \text{für } t \to 0\,.$$

Satz 9.5 *Für alle $t \in \mathbb{R}$ gilt $\varphi_X(-t) = \overline{\varphi_X(t)}$.*

Beweis:

$$\varphi_X(-t) = \mathsf{E}\,\mathrm{e}^{-\mathrm{i}\,tX} = \overline{\mathsf{E}\,\mathrm{e}^{\mathrm{i}\,tX}} = \overline{\varphi_X(t)}\,.$$

□

Folgerung:
Falls X eine symmetrische Verteilung bezüglich 0 hat, dann ist $\varphi_X(t)$ reell, denn

$$\varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}\,.$$

Satz 9.6 *(Umkehrformel) Ist F die Verteilungsfunktion von X und ist φ_X deren c. F., so gilt für alle Stetigkeitspunkte x und y der Funktion F*

$$F(y) - F(x) = \lim_{a \to 0} \int_{-a}^{a} \frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{2\pi\,\mathrm{i}\,t} \varphi_X(t)\,\mathrm{d}t\,.$$

Beweis: Es wird der Cauchy'sche Hauptwert des Integrals berechnet.

□

Satz 9.7 *(Eindeutigkeitssatz) Die Verteilungsfunktion F ist durch ihre c. F. eindeutig bestimmt.*

Beweis: Wegen Satz 9.6 gilt

$$F(y) = \frac{1}{2\pi} \lim_{x\to-\infty} \lim_{a\to\infty} \int\limits_{-a}^{a} \frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t} \varphi_X(t)\,\mathrm{d}t$$

für alle Stetigkeitspunkte x von F. Wir bemerken, daß es höchstens abzählbar unendlich viele Unstetigkeitspunkte von F gibt und $F(x) \to 0$ für $x \to -\infty$ gilt.

□

Beispiel 9.2 $X_j \sim \mathsf{N}(\mu_j, \sigma_j^2)$, unabhängig $\Rightarrow$ Für $S = \sum\limits_{j=1}^{n} X_j$ gilt:

$$\varphi_S(t) = \prod_{j=1}^{n} \mathrm{e}^{\mathrm{i}\,\mu_j t - \frac{1}{2}\sigma_j^2 t^2} = \mathrm{e}^{\mathrm{i}\,(\sum \mu_j)t - \frac{1}{2}(\sum \sigma_j^2)t^2}.$$

Wegen der eineindeutigen Zuordnung von c. F. und Verteilungen folgt hieraus:

$$S \sim \mathsf{N}(\sum \mu_j, \sum \sigma_j^2).$$

Beispiel 9.3 X sei Poisson-verteilt: $\mathsf{P}(X = k) = \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!}$, $k = 0, 1, 2, \ldots$ für $\lambda > 0$. Dann gilt

$$\begin{aligned}
\varphi_X(t) &= \mathsf{E}\,\mathrm{e}^{\mathrm{i}\,tX} = \sum_{k=0}^{\infty} \mathrm{e}^{\mathrm{i}\,kt} \mathsf{P}(X = k) = \sum_{k=0}^{\infty} \mathrm{e}^{\mathrm{i}\,kt} \frac{\lambda^k \mathrm{e}^{-\lambda}}{k!} \\
&= \mathrm{e}^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda\,\mathrm{e}^{\mathrm{i}\,t})^k}{k!} = \mathrm{e}^{-\lambda}\,\mathrm{e}^{\lambda\,\mathrm{e}^{\mathrm{i}\,t}} = \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)}.
\end{aligned}$$

Aus Satz 9.4 berechnen wir Erwartungswert und Varianz einer Poisson-verteilten Variablen:

$$\begin{aligned}
\varphi_X^{(l)}(0) &= \mathrm{i}^{\,l}\,\mathsf{E}(X^l), \qquad \text{d.h.} \\
\mathsf{E}X &= -\mathrm{i}\,\varphi_X^{(1)}(0), \\
\mathsf{Var}\,X &= \mathsf{E}X^2 - (\mathsf{E}X)^2 = -\varphi_X^{(2)}(0) - \left(\mathrm{i}^{\,2}\left(\varphi_X^{(1)}(0)\right)^2\right) \\
&= -\varphi_X^{(2)}(0) + \left(\varphi_X^{(1)}(0)\right)^2 \\
\varphi_X^{(1)}(t) &= \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)} \cdot \lambda\,\mathrm{e}^{\mathrm{i}\,t}\,\mathrm{i} = \mathrm{i}\,\lambda\,\mathrm{e}^{\mathrm{i}\,t}\,\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)} \\
\varphi_X^{(2)}(t) &= \mathrm{i}\,\lambda\,\mathrm{e}^{\mathrm{i}\,t}\,\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)} \cdot \mathrm{i} + \mathrm{i}\,\lambda\,\mathrm{e}^{\mathrm{i}\,t}\,\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)} \cdot \lambda\,\mathrm{e}^{\mathrm{i}\,t} \cdot \mathrm{i} \\
&= -\lambda\,\mathrm{e}^{\mathrm{i}\,t}\,\mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\,t}-1)}\,[1 + \lambda\,\mathrm{e}^{\mathrm{i}\,t}]
\end{aligned}$$

$\Rightarrow \quad \varphi_X^{(1)}(0) = \mathrm{i}\,\lambda$, $\varphi_X^{(2)}(0) - \lambda^2 + (-\lambda) = -\lambda(1 + \lambda)$

$\Rightarrow \quad \mathsf{E}X = \lambda$, $\mathsf{Var}\,X = \lambda(1 + \lambda) - \lambda^2 = \lambda$

Satz 9.8 *Sei φ_X absolut integrierbar über $\mathbb{R}$, $\int |\varphi_X(t)|\,\mathrm{d}t < \infty$, F sei die Verteilungsfunktion von X. Dann ist F stetig differenzierbar, und es gilt*

$$F'(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\,tx} \varphi_X(t)\,\mathrm{d}t\,.$$

Beweis: Aus

$$\frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t} = [1 - \mathrm{i}\,tx + o(t) - 1 + \mathrm{i}\,ty + o(t)]\frac{1}{\mathrm{i}\,t} = y - x + \frac{o(t)}{\mathrm{i}\,t}$$

folgt

$$\left|\frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t}\right| \leq C \qquad \text{für } |t| \leq 1\,.$$

Außerdem gilt für $|t| \geq 1$

$$\left|\frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t}\right| \leq 2\,.$$

Da $\varphi_X(t)$ absolut integrierbar ist, d.h. $\int |\varphi_X(t)|\,\mathrm{d}t < \infty$, folgt

$$\int\limits_{-\infty}^{\infty} \left|\frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t}\right| |\varphi_X(t)|\,\mathrm{d}t < \infty\,.$$

Deshalb kann die Umkehrformel (vergleiche Satz 9.6) als uneigentliches Integral geschrieben werden:

$$F(y) - F(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \frac{\mathrm{e}^{-\mathrm{i}\,tx} - \mathrm{e}^{-\mathrm{i}\,ty}}{\mathrm{i}\,t} \varphi_X(t)\,\mathrm{d}t\,.$$

Sei x_0 fest und sei h derart, daß $x = x_0 - h$, $y = x_0 + h$ Stetigkeitspunkte von F sind. Wir formen um:

$$F(x_0 + h) - F(x_0 - h) =$$

$$\frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \frac{\cos(t(x_0 - h)) - \cos(t(x_0 + h)) - \mathrm{i}\,\sin(t(x_0 - h)) + \mathrm{i}\,\sin(t(x_0 + h))}{\mathrm{i}\,t} \cdot \varphi_X(t)\,\mathrm{d}t\,.$$

Mit

$$\left.\begin{array}{rcl} \cos(tx_0 - th) &=& \cos(tx_0)\cos(th) + \sin(tx_0)\sin(th) \\ \cos(tx_0 + th) &=& \cos(tx_0)\cos(th) - \sin(tx_0)\sin(th) \end{array}\right\} \Rightarrow (I)$$

$$\left.\begin{array}{rcl} \sin(tx_0 - th) &=& \sin(tx_0)\cos(th) - \cos(tx_0)\sin(th) \\ \sin(tx_0 + th) &=& \sin(tx_0)\cos(th) + \cos(tx_0)\sin(th) \end{array}\right\} \Rightarrow (II)$$

folgt

$$\begin{aligned}\text{Realteil} &= (I) &= 2\sin(tx_0)\sin(th)\\ \text{Imaginärteil} &= (II) &= 2\cos(tx_0)\sin(th)\end{aligned}$$

$$\begin{aligned}(I)+\mathrm{i}\,(II) &= 2\sin(th)\sin(tx_0)+2\,\mathrm{i}\,\cos(tx_0)\sin(th)\\ &= 2\sin(th)\,(\sin(tx_0)+\mathrm{i}\,\cos(tx_0))\\ &= 2\,\mathrm{i}\,\sin(th)\,(\cos(tx_0)-\mathrm{i}\,\sin(tx_0))\\ \Rightarrow \frac{(I)+\mathrm{i}\,(II)}{\mathrm{i}\,t} &= \frac{2}{t}\sin(th)\,\mathrm{e}^{-\mathrm{i}\,tx_0}\,.\end{aligned}$$

Wir erhalten schließlich

$$F(x_0+h)-F(x_0-h) \;=\; 2h\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}\,\mathrm{e}^{-\mathrm{i}\,tx_0}\varphi_X(t)\,\mathrm{d}t\,. \tag{1}$$

Wegen $|\frac{\sin th}{th}|\le 1$ haben wir

$$F(x_0+h)-F(x_0-h)\le 2h\frac{1}{2\pi}\int_{-\infty}^{\infty}|\varphi_X(t)|\,\mathrm{d}t<\infty\,,$$

d. h. wir können den Grenzübergang vollziehen und erhalten für $h\to 0$:

$$F(x_0+0)-F(x_0-0)=0\,.$$

Damit ist gezeigt, daß F absolut stetig ist. Ferner gilt

$$\frac{F(x_0+h)-F(x_0-h)}{2h} \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}\,\mathrm{e}^{-\mathrm{i}\,tx_0}\varphi_X(t)\,\mathrm{d}t\,. \tag{2}$$

Aus dem Satz von Lebesgue folgt nun die Umkehrformel für die Dichte:

$$\begin{aligned}\lim_{h\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\sin th}{th}\,\mathrm{e}^{-\mathrm{i}\,tx_0}\varphi_X(t)\,\mathrm{d}t &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{h\to 0}\frac{\sin th}{th}\,\mathrm{e}^{-\mathrm{i}\,tx_0}\varphi_X(t)\,\mathrm{d}t\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\,tx_0}\varphi_X(t)\,\mathrm{d}t=p(x_0)\,.\end{aligned}$$

Es bleibt zu zeigen, daß $p=F'$ stetig ist. Wegen

$$\begin{aligned}p(x_0+h)-p(x_0) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\mathrm{e}^{-\mathrm{i}\,tx_0-\mathrm{i}\,th}-\mathrm{e}^{-\mathrm{i}\,tx_0}\right]\varphi_X(t)\,\mathrm{d}t\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\,tx_0}\left[\mathrm{e}^{-\mathrm{i}\,th}-1\right]\varphi_X(t)\,\mathrm{d}t\end{aligned}$$

folgt

$$|p(x_0 + h) - p(x_0)| \le \varepsilon \qquad \text{für } |h| \le \delta \text{ für ein geeignetes } \delta\,.$$

□

Beispiel 9.4 Wenn c. F. $\varphi_X(t)$ integrierbar über $\mathbb{R}^1$, so existiert Dichte p und es gilt:

$$p(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\,tx} \varphi_X(t)\,\mathrm{d}t\,.$$

Normalverteilung:

$$\begin{aligned} & X \sim \mathsf{N}(0,1) \\ \Rightarrow\quad & \varphi_X(t) = \mathrm{e}^{-\frac{1}{2}t^2} \\ \Rightarrow\quad & \varphi_X \text{ ist integrierbar über } \mathbb{R}^1. \end{aligned}$$

Hier haben wir

$$\begin{aligned} \int\limits_{-\infty}^{\infty} \sin(tx)\,\mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t &= 0 \\ \int\limits_{-\infty}^{\infty} \cos(tx)\,\mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t &= \sqrt{2\pi}\,\mathrm{e}^{-\frac{1}{2}x^2} \\ \Rightarrow \quad \frac{1}{2\pi}\int\limits_{-\infty}^{\infty} \mathrm{e}^{-\frac{1}{2}x^2}\varphi_X(t)\,\mathrm{d}t &= \frac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}x^2}\,. \end{aligned}$$

9.2 Grenzwertsätze für charakteristische Funktionen

Die Zuordnung zwischen c. F. und Verteilungsfunktionen ist nicht nur eineindeutig, sondern auch stetig.

Satz 9.9 *Direkter Grenzwertsatz*
Wenn die Folge von Verteilungsfunktionen $F_1, F_2, \ldots$ im wesentlichen gegen die Verteilungsfunktion F konvergiert, so konvergiert auch die Folge der entsprechenden c. F. $\varphi_1, \varphi_2, \ldots$ gegen die c. F. φ von F. Diese Konvergenz ist in jedem endlichen Intervall gleichmäßig.

Das heißt also

$$\begin{aligned} F_n(x) &\to F(x) \quad \text{in jedem Stetigkeitspunkt von } F. \\ \Rightarrow \quad \varphi_n(t) &= \int \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F_n(x) \xrightarrow{\text{glm.}} \varphi(t) = \int \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F(x) \end{aligned}$$

Satz 9.10 *Umkehrung des direkten Grenzwertsatzes*
Wenn die Folge der c. F. $\varphi_1, \varphi_2, \dots$ gegen eine stetige Funktion φ konvergiert, so konvergiert die Folge der entsprechenden Verteilungsfunktionen $F_1, F_2, \dots$ im wesentlichen gegen eine Verteilungsfunktion F und es ist $\varphi(t) = \int \mathrm{e}^{\mathrm{i}\,tx}\,\mathrm{d}F(x)$.

Das heißt, es seien

$$F_j(x) = \frac{1}{2\pi} \lim_{z\to-\infty} \lim_{a\to\infty} \int_{-a}^{a} \frac{\mathrm{e}^{-\mathrm{i}\,tz} - \mathrm{e}^{-\mathrm{i}\,tx}}{\mathrm{i}\,t} \varphi_j(t)\,\mathrm{d}t\,, \quad j = 1, 2, \dots$$

$\varphi_j \to \varphi$ (stetig) $\Rightarrow$ $F_j \to F$ (im wesentlichen).

9.3 Charakteristische Funktionen eines zufälligen Vektors

Sei X eine p-dimensionale Zufallsgröße $\Rightarrow$ $t'X = \sum\limits_{j=1}^{p} t_j X_j$ ist eine eindimensionale Zufallsgröße.

Definition 9.2 *Als c. F. von X definiert man*

$$\begin{aligned} \varphi_X(t) = \varphi_X(t_1, \dots, t_p) = \mathsf{E}\,\mathrm{e}^{\mathrm{i}\,t'X} &= \int \cdots \int \mathrm{e}^{\mathrm{i}\sum_{j=1}^{p} t_j z_j}\,\mathrm{d}F(z_1, \dots, z_p) \\ &= \int \cdots \int \mathrm{e}^{\mathrm{i}\,t'z}\,\mathrm{d}F(z)\,. \end{aligned}$$

Die Eigenschaften sind analog zu den Eigenschaften der c. F. eindimensionaler Zufallsgrößen:

$$\varphi_X(0) = 1\,, \qquad |\varphi_X(t)| \le 1\,, \qquad \varphi_X(-t) = \overline{\varphi_X(t)}\,.$$

Aus der charakteristischen Funktion von X erhält man die c. F. jeder einzelnen Komponente X_s bzw. einer beliebigen Menge von Komponenten, zum Beispiel

$$\varphi_{X_1}(t_1) = \varphi_X(t_1, 0, \dots, 0)\,.$$

Ähnlich wie im eindimensionalen Fall ist die c. F. einer p-dimensionalen Zufallsgröße im ganzen Raum $\mathbb{R}^p$ gleichmäßig stetig, außerdem gilt für $X = (X_1, \dots, X_p)'$

$$\mathsf{E} X_1^{k_1} \dots X_p^{k_p} = \mathrm{i}^{\sum_{j=1}^{p} k_j} \cdot \left. \frac{\partial^{k_1+\cdots+k_p} \varphi_X(t_1, \dots, t_p)}{\partial t_1^{k_1} \cdots \partial t_p^{k_p}} \right|_{t_1=\dots=t_p=0}.$$

Es folgt auch direkt

Satz 9.11 *Hat X die c.F. φ_X, so ist die c. F. von $(b_1X_1 + a_1, \dots, b_pX_p + a_p)$ gleich*

$$\mathrm{e}^{\mathrm{i}\sum_{j=1}^{p} a_j t_j} \varphi_X(b_1 t_1, \dots, b_p t_p)\,.$$

Satz 9.12 *Ist φ_X die charakteristische Funktion von X, so ist die c. F. der Summe $\sum_{j=1}^{p} X_j$ gleich $\varphi_X(t,\dots,t)$ für $t \in \mathbb{R}^1$.*

(folgt unmittelbar aus der Definition)

Beispiel 9.5 Charakteristische Funktion einer Normalverteilung $\mathsf{N}(\mu, W)$.
Für $X \sim \mathsf{N}(\mu, W)$ gilt: $t'X$ ist eindimensional normalverteilt mit $\mathsf{N}(t'\mu, t'Wt)$.
$\Rightarrow$ Die c. F. von $t'X$ ist

$$\varphi_{t'X}(\tau) = \mathsf{E}\,\mathrm{e}^{\mathrm{i}\,\tau t'X} = \mathrm{e}^{\mathrm{i}\,\tau t'\mu - \frac{1}{2}\tau^2 t'Wt}\,.$$

Für $\tau = 1$ ergibt sich

$$\varphi_X(t) = \mathrm{e}^{\mathrm{i}\,t'\mu - \frac{1}{2}t'\Sigma t}$$

Wichtig ist der folgende Satz:

Satz 9.13 *Die Verteilungsfunktion ist durch ihre charakteristische Funktion eindeutig bestimmt.*

Beweis: Wie im eindimensionalen Fall über die Umkehrformel.

Satz 9.14 *(Umkehrformel) Ist $\varphi(t_1,\dots,t_p)$ die charakteristische Funktion und $F(x_1,\dots,x_p)$ die Verteilungsfunktion von $(X_1,\dots,X_p) = X$, so gilt $\forall\, a, b \in \mathbb{R}^p$*

$$\mathsf{P}(X \le b) - \mathsf{P}(X \le a) = \lim_{T\to\infty} \frac{1}{(2\pi)^p} \int_{-T}^{T} \cdots \int_{-T}^{T} \prod_{j=1}^{p} \frac{\mathrm{e}^{-\mathrm{i}\,t_j a_j} - \mathrm{e}^{-\mathrm{i}\,t_j b_j}}{\mathrm{i}\,t_j} \varphi(t_1,\dots,t_p)\,\mathrm{d}t_1 \cdots \mathrm{d}t_p\,.$$

Es gelten auch für mehrdimensionale c. F. der *direkte GWS* und die *Umkehrung des direkten GWS*.

Satz 9.15 *Die Komponenten von X sind unabhängig genau dann, wenn die charakteristische Funktion von X das Produkt der charakteristischen Funktionen der Komponenten ist, d. h.*

$$\varphi_X(t_1,\dots,t_p) = \prod_{j=1}^{p} \varphi_{X_j}(t_j)\,.$$

Beweis:

$(\Rightarrow)$ $\mathsf{E}\,\mathrm{e}^{\mathrm{i}\,t'X} = \mathsf{E}\prod_{j=1}^{p} \mathrm{e}^{\mathrm{i}\,t_j X_j} = \prod_{j=1}^{p} \mathsf{E}\,\mathrm{e}^{\mathrm{i}\,t_j X_j}.$

($\Leftarrow$) Sei $\varphi_X(t_1, ..., t_p) = \prod\limits_{j=1}^{p} \varphi_{X_j}(t_j)$

Aus der Umkehrformel folgt

$$\begin{aligned}
\mathsf{P}(X \le b) &= \lim_{a\to-\infty} \lim_{T\to\infty} \frac{1}{(2\pi)^p} \prod_{j=1}^{p} \int_{-T}^{T} \frac{\mathrm{e}^{-\mathrm{i}\,t_j a_j} - \mathrm{e}^{-\mathrm{i}\,t_j b_j}}{\mathrm{i}\,t_j} \varphi_{X_j}(t_j)\,\mathrm{d}t_j \\
&= \lim_{a\to-\infty} \prod_{j=1}^{p} [F_j(b_j) - F_j(a_j)] = \prod_{j=1}^{p} F_j(b_j)\,.
\end{aligned}$$

□

Anwendung für die Normalverteilung:

Satz 9.16 *Die Komponenten eines normalverteilten Vektors sind unabhängig genau dann, wenn sie unkorreliert sind.*

Beweis: o. B. d. A. $\mathsf{E}X_i = 0$. Für $X \sim \mathsf{N}(0, W)$ ist $W = \mathrm{diag}(\sigma_{11}, \ldots, \sigma_{nn})$ und die c. F. von X ist das Produkt der charakteristischen Funktionen der X_j.

□

10 Grenzwertsätze

Wir betrachten Folgen von reellwertigen zufälligen Größen $Y_n, n = 1, 2, \ldots$ über $(\Omega, \mathfrak{A}, \mathsf{P})$. F^{Y_n} bezeichne die Verteilungsfunktion von Y_n. Wir untersuchen das Grenzverhalten spezieller Folgen. Wir führen folgende Konvergenzarten ein.

Definition 10.1 *Y_n konvergiert in Verteilung gegen Y, falls $F^{Y_n} \to F^Y$ in allen Stetigkeitspunkten von F^Y.*
Bezeichnung: $Y_n \xrightarrow{\mathrm{d}} Y$.

Dann kann man die Aussagen der Grenzwertsätze bei charakteristischen Funktionen schreiben:
Wenn $Y_n \xrightarrow{\mathrm{d}} Y$, so $\varphi_{Y_n} \to \varphi$.
Wenn $\varphi_{Y_n} \to \varphi$ und φ_Y stetig ist, so $Y_n \xrightarrow{\mathrm{d}} Y$.

Definition 10.2 *Falls für jedes $\varepsilon > 0$ und eine Zufallsgröße Y*

$$\lim_{n\to\infty} \mathsf{P}(|Y_n - Y| \ge \varepsilon) = 0\,,$$

so konvergiert Y_n schwach gegen Y (oder auch: Y_n konvergiert stochastisch gegen Y). (Bezeichnung: $Y_n \xrightarrow{\mathsf{P}} Y$)

Definition 10.3 *Falls für eine Zufallsvariable Y*

$$\mathsf{P}\Big(\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\Big) = 1\,,$$

so konvergiert Y_n stark gegen Y (fast sicher, fast überall, mit Wahrscheinlichkeit 1).
(Bezeichnung: $Y_n \xrightarrow{\text{f.s.}} Y$ oder $Y_n \to Y \quad \mathsf{P}$-f.s.)

Dann folgt:

$$Y_n \xrightarrow{\text{f.s.}} Y \quad \Rightarrow \quad Y_n \xrightarrow{\text{P}} Y \quad \Rightarrow \quad Y_n \xrightarrow{\text{d}} Y\,.$$

Satz 10.1 *Folgende Bedingungen sind äquivalent:*
a) $\mathsf{P}(\omega : \lim Y_n(\omega) = Y(\omega)) = 1$
b) $\mathsf{P}(\omega : \sup_{n\geq N} |Y_n(\omega) - Y(\omega)| < \varepsilon) \to 1 \quad$ *für* $\quad N \to \infty$.

Beweis: Sei $B_N = B_N(\varepsilon) = \{\omega : \sup_{n\geq N} |Y_n(\omega) - Y(\omega)| < \varepsilon\}$. Dann ist $\{B_N\}$ eine aufsteigende Folge. Für $A = \{\omega : \lim Y_n(\omega) = Y(\omega)\}$ gilt $A \subseteq B_\infty = \bigcup_N B_N$.

a) $\Rightarrow \mathsf{P}(A) = 1 \Rightarrow \mathsf{P}(B_\infty = 1)$. Aus dem Satz über die aufsteigende Konvergenz folgt: $\mathsf{P}(B_N) \to 1$, d.h. b)

b) Wenn $\omega \in B_N$, so folgt $|Y_n(\omega) - Y(\omega)| < \varepsilon \quad \forall n \geq N$. Wenn a) nicht gilt, so hat $\{\omega : |Y_N(\omega) - Y(\omega)| > \tilde{\varepsilon}\}$ in der Grenze ein P-Maß > 0. Folglich gilt $\{\dots\} \nsubseteq B_N(\tilde{\varepsilon})$.
$\Rightarrow \quad \mathsf{P}(B_N) \nrightarrow 1$.

□

Derartige Konvergenzen sind von Interesse vom mathematischen und praktischen Gesichtspunkt.

10.1 Zentraler Grenzwertsatz für unabhängige, identisch verteilte Zufallsgrößen

Satz 10.2 *Es seien $Y_1, Y_2, \dots$ unabhängige und identisch verteilte Zufallsgrößen mit* $\mathsf{E}Y_i = \mu$, $\mathsf{Var}\, Y_i = \sigma^2 < \infty$. *Dann gilt für die standardisierte Summe*

$$V_n = \frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

für $S_n = \sum_{i=1}^{n} Y_i$:

$$\mathsf{P}(V_N \leq x) \to \Phi(x) \qquad \textit{gleichmäßig in } x,$$

wobei Φ die Verteilungsfunktion der $\mathsf{N}(0,1)$-*Verteilung ist.*

Beweis: Sei φ_0 die c. F. einer $\mathsf{N}(0,1)$-verteilten Variablen, d. h. $\varphi_0(t) = \mathrm{e}^{-\frac{t^2}{2}}$. Es genügt zu zeigen: $\varphi_{V_n}(t) \to \varphi_0(t)$.

Wir haben

$$\varphi_{V_n}(t) = \left[\varphi_{Y_1-\mu}\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n .$$

$\varphi_{Y_1-\mu}$ besitzt eine Taylorreihenentwicklung:

$$\varphi_{Y_1-\mu}(t) = 1 + \sum_{j=1}^{2} \frac{\mathrm{i}^j t^j}{j!}\varphi^{(j)}_{Y_1-\mu}(0) + o(t^2) = 1 - \frac{t^2\sigma^2}{2} + o(t^2)\,,$$

wobei $\varphi^{(j)}_{Y_1-\mu}(0) = \mathrm{i}^j\mathsf{E}(Y_1-\mu)^j$. Damit erhält man

$$\varphi_{V_n}(t) = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n .$$

Ferner gilt

$$\ln\left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right)^n = n\ln\left(1 - \frac{t^2}{2n} + o(\frac{t^2}{n})\right) \to -\frac{t^2}{2}\,,$$

d. h. $\varphi_{V_n}(t) \to \varphi_0(t)$.

□

Anwendung des Zentralen Grenzwertsatzes

Wenn $V_n \xrightarrow{\mathrm{d}} V$, so

$$\mathsf{P}(a < V_n \le b) \to F(b) - F(a)\,.$$

Hat V speziell eine Dichte f, so

$$\mathsf{P}(x < V_n \le x + \Delta x) \approx f(x)\Delta x\,.$$

Beispiel 10.1 Zwölfer-Regel

Es seien Y_i gleichverteilt über $[0,1]$. Dann gilt $\mathsf{E}Y_i = 0{,}5$, $\mathsf{Var}\,Y_i = 1/12$. Für $n = 12$ gilt dann $V_n = S_n - 6$ und V_n ist eine Approximation für eine standardisiert normalverteilte Variable.

Beispiel 10.2 Binomialverteilung

Es seien $Y_i \sim B(1,\vartheta)$. Dann ist nach dem Zentralen GWS die Verteilung von S_n approximativ $\mathsf{N}(n\vartheta, n\vartheta(1-\vartheta))$, d.h.

$$\mathsf{P}(S_n = k) \approx \frac{1}{\sqrt{2\pi}\sqrt{n\vartheta(1-\vartheta)}}\exp\left(-\frac{(k-n\vartheta)^2}{2n\vartheta(1-\vartheta)}\right) .$$

(Nutzung zur Beschreibung des Nagelbrettes)

10.2 Gesetze der großen Zahlen

Es sei $\{Y_n\}$ eine Folge von Zufallsgrößen über $(\Omega, \mathfrak{A}, \mathsf{P})$.

Definition 10.4 *Sei* $\bar{Y}_n = n^{-1} \sum_{i=1}^{n} Y_i$*, wobei* Y_i *unabhängige, reellwertige Zufallsgrößen sind. Wenn eine Folge* $\{a_n\}$ *reeller Zahlen existiert, so daß* $\bar{Y}_n - a_n$ *fast sicher gegen Null konvergiert, so erfüllt die Folge* $\{Y_n\}$ *das starke Gesetz der großen Zahlen, konvergiert* $Y_n - a_n$ *nur in Wahrscheinlichkeit gegen Null, so erfüllt* $\{Y_n\}$ *das schwache Gesetz der großen Zahlen.*

Satz 10.3 *(schwaches Gesetz der großen Zahlen)*
Es sei Y_i *Zufallsvariable mit* $\mathsf{E}Y_i = \mu_i$*,* $\mathsf{Var}\, Y_i = \sigma_i^2$*,* $\mathsf{Cov}\,(Y_i, Y_j) = 0$ *für* $i \neq j$ *mit* $\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 = 0$*. Dann gilt*

$$\bar{Y}_n - \bar{\mu}_n \xrightarrow{\mathsf{P}} 0\,.$$

Beweis: Tschebyscheff $\Rightarrow$

$$\mathsf{P}(|\bar{Y}_n - \bar{\mu}_n| \geq \underbrace{\varepsilon}_{\lambda\tilde{\sigma}_n}) \leq \frac{1}{\varepsilon^2}\tilde{\sigma}_n^2 \qquad \text{falls } \tilde{\sigma}_n^2 = \mathsf{Var}\,\bar{Y}_n,$$

$$\varepsilon = \lambda\tilde{\sigma}_n \Rightarrow \lambda = \frac{\varepsilon}{\tilde{\sigma}_n}$$

$$\tilde{\sigma}_n^2 = \mathsf{Var}\,\bar{Y}_n = \frac{1}{n^2}\sum_{i=1}^{n} \mathsf{Var}\,Y_i = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$$

$$\Rightarrow \qquad \mathsf{P}(|\bar{Y}_n - \bar{\mu}_n| \geq \varepsilon) \leq \frac{1}{n^2\varepsilon^2}\sum_{i=1}^{n}\sigma_i^2 \to 0\,.$$

□

Folgerung: Seien Y_i unabhängig, identisch verteilt mit $\mathsf{Var}\,Y_i = \sigma^2$. Dann konvergiert $\bar{Y}_n$ schwach gegen μ. Aus dem folgenden 1. Kolmogorowschen Satz folgt sogar die starke Konvergenz.

Satz 10.4 (1. Kolmogorowscher Satz) *Seien* $Y_1, Y_2, \ldots$ *unabhängige Zufallsgrößen mit* $\mathsf{E}Y_i = \mu_i$ *und* $\mathsf{Var}\,Y_i = \sigma_i^2 < \infty$*.*
Wenn

$$\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2} < \infty\,,$$

dann gilt für $\bar{Y}_n$ das starke Gesetz der großen Zahlen, d.h.

$$\bar{Y}_n - \bar{\mu}_n \xrightarrow{f.s.} 0\,, \qquad \textit{wobei} \quad \bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i$$

Beweis: Wesentliches Hilfsmittel ist die Ungleichung von Hajek-Rényi (1955): Seien $Y_1, Y_2, \ldots$ unabhängige Zufallsgrößen mit $\mathsf{E}Y_i = 0$, $\mathsf{Var}\,Y_i = \sigma_i^2 < \infty$. Wenn $c_1, c_2, \ldots$ eine positive, nicht wachsende Folge reeller Zahlen ist, so gilt für beliebige ganze Zahlen m, n $(m < n)$ und beliebige $\varepsilon > 0$:

$$\mathsf{P}\left(\max_{m\le k\le n} c_k|Y_1 + \cdots + Y_k| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}\left(c_m^2\sum_{i=1}^{m}\sigma_i^2 + \sum_{i=m+1}^{n} c_i^2\sigma_i^2\right)$$

(Gnedenko: Lehrbuch der Wahrscheinlichkeitsrechnung).

Die Anwendung dieser Ungleichung für $Y_i = Y_i - \mu_i$, $c_i = \frac{1}{i}$ liefert

$$\mathsf{P}\left(\max_{m\le k\le n} |\bar{Y}_k| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{i=1}^{m}\sigma_i^2 + \sum_{i=m+1}^{n} \frac{1}{i^2}\sigma_i^2\right).$$

Es folgt

$$\mathsf{P}\left(\sup_{m\le k} |\bar{Y}_k| \ge \varepsilon\right) \le \frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{i=1}^{m}\sigma_i^2 + \sum_{i=m+1}^{\infty} \frac{1}{i^2}\sigma_i^2\right) \underset{m\to\infty}{\longrightarrow} 0\,,$$

das heißt die starke Konvergenz der Folge $\bar{Y}_n - \bar{\mu}_n$ gegen Null ist gezeigt.

$\Rightarrow \quad \bar{Y}_n - \bar{\mu}_n \xrightarrow{f.s.} 0.$

□

Folgerung: Wenn $Y_1, Y_2, \ldots$ unabhängig mit $\mathsf{E}Y_i = \mu_i$, $\mathsf{Var}\,Y_i \le c < \infty$, so genügt $\bar{Y}_n$ einem starken Gesetz der großen Zahlen.

Satz 10.5 (2. Kolmogorowscher Satz) *Seien $Y_1, Y_2, \ldots$ unabhängige, identisch verteilte Zufallsgrößen.*
Dann gilt:
$\bar{Y}_n$ konvergiert mit Wahrscheinlichkeit 1 gegen μ genau dann, wenn $\mathsf{E}Y_i$ existiert und $\mathsf{E}Y_i = \mu$ gilt.

ohne Beweis

Beispiel 10.3 Ein vorteilhaftes Spiel, bei dem man auf Dauer verliert.
Ein Spiel, bei dem in jeder Runde der Erwartungswert des Gewinns gleich dem des Verlustes ist, heißt fair, eins, bei dem der erwartete Gewinn jeweils größer ist als der erwartete Verlust, heißt vorteilhaft. Das folgende Spiel ist vorteilhaft: Man beginnt mit

dem Startkapital $Y_0 = 1$. In jeder Runde wird eine Münze geworfen. Das Kapital Y_n nach der n-ten Runde sei $Y_{n-1}/2$, wenn der n-te Wurf Kopf ist, und $5Y_{n-1}/3$, wenn im n-ten Wurf Zahl kommt. (Man verliert also $Y_{n-1}/2$ mit Wahrscheinlichkeit $1/2$, und man gewinnt $2Y_{n-1}/3$ mit Wahrscheinlichkeit $1/2$, so daß das Spiel vorteilhaft ist.)

Sei $Y_n = 1/2$, wenn im n-ten Wurf Kopf kommt, und $Y_n = 5/3$, wenn im n-ten Wurf Zahl kommt. Dann ist $Y_n = Y_1 \cdot Y_2 \cdots Y_n$. Die Y_i sind unabhängig und identisch verteilt. Aus

$$\mathsf{E}(Y_i) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{5}{3} = \frac{13}{12} > 1$$

folgt

$$\mathsf{E}Y_n = (\mathsf{E}Y_1)(\mathsf{E}Y_2) \cdots (\mathsf{E}Y_n) = \left(\frac{13}{12}\right)^n \to \infty\,.$$

Sei $\mu = \mathsf{E}(\log Y_i)$. Dann ist $\mu = \frac{\log\frac{1}{2} + \log\frac{5}{3}}{2} < \frac{\log\frac{1}{2} + \log 2}{2} = 0$. Für beliebige $\varepsilon > 0$ gilt nach dem starken Gesetz der großen Zahlen

$$\mathsf{P}\left(\sup_{n \geq N} \left|\frac{1}{n}(\log Y_i + \cdots + \log Y_n) - \mu\right| \leq \varepsilon\right) \to 1 \quad \text{für } N \to \infty\,.$$

Speziell setzen wir $\varepsilon = -\frac{\mu}{2}$. Es folgt $\mathsf{P}(\frac{1}{n}\log Y_n - \mu \leq -\frac{\mu}{2}) \to 1$. Für großes n ist also mit Wahrscheinlichkeit nahe bei 1 $Y_n \leq \mathrm{e}^{\frac{\mu n}{2}}$. Wegen $\mu < 0$ strebt $\mathrm{e}^{\frac{\mu n}{2}}$ (exponentiell schnell) gegen 0. Auf lange Sicht strebt der Kapitalstand ziemlich schnell gegen 0. ($\mathsf{E}Y_n$ ist trotz der kleinen Gewinnwahrscheinlichkeit groß, weil man mit kleiner Wahrscheinlichkeit enorm große Gewinne machen kann.) Hierbei haben wir $\mathsf{E}Y_n \to \infty$, $\mathsf{Var}\, Y_n \to \infty$ und trotzdem $Y_n \xrightarrow{\mathsf{P}} 0$.

□

Beispiel 10.4 Das arithmetische Mittel unabhängiger, identisch verteilter Zufallsgrößen strebt gegen den Erwartungswert:

$$\mathsf{P}(|\bar{Y}_n - \mu| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2$$

Wir betrachten speziell einen Würfel. Wir setzen

$$Y_i = \begin{cases} 1 & \text{für Würfeln einer 6} \\ 0 & \text{für Würfeln einer } 1, \ldots, 5\,. \end{cases}$$

$\bar{Y}_n$ = relative Häufigkeit der Erfolge.

$$\begin{aligned} \Rightarrow \qquad & \mathsf{P}(|\bar{Y}_n - \frac{1}{6}| \geq \varepsilon) \leq \frac{1}{n\varepsilon^2}\sigma^2 \\ & \mathsf{Var}\, Y_i = \vartheta(1 - \vartheta) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36} \leq \frac{1}{6} \\ \Rightarrow \qquad & \mathsf{P}(|\bar{Y}_n - \frac{1}{6}| \geq \varepsilon) \leq \frac{1}{6n\varepsilon^2} = \text{klein}\,. \end{aligned}$$

„Schwache Konvergenz“ besagt, daß Y_n für nur wenige ω weit weg von Y liegt. Es besagt aber nicht, daß $Y_n(\omega) \to Y(\omega)$ für ein ω gilt.

Wir hatten die starke Konvergenz (fast sichere Konvergenz) wie folgt definiert:

$$\mathsf{P}(\lim_n Y_n(\omega) = Y(\omega)) = 1\,.$$

Satz 10.6 *Aus der starken Konvergenz folgt die schwache Konvergenz.*

BEWEIS: Sei $\varepsilon > 0$. Definiere $B_N = \{\omega : \omega \in \Omega, |Y_n(\omega) - Y(\omega)| < \varepsilon$ für alle $n \geq N\}$.
Das ist eine aufsteigende Folge $B_N \subseteq B_{N+1} \subseteq \ldots$
Sei $A = \{\omega : \lim Y_n(\omega) = Y(\omega)\}$, $B_\infty = \bigcup_N B_N \Rightarrow \quad A \subseteq B_\infty$.
Wegen $\mathsf{P}(A) = 1$ folgt $\mathsf{P}(B_\infty) = 1$.
Aus dem Satz über aufsteigende Konvergenz: $\mathsf{P}(B_N) \to 1$.
Wegen

$$\mathsf{P}(\underbrace{|Y_N - Y| \geq \varepsilon}_{\subseteq \bar{B}_N}) \leq \mathsf{P}(\bar{B}_N)$$

und $\mathsf{P}(\bar{B}_N) \to 0$ folgt die Behauptung.

Folgerung: Falls $Y_n \to Y_n$ fast sicher, so

$$\mathsf{P}\left(\sup_{n \geq N} |Y_n(\omega) - Y(\omega)| < \varepsilon\right) \to 1$$

$$\text{bzw.} \qquad \mathsf{P}\left(\sup_{n \geq N} |Y_n(\omega) - Y(\omega)| \geq \varepsilon\right) \to 0 \qquad \text{für } N \to \infty\,.$$

□

Andererseits folgt aus $\sup_{n \geq N} |Y_n(\omega) - Y(\omega)| < \varepsilon$ für großes N

$$\lim Y_n(\omega) = Y(\omega)\,.$$

$\Rightarrow$ Die folgenden Bedingungen sind äquivalent:

$$\mathsf{P}(\lim Y_n = Y) = 1$$

$$\mathsf{P}\left(\sup_{n \geq N} |Y_n - Y| < \varepsilon\right) \to 1$$

$$\mathsf{P}\left(\sup_{n \geq N} |Y_n - Y| \geq \varepsilon\right) \to 0\,.$$

Ein wichtiges Ergebnis ist das folgende Lemma.

Lemma 3 *(Borel-Cantelli)*
Für eine Folge $A_1, A_2, \ldots$ von Ereignissen ($A_i \in \mathfrak{A}$) sei

$$A^* = \{\omega \in \Omega : \omega \in A_k \text{ für unendlich viele } k\}\,.$$

1. *Gilt* $\sum\limits_{k=1}^{\infty} \mathsf{P}(A_k) < \infty$, *so ist* $\mathsf{P}(A^*) = 0$.

2. *Sind die* A_k *unabhängig und ist* $\sum\limits_{k=1}^{\infty} \mathsf{P}(A_k) = \infty$, *so ist* $\mathsf{P}(A^*) = 1$.

Beweis:

1. $\omega \in \Omega$ gehört zu A^*, wenn für jedes n ein $k = k(\omega) \geq n$ mit $\omega \in A_k$ existiert.
   $\Rightarrow \quad A^* = \bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k \geq n} A_k$.
   n beliebig $\Rightarrow \mathsf{P}(A^*) \leq \mathsf{P}\left(\bigcup\limits_{k=n}^{\infty} A_k\right) \leq \sum\limits_{k=n}^{\infty} \mathsf{P}(A_k) \to 0$ bei $\sum\limits_{k=1}^{\infty} \mathsf{P}(A_k) < \infty$.

2. A_k unabhängig $\Rightarrow \bar{A}_k$ unabhängig.

$$\mathsf{P}\left(\bigcap_{k=n}^{N} \bar{A}_k\right) = \prod_{k=n}^{N} \mathsf{P}(\bar{A}_k) = \prod_{k=n}^{N} (1 - \mathsf{P}(A_k))$$

   Für $0 \leq \alpha_i \leq 1$ gilt $\ln(1-\alpha_i) \leq -\alpha_i \qquad \Rightarrow$

$$\begin{array}{rrcl} & \ln\left(\prod\limits_{k=n}^{N}(1-\alpha_k)\right) & \leq & -\sum\limits_{k=n}^{N} \alpha_k \\ & \prod\limits_{k=n}^{N}(1-\alpha_k) & \leq & \mathrm{e}^{-\sum_{k=n}^{N} \alpha_k} \\ \Rightarrow & \mathsf{P}\left(\bigcap\limits_{k=n}^{N} \bar{A}_k\right) & \leq & \mathrm{e}^{-\sum_{k=n}^{N} \mathsf{P}(A_k)}\,. \end{array}$$

   Betrachte $N \to \infty$:

$$\mathrm{e}^{-\sum_{k=n}^{N} \mathsf{P}(A_k)} \to 0 \qquad \Rightarrow \qquad \mathsf{P}\left(\bigcap_{k=n}^{\infty} \bar{A}_k\right) = 0\,.$$

   Es gilt $\overline{A^*} = \bigcup\limits_{n=1}^{\infty} \underbrace{\bigcap\limits_{k=n}^{\infty} \bar{A}_k}_{\text{wachsend}}$.

$$\begin{array}{rrcll} \Rightarrow & \mathsf{P}(\overline{A^*}) & = & 0 & \text{wegen } \sigma\text{-Additivität von } \mathsf{P}. \\ \Rightarrow & \mathsf{P}(A^*) & = & 1\,. & \end{array}$$

$\square$

Satz 10.7 *(starkes Gesetz der großen Zahlen)*
Sei $Y_1, Y_2, \ldots$ Folge reellwertiger unkorrelierter Zufallsvariablen und $\mathsf{Var}\, Y_i \leq M < \infty \quad \forall i$. Dann konvergiert die Folge

$$Z_n = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathsf{E} Y_i)$$

fast sicher gegen 0.

BEWEIS:

1. Wir zeigen, daß Z_{n^2} fast sicher gegen 0 konvergiert. Sei $Y_i := Y_i - \mathsf{E}Y_i$, so daß $Z_n = \frac{1}{n} \sum\limits_1^n Y_i$ mit $\mathsf{E}Y_i = 0$.

$$\Rightarrow \qquad \mathsf{Var}\, Z_{n^2} = \mathsf{Var}\, \left(\frac{1}{n^2} (Y_1 + \cdots + Y_{n^2}) \right) = \frac{1}{n^4} \sum_1^{n^2} \mathsf{Var}\, Y_i \leq \frac{M}{n^2}.$$

Tschebyscheff $\Rightarrow$ für beliebige $\varepsilon > 0$: $\quad \mathsf{P}(|Z_{n^2}| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \mathsf{Var}\, Z_{n^2} \leq \frac{M}{n^2 \varepsilon^2}$.

Setzen $A_{n,\varepsilon} = \{\omega : |Z_{n^2}| \geq \varepsilon\}$. $\Rightarrow \sum\limits_{n=1}^{\infty} \mathsf{P}(A_{n,\varepsilon}) < \infty$.
Aus Borel-Cantelli folgt: Fast jedes ω gehört nur zu endlich vielen $A_{n,\varepsilon}$.
Sei $\mathsf{E}_k \;=\; \{\omega : |Z_{n^2}| \geq \frac{1}{k}$ für unendlich viele $n\}$
$\qquad\;\; = \; \{\omega : \omega \in A_{n,\frac{1}{k}}$ für unendlich viele $n\}$

$$\Rightarrow \qquad \mathsf{P}(\mathsf{E}_k) = 0 \quad \forall k \qquad \Rightarrow \qquad \mathsf{P}\left(\bigcup_k \mathsf{E}_k \right) = 0.$$

$$\Rightarrow \qquad \left(\omega : \sup_{n \geq N} |Z_{n^2}| \geq \varepsilon \right) \subseteq \bigcup_{n \geq N} \{\omega : |Z_{n^2}(\omega)| \geq \varepsilon\}.$$

Sei k so groß, daß $\varepsilon \geq \frac{1}{k}$.

$$\Rightarrow \qquad \bigcup_{n \geq N} \{\omega : |Z_{n^2}(\omega)| \geq \varepsilon\} \subseteq \bigcup_{n \geq N} \left\{ \omega : |Z_{n^2}(\omega)| \geq \frac{1}{k} \right\}$$

$$\Rightarrow \qquad \mathsf{P}\left(\omega : \sup_{n \geq N} |Z_{n^2}(\omega)| \geq \varepsilon \right) \leq \sum_{n=N}^{\infty} \mathsf{P}\left(\omega : |Z_{n^2}(\omega)| \geq \frac{1}{k} \right) \to 0$$

$\Rightarrow \qquad Z_{n^2}$ konvergiert fast sicher gegen 0.

2. Betrachte $m \in \mathbb{N}$, sei $n = n(m)$ diejenige natürliche Zahl mit $n^2 \leq m < (n+1)^2$. Wir vergleichen Z_{n^2} mit Z_m. Sei $S_k = \sum_{i=1}^{k} Y_i$.

$$\Rightarrow \qquad \mathsf{Var}\,(S_m - S_{n^2}) = \sum_{i=n^2+1}^{m} \mathsf{Var}\,Y_i \leq M(m-n^2)$$

Tschebyscheff: $\mathsf{P}(|S_m - S_{n^2}| \geq \varepsilon n^2) \leq \frac{M}{\varepsilon^2 n^4}(m-n^2)$
Jetzt summieren wir über m:

$$\begin{aligned}
\sum_{m=1}^{\infty} \mathsf{P}\left(\frac{1}{n(m)^2}|S_m - S_{n^2}| \geq \varepsilon\right) &\leq \frac{M}{\varepsilon^2}\sum_{m=1}^{\infty}\frac{m-n(m)^2}{n(m)^4} \\
&= \frac{M}{\varepsilon^2}\sum_{n=1}^{\infty}\sum_{m=n^2}^{(n+1)^2-1}\frac{m-n^2}{n^4} \\
&= \frac{M}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{n^4}(1+2+\cdots+2n) \\
&= \frac{M}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{2n\cdot(2n+1)}{2}\frac{1}{n^4} < \infty\,.
\end{aligned}$$

Nach Borel-Cantelli: Für $B_m = \{\omega : \frac{1}{n(m)^2}|S_m - S_{n(m)^2}| \geq \varepsilon\}$ und $B^* = \{\omega : \omega \in B_m$ für unendlich viele $m\}$ gilt $\mathsf{P}(B^*) = 0$.
$\Rightarrow$ Fast alle ω liegen nicht in B^*.
$\Rightarrow$ Für fast alle ω existiert $m_\varepsilon(\omega)$, so daß für alle $m \geq m_\varepsilon(\omega)$ gilt

$$\frac{1}{n(m)^2}|S_m(\omega) - S_{n(m)^2}(\omega)| < \varepsilon\,.$$

In 1. wurde gezeigt

$$\sum_{n=1}^{\infty}\mathsf{P}(A_{n,\varepsilon}) < \infty \qquad \text{für } A_{n,\varepsilon} = \{\omega : |Z_{n^2}(\omega)| \geq \varepsilon\}\,.$$

Borel-Cantelli $\Rightarrow$
Mit $A_\varepsilon^* = \{\omega : \omega \in A_{n,\varepsilon}$ für unendlich viele $n\}$ gilt $\mathsf{P}(A_\varepsilon^*) = 0$. $\Rightarrow$ Für fast alle ω für hinreichend großes m gilt

$$\left|\underbrace{\frac{1}{n(m)^2}S_{n(m)^2}(\omega)}_{Z_{n(m)^2}(\omega)}\right| < \varepsilon\,.$$

$\Rightarrow$ Beides gilt für fast alle ω, d. h.

$$\begin{aligned}\left|\frac{1}{n(m)^2} S_{n(m)^2}(\omega)\right| &< \varepsilon\,,\\ \frac{1}{n(m)^2}\left|S_m(\omega) - S_{n(m)^2}(\omega)\right| &< \varepsilon\,.\end{aligned}$$

$\Rightarrow \quad \dfrac{1}{n(m)^2}|S_m(\omega)| < 2\varepsilon\,.$ Wir haben $m \geq n(m)^2$.

$\Rightarrow \quad Z_m(\omega) = \dfrac{1}{m}|S_m(\omega)| < 2\varepsilon$ für fast alle ω, m hinreichend groß.

$\Rightarrow$ Für fast alle ω gilt $\lim Z_m(\omega) = 0\,.$

□

Beispiel 10.5 Seien $Y_1, Y_2, \ldots$ eine unendliche Folge von Bernoulli-Versuchen $Y_i \sim B(1, \vartheta)$.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i \to \vartheta \qquad \text{fast sicher}\,.$$

Nehmen wir z. B. beim Würfelwurf $Y_i = \begin{cases} 1, & \text{falls Würfel} = 6\\ 0, & \text{sonst.}\end{cases}$

$H_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ ist relative Häufigkeit für eine 6 bei n Würfeln. Aus dem schwachen Gesetz der großen Zahlen: $\mathsf{P}(\omega\,:\, H_n(\omega) - \frac{1}{6} > \varepsilon) \to 0$. Wenn nur das bekannt wäre, so könnte es vorkommen, daß für jedes ω ein hinreichend großes n mit $H_n(\omega) \approx 1$ ist. Aus dem starken Gesetz der großen Zahlen gilt aber $H_n(\omega) \to \frac{1}{6}$ für fast alle ω. Das heißt, lange Sechserfolgen treten höchstens dann auf, wenn $H_n(\omega) \approx \frac{1}{6}$ schon ist.

(Noch ein Beispiel für das schwache Gesetz der großen Zahlen: Krengel S. 60.)

10.3 Approximation der Binomialverteilung

$Y_1, Y_2, \ldots$ unabhängig und identisch verteilt, $Y_i \sim B(1, \vartheta)$

$\mathsf{E}Y_i = \vartheta$, $\mathsf{Var}\, Y_i = \vartheta(1-\vartheta)$

$$S_n = \sum_{i=1}^{n} Y_i$$

$\Rightarrow \quad \mathsf{P}(S_n = k) = \dbinom{n}{k}\vartheta^k(1-\vartheta)^{n-k} = b_{n,\vartheta}(k)$ Werte der Binomialverteilung $B(n, \vartheta)$.

Wie wahrscheinlich ist es nun, bei 80 Würfen einer Münze $k = 40$ mal Kopf zu haben?

$$\binom{80}{40} 2^{-80}$$

Noch nicht einmal die Größenordnung ist bekannt.

Für große n sind k von Interesse, für die $\frac{k}{n}$ ungefähr ϑ ist. Wir betrachten Folgen k_n mit

$$\frac{k_n}{n} \to \vartheta \in (0,1)$$

$$\Rightarrow \quad k_n \to \infty, \quad n - k_n \to \infty.$$

Wir benutzen die Stirlingsche Formel:

$$n! \sim \eta_n := \sqrt{2\pi n} \left(\frac{n}{\mathrm{e}}\right)^n$$

Hier bedeutet $a_n \sim b_n$, falls $\frac{a_n}{b_n} \to 1$.

Dann gilt

$$\begin{aligned}
b_{n,\vartheta}(k_n) &= \binom{n}{k_n} \vartheta^{k_n} (\underbrace{1-\vartheta}_{q})^{n-k_n} \\
&= \frac{n!}{k_n!(n-k_n)!} \vartheta^{k_n} q^{n-k_n} \\
&= \frac{\sqrt{2\pi}\, n^{n+\frac{1}{2}}\, \mathrm{e}^{-n}}{2\pi k_n^{k_n+\frac{1}{2}}\, \mathrm{e}^{-k_n} (n-k_n)^{n-k_n+\frac{1}{2}}\, \mathrm{e}^{-(n-k_n)}} \vartheta^{k_n} q^{n-k_n} \\
&= \frac{1}{\sqrt{2\pi}} \sqrt{\frac{n}{k_n(n-k_n)}} \left(\frac{n\vartheta}{k_n}\right)^{k_n} \left(\frac{nq}{n-k_n}\right)^{n-k_n}
\end{aligned}$$

$$k_n \sim n\vartheta, \quad n - k_n \sim nq \quad \Rightarrow \quad \frac{n}{k_n(n-k_n)} \sim \frac{n}{n^2\vartheta q} = \frac{1}{n\vartheta q}$$

$$\Rightarrow \quad b_{n,\vartheta}(k_n) \sim \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{n\vartheta q}} \underbrace{\left(\frac{n\vartheta}{k_n}\right)^{k_n} \left(\frac{nq}{n-k_n}\right)^{n-k_n}}_{=:\chi_{n,k_n}}$$

Es gilt

$$\begin{aligned}
\mathsf{P}(S_n = k_n) = b_{n,\vartheta}(k_n) &= \binom{n}{k_n} \vartheta^{k_n} (1-\vartheta)^{n-k_n} \\
&= \binom{n}{k_n} \vartheta^{k} q^{n-k_n}
\end{aligned}$$

Lemma 4 *Für* $v(n,k_n) = \frac{k_n - n\vartheta}{\sqrt{n\vartheta q}}$ *und* $n\psi(t_n - \vartheta) \to 0$ *gilt*

$$\lim \frac{\chi_{n,k_n}}{\mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}} = 1\,, \quad d.\,h.$$

$$\chi_{n,k_n} \sim \mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}\,.$$

Beweis: Wir setzen $t_n = \frac{k_n}{n} \quad \Rightarrow \quad t_n \to \vartheta$. Dann folgt

$$\begin{aligned}
\ln \chi_{n,k_n} &= k_n \ln\left(\frac{n\vartheta}{k_n}\right) + (n-k_n)\ln\left(\frac{nq}{n-k_n}\right) \\
-\ln \chi_{n,k_n} &= k_n \ln\left(\frac{k_n}{n\vartheta}\right) + (n-k_n)\ln\left(\frac{n-k_n}{nq}\right) \\
&= n\left[\underbrace{t_n \ln\left(\frac{t_n}{\vartheta}\right) + (1-t_n)\ln\left(\frac{1-t_n}{q}\right)}_{g(t_n)}\right]
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \qquad g(\vartheta) &= \vartheta\cdot 0 + (1-\vartheta)\cdot 0 = 0\,. \\
g'(t) &= \ln\left(\frac{t}{\vartheta}\right) + t\cdot\frac{\vartheta}{t}\cdot\frac{1}{\vartheta} - \ln\left(\frac{1-t}{q}\right) + (1-t)\cdot\frac{q}{1-t}\cdot\frac{-1}{q} \\
&= \ln\frac{t}{\vartheta} + 1 - \ln\left(\frac{1-t}{q}\right) - 1 = \ln\frac{t}{\vartheta} - \ln\frac{1-t}{q} \\
g'(\vartheta) &= 0 \\
g''(t) &= \frac{\vartheta}{t}\cdot\frac{1}{\vartheta} - \frac{q}{1-t}\cdot\frac{-1}{q} = \frac{1}{t} + \frac{1}{1-t} = \frac{1}{t(1-t)} \\
g''(\vartheta) &= \frac{1}{\vartheta q}
\end{aligned}$$

$\Rightarrow$ Taylorentwicklung $\quad g(t) = (t-\vartheta)^2\cdot\frac{1}{2\vartheta q} + \psi(t-\vartheta)$ mit $|\psi(t-\vartheta)| \le c\,|t-\vartheta|^3$

Wir setzen $v(n,k_n) = \frac{k_n - n\vartheta}{\sigma_n}$ mit $\sigma_n^2 = \frac{k_n(n-k_n)}{n} \sim n\vartheta q$.

Wir haben $n\cdot g(t_n) \;=\; n(t_n-\vartheta)^2\cdot\frac{1}{2\vartheta q} + n\psi(t_n-\vartheta)$

$\qquad\qquad\qquad\sim\; \frac{1}{2}v^2(n,k_n) + n\psi(t_n-\vartheta)$ und deshalb

$$\frac{n(t_n-\vartheta)^2}{2\vartheta q} = \frac{n(\frac{k_n}{n}-\vartheta)^2}{2\vartheta q} = \frac{(k_n-n\vartheta)^2}{2n\vartheta q} \sim \frac{1}{2}\,v^2(n,k_n) \tag{3}$$

$$\Rightarrow \qquad -\ln\chi_{n,k_n} - \frac{1}{2}\,v^2(n,k_n) \sim n\psi(t_n-\vartheta)$$

$$\Rightarrow \qquad \ln\chi_{n,k_n} + \frac{1}{2}v^2(n,k_n) \sim -n\psi(t_n-\vartheta) = \ln\left[\frac{\chi_{n,k_n}}{\mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}}\right]$$

$$\Rightarrow \qquad -n|\psi(t_n-\vartheta)| \le \ln\left[\frac{\chi_{n,k_n}}{\mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}}\right] \le n|\psi(t_n-\vartheta)|$$

$$\Rightarrow \qquad \mathrm{e}^{-n|\psi(t_n-\vartheta)|} \le \frac{\chi_{n,k_n}}{\mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}} \le \mathrm{e}^{n|\psi(t_n-\vartheta)|} \qquad (*)\,.$$

$\square$

Folgerung: Unmittelbar aus $(*)$ erhält man:

Sind α_n, β_n zwei Folgen mit

$$\alpha_n \leq \beta_n \text{ und } n\psi\left(\frac{\alpha_n}{n} - \vartheta\right) \to 0\,,\; n\psi\left(\frac{\beta_n}{n} - \vartheta\right) \to 0\,, \tag{4}$$

so folgt wegen $|\psi(t-\vartheta)| \leq c\,|t-\vartheta|^3$:

$$n\psi(\tau_n - \vartheta) \to 0 \qquad \text{gleichmäßig für} \quad \frac{\alpha_n}{n} \leq \tau_n \leq \frac{\beta_n}{n}\,.$$

Lemma 5 *Gilt* $\frac{1}{\sqrt{n}} v^3(n,k_n) \to 0$*, so gilt*

$$\lim \frac{\chi_{n,k_n}}{\mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}} = 1\,,\; \textit{d. h.}$$
$$\chi_{n,k_n} \sim \mathrm{e}^{-\frac{1}{2}v^2(n,k_n)}\,.$$

Beweis: Aus $\frac{1}{\sqrt{n}} v^3(n,k_n) \to 0$, folgt $n\psi(\frac{k_n}{n} - \vartheta) \to 0$.

Denn aus (3):

$$\begin{aligned}
\frac{n^3(t_n-\vartheta)^6}{(2\vartheta q)^3} \quad &\sim \quad \frac{1}{2^3}\, v^6(n,k_n) \\
\frac{1}{n} \cdot \frac{n^3(t_n-\vartheta)^6}{(2\vartheta q)^3} \quad &\sim \quad \frac{1}{2^3} \cdot \frac{1}{n}\, v^6(n,k_n) \\
= \frac{n^2(t_n-\vartheta)^6}{(2\vartheta q)^3} \quad & \quad\;\Rightarrow \quad \text{mit Lemma 4 Behauptung.}
\end{aligned}$$

Folgerung: Falls $\alpha_n \leq \beta_n$ und $\frac{1}{\sqrt{n}} v^3(n,\alpha_n) \to 0$, $\frac{1}{\sqrt{n}} v^3(n,\beta_n) \to 0$, so

$$n\psi(\tau_n - \vartheta) \to 0 \qquad \text{gleichmäßig für} \quad \frac{\alpha_n}{n} \leq \tau_n \leq \frac{\beta_n}{n}\,.$$

$\square$

Dann gilt der folgende Satz:

Satz 10.8 *(lokaler Grenzwertsatz für die Binomialverteilung)*
Ist $0 < \vartheta < 1$ *und* k_n *eine Folge mit* $\frac{1}{\sqrt{n}}\, v^3(n,k_n) \to 0$*, so gilt*

$$b_{n,\vartheta}(k_n) \sim \frac{1}{\sigma_n}\, \varphi(v(n,k_n))$$

für $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, \mathrm{e}^{-\frac{x^2}{2}}$.

Sind α_n, β_n *Folgen wie unter (4), so konvergiert* $\frac{b_{n,\vartheta}(k_n)}{\frac{1}{\sigma_n}\varphi(v(n,k_n))}$ *gleichmäßig gegen 1 für alle* k_n *mit* $\alpha_n \leq k_n \leq \beta_n$.

Die Binomialverteilung stellt man üblicherweise durch ein Histogramm dar. Dann hat das Rechteck in $(k_n - \frac{1}{2}, k_n + \frac{1}{2})$ mit der Höhe $b_{n,\vartheta}(k_n)$ den Flächeninhalt $b_{n,\vartheta}(k_n)$. $\sum\limits_{k=0}^{n} b_{n,\vartheta}(k) = 1$ entspricht dann dem Integral über das Histogramm.

10.4 Satz von Moivre-Laplace

Wir erinnern uns noch einmal. Für

$$S_n = \sum_{i=1}^{n} Y_i \text{ gilt } \mathsf{E} S_n = n\vartheta\,, \quad \mathsf{Var}\, S_n = n\vartheta q = n\vartheta(1-\vartheta)\,.$$

Nach den allgemeinen Rechenregeln gilt für $S_n^* = \frac{1}{\sqrt{n\vartheta q}}(S_n - n\vartheta)$:

$$\begin{array}{lll} \mathsf{E} S_n^* = 0\,, & \mathsf{Var}\, S_n^* = 1 & \text{und} \\ S_n = k_n & \Leftrightarrow \quad S_n^* = v(n, k_n) & \\ & \quad \uparrow & \\ & \quad \text{standardisierte Form von } S_n. & \end{array}$$

Üblicherweise bezeichnet man mit Φ die Verteilungsfunktion der standardisierten Normalverteilung

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}\xi^2}\,\mathrm{d}\xi = \int_{-\infty}^{x} \varphi(\xi)\,\mathrm{d}\xi\,.$$

Satz 10.9 *(Moivre-Laplace)*
Sei $0 < \vartheta < 1$ *und* $S_n \sim B(n, \vartheta)$. *Dann gilt für alle* $a < b$

$$\lim_{n\to\infty} \mathsf{P}(a \le S_n^* \le b) = \Phi(b) - \Phi(a)\,.$$

Heuristisch ist das klar, denn wir wissen:

$$\begin{aligned} \mathsf{P}(S_n^* \le b) &= \mathsf{P}(S_n \le b\sqrt{n\vartheta q} + n\vartheta) \\ &= \sum_{k=0}^{[b\sqrt{n\vartheta q}+n\vartheta]} b_{n,\vartheta}(k) \\ &\sim \frac{1}{\sqrt{2\pi}\sigma_n} \int_{0}^{b\sigma_n + n\vartheta} \mathrm{e}^{-\frac{1}{2n\vartheta q}(x-n\vartheta)^2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} \mathrm{e}^{-\frac{1}{2}\xi^2}\,\mathrm{d}\xi \end{aligned}$$

BEWEIS: (exakt)
$a \le S_n^* \le b$ genau dann, wenn $a\sigma_n + n\vartheta \le S_n \le b\sigma_n + n\vartheta$.

α_n sei kleinste ganze Zahl mit $\alpha_n \ge a\sigma_n + n\vartheta$,
β_n sei größte ganze Zahl mit $\beta_n \le b\sigma_n + n\vartheta$.

$\Rightarrow \qquad \{a \le S_n^* \le b\} = \{\alpha_n \le S_n \le \beta_n\}$.

Wir haben

$$\begin{aligned} v(n,\alpha_n) &= \tfrac{\alpha_n - n\vartheta}{\sigma_n} \quad < a + \frac{1}{\sigma_n} \\ b - \frac{1}{\sigma_n} \le \quad v(n,\beta_n) &= \tfrac{\beta_n - n\vartheta}{\sigma_n} \quad \le b\,. \end{aligned}$$

Es gilt $\alpha_n - 1 < a\sigma_n + n\vartheta \quad \Rightarrow \quad \alpha_n < a\sigma_n + n\vartheta + 1$

$$\Rightarrow \qquad |v(n,\alpha_n) - a| < \frac{1}{\sigma_n}$$

$\Rightarrow \quad v(n,\alpha_n), v(n,\beta_n)$ beschränkt.
$\Rightarrow \quad \frac{1}{\sqrt{n}} v^3(n,\alpha_n) \to 0,\ \frac{1}{\sqrt{n}} v^3(n,\beta_n) \to 0$.

Wegen der Gleichmäßigkeit der Konvergenz von $\frac{b_{n,\vartheta}(k_n)}{\frac{1}{\sigma_n}\varphi(v(n,k_n))}$ folgt:
Es existiert eine Folge $\varepsilon_n \to 0$ mit

$$1 - \varepsilon_n \le \frac{b_{n,\vartheta}(k)}{\frac{1}{\sigma_n}\varphi(v(n,k))} \le 1 + \varepsilon_n$$

für $\alpha_n \le k \le \beta_n$.

$(1-\varepsilon_n)\frac{1}{\sigma_n}\varphi(v(n,k)) \le b_{n,\vartheta}(k) \le (1+\varepsilon_n)\frac{1}{\sigma_n}\varphi(v(n,k))$

$$\mathsf{P}(a \le S_n^* \le b) = \sum_{k=\alpha_n}^{\beta_n} b_{n,\vartheta}(k)$$

$\Rightarrow \quad (1-\varepsilon_n)R_n \le \mathsf{P}(a \le S_n^* \le b) \le (1+\varepsilon_n)R_n$

$$\begin{aligned} \text{für } R_n &= \sum_{k=\alpha_n}^{\beta_n} \tfrac{1}{\sigma_n}\varphi(v(n,k)) \\ &= \tfrac{1}{\sigma_n}\tfrac{1}{\sqrt{2\pi}} \sum_{k=\alpha_n}^{\beta_n} \mathrm{e}^{-\frac{1}{2\sigma_n^2}(k-n\vartheta)^2} \end{aligned}$$

$k = \alpha_n \quad \Rightarrow \quad \frac{(k-n\vartheta)^2}{\sigma_n^2} = \frac{(\alpha_n - n\vartheta)^2}{\sigma_n^2} = \quad v^2(n,\alpha_n)$
$k = \beta_n \quad \Rightarrow \quad\quad v^2(n,\beta_n) \qquad$ beide beschränkt
d.h. mit $n \to \infty$ wird das Intervall $[v(n,\alpha_n), v(n,\beta_n)]$ in Intervalle unterteilt. Für $n \to \infty$ gilt für diese Riemann-Summen

$$\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_n}\sum_{k=\alpha_n}^{\beta_n} \mathrm{e}^{-\frac{1}{2\sigma_n^2}(k-n\vartheta)^2} - \int\limits_{v(n,\alpha_n)}^{v(n,\beta_n)} \varphi(\xi)\,\mathrm{d}\xi \to 0\,.$$

$$\begin{array}{l} v(n,\beta_n) \to b \\ v(n,\alpha_n) \to a \end{array} \qquad \text{(Integral über Histogramm)}$$

$$\Rightarrow \qquad \mathsf{P}(a \le S_n^* \le b) \to \Phi(b) - \Phi(a)\,.$$

□

Teil II
Statistik

11 Vorbemerkungen

Bisher betrachteten wir Zufallsgrößen $Y_1, Y_2, \ldots$ und wir definierten Werte oder Funktionen, die mit $Y = (Y_1, \ldots, Y_n)$ zusammenhingen. $(\Omega, \mathfrak{A}) \xrightarrow{Y} (\mathcal{Y}, \mathfrak{C})$. Für $(\mathcal{Y}, \mathfrak{C}) \subseteq (\mathbb{R}^n, \mathfrak{B})$ war

$$\begin{aligned}
\mathsf{P}(Y \le y) = F^Y(y) &\quad \text{die Verteilungsfunktion,} \\
\mathsf{P}(Y \le y) = \int_{-\infty}^{y} f^Y(\xi)\,\mu(\,\mathrm{d}\xi) &\quad \text{für die } \mu\text{–Dichte } f^Y, \\
\mathsf{E}e^{it'Y} = \varphi_Y(t) = \int e^{it'y}\,\mathrm{d}F^Y(y) &\quad \text{die charakteristische Funktion,} \\
\mathsf{E}Y^k \text{ bzw. } \mathsf{E}(Y - \mathsf{E}Y)^k &\quad \text{für } n = 1 \text{ das } k\text{–te nichtzentrale} \\
&\qquad \text{bzw. zentrale Moment von } Y, \\
\mathsf{E}Y &\quad \text{der Erwartungswert von } Y, \\
\mathsf{Var}\,Y &\quad \text{die Varianzmatrix von } Y.
\end{aligned}$$

Es existieren Bedingungen für den zentralen Grenzwertsatz

$$\mathsf{P}\left(\frac{\sum_{i=1}^n Y_i - n\mathsf{E}Y_1}{\sqrt{n\mathsf{Var}\,Y_1}} \le y\right) \to \Phi(y)$$

oder das Gesetz der großen Zahlen,

$$\overline{Y}_n \to \mu, \text{schwach oder stark.}$$

Dabei lassen sich F^Y, φ_Y, f^Y, $\mathsf{E}Y^k$, $\mathsf{E}Y$, $\mathsf{Var}\,Y$ berechnen, der zentrale Grenzwertsatz anwenden, wenn P^Y oder äquivalente Größen bekannt sind. Meist ist dies nicht der Fall. Wir haben $(\Omega, \mathfrak{A}, \mathsf{P})$ als Wahrscheinlichkeitsraum. Y als Funktion von ω ist uns nicht gegeben. Wenn es bekannt wäre, so würde gelten

$$F^Y(y) = \mathsf{P}(Y \le y) = \mathsf{P}^Y((-\infty, y]) \quad \text{oder}$$

$$\mathsf{E}Y = \int Y(\omega)\,\mathsf{P}(\,\mathrm{d}\omega), \ldots$$

In der Wahrscheinlichkeitstheorie werden Gesetze und Darstellungen bestimmt, wenn $\mathsf{P}, Y, \ldots$ bekannt sind.

So sagt z.B. das Gesetz der großen Zahlen:

$$\overline{Y}_n \overset{\text{f.s.}}{\to} \mathsf{E}Y_1 \qquad \text{für unabhängige, identisch verteilte } Y_i, \text{ wenn } \mu = \mathsf{E}Y_1 \text{ existiert.}$$

D.h. $\lim \overline{Y}_n(\omega) = \mu$ P–fast sicher.

Die Statistik behandelt das zur Wahrscheinlichkeitstheorie inverse Problem. Gegeben seien Werte $y_1, y_2, \ldots, y_n$, sie werden betrachtet als Realisierungen von Zufallsgrößen $Y_1, Y_2, \ldots$ d.h. $y_i = Y_i(\omega)$. Kann man $F^Y, \mathsf{E}Y, \ldots$ für $Y = (Y_1, \ldots, Y_n)$ approximieren oder bestimmen?

Zum Beispiel: Es werde ein Prozeß mehrere Male $(= n)$ durchgeführt, und am Ende das interessierende Ergebnis abgelesen: $y_1, \ldots, y_n$. Wir nehmen an, $y_i = Y_i(\omega)$ für i.i.d. Y_i. Falls $\mathsf{E}Y_1$ existiert, so gilt

$$\begin{aligned} \mathsf{E}Y_1 &= \int y\, \mathsf{P}^Y(\,\mathrm{d}y), \\ F^Y(y) &= \mathsf{P}^y((-\infty, y]). \end{aligned}$$

Als Approximation für diese Werte würde aus dem starken Gesetz der großen Zahlen folgen:

$$\frac{1}{n}(Y_1 + \cdots + y_n) \approx \mathsf{E}Y_1$$

$\Rightarrow$ Eine Approximation für den Erwartungswert einer Zufallsgrße Y_1 ist das arithmetische Mittel $\overline{Y}_n$, denn es gilt $\lim \overline{Y}_n = \mathsf{E}Y_1$ P–fast sicher.

$$\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i \Rightarrow \mathsf{E}\overline{Y}_n = \mathsf{E}Y_1,$$

d.h. „im Mittel“ ist die Approximation gleich $\mathsf{E}Y_1$.
Nun zur Approximation von $F^{Y_1}(z)$. Wir haben die i.i.d. $Y_1, \ldots, Y_n$. Die Indikatorfunktion von $(-\infty, z]$ ist

$$I_{(-\infty,z]}(y).$$

Das ist für festes $z \in \mathbb{R}^1$ als Funktion von y eine meßbare Funktion.

$$\begin{aligned} &\Rightarrow I_{(-\infty,z]}(Y_1) \quad \text{ist Zufallsgröße,} \\ &\mathsf{E}I_{(-\infty,z]}(Y_1) = \int_{-\infty}^{z} \mathsf{P}^{Y_1}(\,\mathrm{d}\xi) = F^{Y_1}(z). \\ &\Rightarrow I_{(-\infty,z]}(Y_1), \ldots, I_{(-\infty,z]}(Y_n) \\ &\quad \text{sind i.i.d. Zufallsgrößen mit Erwartungswert} F^{Y_1}(z), \\ &\text{Varianz} = F^{Y_1}(z)(1 - F^{Y_1}(z)). \end{aligned}$$

Zur Approximation der Verteilungsfunktion F^{Y_1} wählt man nach dem starken Gesetz der großen Zahlen:

$$\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,z]}(Y_i)$$

Hierfür gilt:

Satz 11.1 *Für die durch $\hat{F}_n$ definierte Zufallsgröße gilt*

$$\mathsf{E}\hat{F}_n(z) = F^{Y_1}(z),$$
$$\hat{F}_n(z) \overset{f.s.}{\Rightarrow} F^{Y_1}(z) \qquad \text{für jedes } z.$$

Beweis ist oben gegeben.

Diese Beispiele zeigen uns, daß aus Realisierungen von Zufallsgrößen der Erwartungswert der Zufallsgröße sowie deren Verteilungsfunktion approximiert werden können. Das Ziel der Statistik besteht darin, aus Realisierungen Aussagen über charakterisierende Werte oder Verteilungen von Zufallsgrößen abzuleiten.

Grundvoraussetzungen:

Daten $y_1, \ldots, y_n$ sind Realisierungen von Zufallsgrößen $Y_1, \ldots, Y_n$,

$(\Omega, \mathfrak{A}, \mathsf{P})$ Wahrscheinlichkeitsraum, $(\Omega, \mathfrak{A}) \overset{Y_i}{\to} (\mathcal{Y}, \mathfrak{C})$,

$Y = (Y_1, \ldots, Y_n)$ Zufallsgröße,

$Y(\omega) = (Y_1(\omega), \ldots, Y_n(\omega))$ heißt Stichprobe.

Falls $Y_1, \ldots, Y_n$ unabhängige und identisch verteilte Zufallsgrößen sind, so heißt Y Stichprobe vom Umfang n.

Die Verteilung P^Y ist im allgemeinen unbekannt, man kennt bestenfalls eine Klasse von möglichen Verteilungen, in der P^Y enthalten sein wird. Eine solche Klasse $\mathcal{P}$ heißt *statistisches Modell*. Jedes solches Modell kann parametrisiert werden , zur einfacheren Beschreibung ist das günstig: $\mathcal{P} = \{\mathsf{P}_\vartheta : \vartheta \in \Theta\}$ Θ heißt Parameterraum.

Beispiel 11.1 $\qquad \mathcal{P} = \{\mathsf{P}_\vartheta(y_1, \ldots, y_n) = \prod_{i=1}^{n} \mathsf{P}_\vartheta^{Y_i}(y_i), \vartheta \in [0,1], \mathsf{P}_\vartheta^{Y_i} = B(1, \vartheta)\}.$

Das heißt, als statistisches Modell betrachten wir unabhängige, binomialverteilte Zufallsgrößen.

Beispiel 11.2 $\mathcal{P} = \{\mathsf{P}_\vartheta = \mathsf{N}_n(\mu, \Sigma), \mu \in \mathbb{R}^n, \Sigma \in \mathcal{M}_{n\times n}^{>}\}.$

Die Konstruktion oder defi eines statistischen Modells ist der erste Schritt. Ziel ist es, aus den zufallsabhängigen Beobachtungen Aussagen oder Schlüsse über die Verteilungen P^Y

herzuleiten. Solche Aussagen nennt man Entscheidungen. Daraus folgt, Entscheidungen sind zufällig, da eine sinnvolle Entscheidung von der Stichprobe abhängig sein muß.

Sei Δ die Menge der Entscheidungen, mit der σ–Algebra $\tilde{\mathcal{D}}$, so daß $(\Delta, \tilde{\mathcal{D}})$ ein meßbarer Raum (Entscheidungsraum) ist.

Definition 11.1 *Unter einer (nichtrandomisierten) Entscheidungsfunktion versteht man eine meßbare Abbidung:* $(\mathcal{Y}, \mathfrak{C}) \to (\Delta, \tilde{\mathcal{D}})$.

12 Entscheidungsprobleme

12.1 Testen von Hypothesen

Das statistische Modell sei $\mathcal{P} = \{\mathsf{P}_\vartheta : \ \vartheta \in \Theta\}$. Es werden zwei Annahmen über die Verteilungen formuliert:

Hypothese: $Y \sim \mathsf{P}_\vartheta, \quad \vartheta \in \Theta_H \subset \Theta$,

Alternative: $Y \sim \mathsf{P}_\vartheta, \quad \vartheta \in \Theta_K = \Theta \setminus \Theta_H$.

Beide Annahmen (Hypothese H und Gegenhypothese K) werden durch Angaben über ϑ beschrieben.

Das Problem des Testens (Prüfens) von

$$H : \vartheta \in \Theta_H \qquad \text{gegen} \qquad K : \vartheta \in \Theta_K$$

heißt Testproblem.

Der Stichprobenraum ist $(\mathcal{Y}, \mathfrak{C})$, der Entscheidungsraum ist $(\Delta, \tilde{\mathcal{D}})$ für $\Delta = \{d_H, d_K\}, \tilde{\mathcal{D}} = \mathcal{P}(\Delta)$, dabei ist d_H Entscheidung für H, d_K Entscheidung für K. Ist T die gewählte Entscheidungsfunktion, so kann $T(y) = d_H$ oder $T(y) = d_K$ gelten.

Definition 12.1 *Eine meßbare Funktion* $T : (\mathcal{Y}, \mathfrak{C}) \to (\Delta, \tilde{\mathcal{D}})$ *heißt Testfunktion.*

Beispiel 12.1 Es sollen Eigenschaften eines elektrischen Netzgerätes untersucht werden. Gemessen werden in unabhängigen Versuchen die Werte

$$221.3,\ 220.8,\ 223.1,\ 223.0,\ 220.9,\ 221.4,\ 221.9,\ 223.0,\ 224.5,\ 219.5.$$

Es soll geprüft werden, ob die Meßwerte normalverteilt sind mit Erwartungswert $= 220$ und einer Varianz 2.

Lösungsmöglichkeit: Berechnung der empirischen Verteilungsfunktion $\hat{F}_{10}(y)$ und Vergleich mit $\Phi(\frac{y-220}{\sqrt{2}})$,

$$\sup_y \left| \hat{F}_{10}(y) - \Phi\left(\frac{y-220}{\sqrt{2}}\right) \right| > c,$$

so H ablehnen.

Das geschieht, falls man annimmt, daß die Verteilung von den Beobachtungen beliebig sein kann. Wenn man voraussetzt, daß die y_i Realisierungen normalverteilter Variabler sind, so heißt das

$$\begin{aligned}
&Y_i \sim \mathsf{N}(\mu, \sigma^2) \qquad \text{und}\\
&H : \mu = 220, \qquad \sigma^2 = 2\\
&K : \mu \neq 220 \qquad \text{oder} \quad \sigma^2 \neq 2.
\end{aligned}$$

Es gilt $\hat{\mu} = 221.94, s^2 = 2.13$. Wenn $(\hat{\mu}, s^2)$ weit von $(220, 2)$ liegt, so wird H abgelehnt.

Bei einem Test können zwei mögliche Fehler gemacht werden:

Fehler 1. Art: Entscheidung für K, obwohl H richtig ist;

Fehler 2. Art: Entscheidung für H, obwohl K richtig ist.

Wahrscheinlichkeit für Fehler 1. Art = Irrtumswahrscheinlichkeit 1. Art (Signifikanzniveau)

$$\mathsf{P}_\vartheta(T(y) = d_K), \qquad \vartheta \in \Theta_H$$

Wahrscheinlichkeit für Fehler 2. Art = Irrtumswahrscheinlichkeit 2. Art

$$\mathsf{P}_\vartheta(T(y) = d_H) = 1 - \mathsf{P}_\vartheta(T(y) = d_K), \qquad \vartheta \in \Theta_K$$

$\Rightarrow$ Beide Fehler sind beschrieben durch die Gütefunktion

$$\beta_T(y) := \mathsf{P}_\vartheta(T(y) = d_K).$$

Bisher ist keiner der Fehler ausgezeichnet. Es erweist sich als günstig, eine nichtsymmetrische Behandlung der Fehler vorzunehmen.

Definition 12.2 *Zu vorgegebenem α, $0 < \alpha < 1$, heißt eine Testfunktion T α–Testfunktion oder kurz α–Test, falls $\beta_T(\vartheta) \leq \alpha$ für alle $\vartheta \in \Theta_H$ gilt, d.h. die Irrtumswahrscheinlichkeit 1. Art ist $\leq \alpha$. Man wählt $\alpha = 0.05$ oder $0,01$.*

$\Rightarrow$ Wenn T ein α–Test ist, so ist die Entscheidung für K, obwohl H richtig ist, nur in $100\% \cdot \alpha$ der Fälle zu erwarten, d.h. nur in 5 bzw. 1 Fällen. Das bedeutet, daß eine Entscheidung für K faktisch nur zu erwarten ist, falls K auch richtig ist. Damit ist die Entscheidung für K statistisch abgesichert (mit dem Signifikanzniveau α).

Beispiel 12.2 Es gibt ein erfolgreiches Medikament für eine Krankheit. Die Pharmaindustrie hat ein neues Medikament für eben diese Krankheit entwickelt. Man hat also bei einer Krankheit die Möglichkeit zu zwei Behandlungen I, II. Sinnvoll ist es, das neue Medikament nur dann zuzulassen, wenn es besser ist.

Beurteilungsmaß für den Krankheitszustand sei eine Größe y. Eine Personengruppe von n_I Patienten erhält Behandlung I, eine andere mit n_{II} Patienten Behandlung II. Die Meßwerte seien $y_{\mathrm{I}1}, \ldots, y_{\mathrm{I}n_1}$ und $y_{\mathrm{II}1}, \ldots, y_{\mathrm{II}n_2}$.

$H : Y_{\mathrm{II}i}$ und Y_{Ij} sind identisch verteilt.

$K : Y_{\mathrm{II}i}$ ist durchschnittlich größer als Y_{Ij}.

K abgesichert bedeutet, daß die Entscheidung für das neue Medikament wirklich nur dann erfolgt, wenn es besser ist.

Lösungsmöglichkeit:

$$\begin{aligned}\mathcal{P} &= \{\text{Normalverteilungen}\mathsf{N}(\vartheta, \sigma_0^2), \quad \sigma_0^2\text{bekannt}\}. \\ &= \{\mathsf{P}_\vartheta : \ \mathsf{P}_\vartheta = \mathsf{N}(\vartheta, \sigma_0^2), \quad \vartheta \in \mathbb{R}^1\}.\end{aligned}$$

Da das Medikament I schon lange genutzt wird, weiß man, daß $\vartheta \leq \vartheta_0$ unter I für bekanntes ϑ_0 gilt. Dann würde K bedeuten: $\vartheta > \vartheta_0$.

Das Testproblem

$$H : \vartheta \leq \vartheta_0 \quad \text{gegen} K : \vartheta > \vartheta_0$$

unter $\mathsf{P}_\vartheta = \mathsf{N}(\vartheta, \sigma_0^2)$ heißt einseitiges Gauß–Testproblem.

Satz 12.1 *Sei $\mathsf{P}_\vartheta = \mathsf{N}(\vartheta, \sigma_0^2)$. Zu einer Stichprobe vom Umfang n ist der Test*

$$T(y) = \begin{cases} d_K & \\ & \textit{für} t(y) = \frac{\sqrt{n}}{\sigma_0}(\overline{y} - \vartheta_0) \\ d_H & \end{cases} \begin{matrix} > c_\alpha \\ \\ \leq c_\alpha \end{matrix}$$

ein α–Test und er ist unverfälscht (d.h. $\mathsf{P}_\vartheta(T(y) = d_K) \geq \alpha$ für $\vartheta > \vartheta_0$) für das α–Fraktil c_α der Normalverteilung, d.h. $1 - \alpha = \Phi(c_\alpha)$.

Beweis: $Y_1, \ldots, Y_n$ sind i.i.d., $Y_i \sim \mathsf{N}(\vartheta, \sigma_0^2)$.

$$\begin{aligned}&\Rightarrow \overline{Y} = \frac{1}{n} \sum Y_i \sim \mathsf{N}(\vartheta, \frac{1}{n}\sigma_0^2) \\ &\Rightarrow t(Y) = \frac{\sqrt{n}}{\sigma_0}(\overline{Y} - \vartheta_0) \sim \mathsf{N}(\frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0), 1).\end{aligned}$$

$$\begin{aligned}\mathsf{P}_\vartheta(t(y) > c) &= \frac{1}{\sqrt{2\pi}} \int_c^\infty e^{-\frac{(x - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0))^2}{2}} \, \mathrm{d}x \\ &= 1 - \Phi(c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)) \\ &\Rightarrow \mathsf{P}_\vartheta(t(y) > c) \leq 1 - \Phi(c) \quad \text{für } \vartheta \leq \vartheta_0, \\ &\Rightarrow \text{Falls} \quad 1 - \alpha = \Phi(c_\alpha) \quad \text{folgt}\end{aligned}$$

$$\mathsf{P}_\vartheta(t(y) > c_\alpha) \leq \alpha \quad \text{für } \vartheta \leq \vartheta_0.$$

Ein anderes c_α führt entweder zu keinem α–Test oder er ist nicht unverfälscht.

Die Gütefunktion $\beta_T(\vartheta) = \mathsf{P}_\vartheta(t(y) > c_\alpha) = 1 - \Phi(c_\alpha - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0))$ ist wachsend in ϑ.
$\Rightarrow$ für c_α mit $1 - \Phi(c_\alpha) = \alpha$ gilt

$$1 - \Phi(c_\alpha - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)) \geq \alpha.$$

$\square$

Beispiel 12.3 Es seien die Y_i normalverteilt. Gehören die 10 Daten zu einer Stichprobe von $\mathsf{N}(220, 2)$? Das führt zum Testproblem

$$\begin{aligned} H &: \vartheta \leq 220 \\ K &: \vartheta > 220. \end{aligned}$$

Wir haben $t(y) = \frac{\sqrt{10}}{\sqrt{2}}(\overline{y} - 220) = 4.34$, denn $\overline{y} = 221.94$. Für $\alpha = 0.05$ gilt $c_\alpha = 1.64$.
$\Rightarrow K$ wird angenommen.

Zweiseitiger Gaußtest: Es gelte $\mathsf{P}_\vartheta = \mathsf{N}(\vartheta, \sigma_0^2)$. Zu testen ist

$$H : \vartheta = \vartheta_0 \quad \text{gegen} \quad K : \vartheta \neq \vartheta_0.$$

Wesentlich ist hier der Abstand $|\vartheta - \vartheta_0|$. Darüber gibt die Funktion $|t(y)|$ Auskunft.

Satz 12.2 *Sei $\mathsf{P}_\vartheta = \mathsf{N}(\vartheta, \sigma_0^2)$. Für den zweiseitigen Gaußtest ist*

$$T(y) = \begin{cases} d_K & > c_{\alpha/2} \\ & \text{für } |t(y)| \\ d_H & \leq c_{\alpha/2} \end{cases}$$

ein α–Test und er ist unverfälscht. Hier ist $c_{\alpha/2}$ das $\frac{\alpha}{2}$-Fraktil der Normalverteilung.

BEWEIS:

$$\begin{aligned} t(Y) &\sim N(\frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0), 1) \\ \Rightarrow \mathsf{P}_\vartheta(|t(y)| > c) &= \mathsf{P}_\vartheta(t(y) < -c \quad \text{oder} \quad t(y) > c) \\ &= \mathsf{P}_\vartheta(t(y) < -c) + \mathsf{P}_\vartheta(t(y) > c) \\ &= \Phi\left(-c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right) + 1 - \Phi\left(c - \frac{\sqrt{n}}{\sigma_0}(th - \vartheta_0)\right). \end{aligned}$$

Unter H:

$$\mathsf{P}_{\vartheta_0}(|t(y)| > c) = \Phi(-c) + 1 - \Phi(c) = 2(1 - \Phi(c)) \leq \alpha \quad \text{für}$$
$$c \geq c_{\alpha/2}$$

$$\text{Sei} \quad H(\vartheta) := \Phi\left(-c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right) + 1 - \Phi\left(c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right)$$
$$\begin{aligned}\Rightarrow H'(\vartheta) &= \varphi\left(-c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right) \cdot \left(-\frac{\sqrt{n}}{\sigma_0}\right) \\ &\quad + \varphi\left(c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right)\left(\frac{\sqrt{n}}{\sigma_0}\right) \\ &= \frac{\sqrt{n}}{\sigma_0}\left[\varphi\left(c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0)\right) - \varphi\left(-c - \frac{\sqrt{n}}{\sigma_0}(\vartheta - \vartheta_0\right)\right],\end{aligned}$$

wobei φ die Dichte der $\mathsf{N}(0,1)$–Verteilung ist.

$$\Rightarrow H'(\vartheta_0) = 0,$$
$$H'(\vartheta) \begin{cases} < 0 & \text{für} \quad \vartheta < \vartheta_0, \\ > 0 & \text{für} \quad \vartheta > \vartheta_0. \end{cases}$$

$\Rightarrow c = c_{\alpha/2}$ liefert unverfälschten α–Test.

□

Fazit: Die speziell gewählten Hypthesen und Alternativen haben entscheidenden Einfluß auf den zu wählenden Test.

Die Funktion $t(y)$ bzw. $|t(y)|$ heißt Prüfgröße, c_α oder $c_{\alpha/2}$ heißen kritische Werte.

Beispiel 12.4 Für die 10 Meßwerte: $t(y) = 4.34, \alpha = 0.05, c_{\alpha/2} = 1.96$.
Also sind H abgelehnt.

Nach dem Test hat man die Situation: H oder K gilt. Wenn K gilt, so interessiert man sich für den Wert ϑ, der die Verteilung der Zufallsgröße bestimmt.

12.2 Punktschätzungen

Sei

$$\mathcal{P} = \{\mathsf{P}_\vartheta, \vartheta \in \Theta\}$$

Man interessiert sich für ϑ oder $\gamma(\vartheta)$. Zum beispiel bei Normalverteilungen

$$\mathcal{P} = \{\mathsf{N}(\mu, \sigma^2) : \vartheta = (\mu, \sigma^2) \in \mathbb{R}^1 \times \mathbb{R}^+\}$$
$$\gamma(\vartheta) = \mu \quad \text{oder} \quad \gamma(\vartheta) = \sigma^2 \quad \text{oder} \quad \gamma(\vartheta) = (\mu, \sigma^2)$$
$$\gamma(\Theta) \subseteq \Gamma.$$

Definition 12.3 *Eine meßbare Abbildung* $\hat{\gamma} : (\mathcal{Y}, \mathfrak{C}) \to (\Gamma, \mathfrak{G})$ *heißt Punkt–Schätzfunktion.*

> $\hat{\gamma}$ *Schätzfunktion, Schätzer*
>
> $\hat{\gamma}(y)$ *Schätzwert oder Schätzung*
>
> $\hat{\gamma}(Y)$ *Schätzvariable.*

Eigenschaften von Schätzern:

Definition 12.4 $\hat{\gamma}$ *heißt erwartungstreu für* $\gamma(\vartheta)$*, falls*

$$\mathsf{E}_\vartheta \hat{\gamma}(Y) = \gamma(\vartheta) \quad \forall \vartheta \in \Theta,$$

$b_\vartheta(\hat{\gamma}) = \mathsf{E}_\vartheta \hat{\gamma}(Y) - \gamma(\vartheta)$ *heißt Verzerrung von* $\hat{\gamma}$.

Beispiel 12.5 $Y_i \sim \mathsf{N}(\vartheta, \sigma^2), i = 1, \ldots, n.$

> Ist $\hat{\gamma}(y) = c = \text{const}$ ist ein Schätzer, aber kein erwartungstreuer Schätzer.
>
> $\overline{Y} = \frac{1}{n} \sum Y_i$ ist erwartungstreuer Schätzer, $\mathsf{Var}\, \overline{Y} = \frac{1}{n}\sigma^2$.

Beispiel 12.6 Geometrische Verteilung.
Es wird ein erwartungstreuer Schätzer hergeleitet.

Beispiel 12.7 Binomialverteilung.
$\overline{Y} = \frac{1}{n} \sum Y_i$ ist erwartungstreuer Schätzer mit der Varianz $\frac{1}{n}\vartheta(1 - \vartheta)$.

Bei der Normalverteilung nehmen wir $\overline{Y}$ zur Schätzung des Erwartungswertes $= \int \xi \ \mathrm{d}F(\xi)$. Erinnern wir uns an die Zufallsgröße

$$\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, z]}(Y_i), \quad \text{so gilt}$$

$$\int \xi \ \mathrm{d}\hat{F}_n(\xi) = \frac{1}{n} \sum Y_i = \overline{Y}.$$

Nun soll die Varianz geschätzt werden.

$$\mathsf{Var}\, Y = \int \ \underset{\overline{Y}}{\underset{\uparrow}{(\xi - \mu)^2}} \ \underset{d\hat{F}_n(\xi)}{\underset{\uparrow}{\mathrm{d}F(\xi)}}$$

$$\hat{\sigma^2} = \int (\xi - \overline{Y})^2 \ \mathrm{d}\hat{F}_n(\xi) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

Erwartungswert von $\hat{\sigma}^2$:

$$\mathsf{E}_\vartheta \hat{\sigma}^2 = \frac{n-1}{n}\sigma^2 \Rightarrow \quad \text{erwartungstreu oder nicht?}$$

13 Literatur

Bauer, H.: Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. Berlin 1968.
Devore, J., R. Peck: Statistics. Belmont 1993.
Fisz, M.: Wahrscheinlichkeitsrechnung und Mathematische Statistik. Berlin 1989.
Gnedenko: Wahrscheinlichkeitstheorie. Berlin
Humak, K.M.S.: Statistische Methoden der Modellbildung. Berlin 1977.
Krengel, U.: Einf. in die Wahrscheinlichkeitstheorie und Statistik. Vieweg 1991
Pitman, J.: Probability. New York 1993.
Schmetterer, L.: Einführung in die Mathematische Statistik. Wien 1966.
Witting, H.: Mathematische Statistik. Stuttgart, Teubner 1978.