Stochastik I

Matthias Hahne und das latexki-Team

Dieses Dokument ist eine persönliche Vorlesungsmitschrift der Vorlesung Stochastik I im Sommersemester 2005 bei Prof. Dr. Bäuerle.

Diese Version des Skriptes ist angepasst an die Vorlesung von Prof. Dr. Bäuerle im Wintersemester 05/06 an der Universität Karlsruhe. Koordiniert wurde diese Arbeit über http://mitschriebwiki.nomeata.de/, einem LaTeX-Wiki von Joachim Breitner.

Stand: 3. Mai 2017

Inhaltsverzeichnis

Kapitel 1

Ereignisse und Wahrscheinlichkeiten

In der Stochastik werden zufallsabhängige Phänomene mathematisch modelliert und analysiert. (z.B. würfeln)
Ω = Ergebnisraum (z.B. $\Omega = \{1, 2, 3, 4, 5, 6\}$)
$A \subset \Omega$ Ereignis (z.B. A={2,4,6} $\hat{=}$ gerade Zahl fällt)
Ist $\omega \in \Omega$, so heißt $\{\omega\}$ Elementarereignis

Beispiel 1.1

a) Zuerst wird eine Münze geworfen. Fällt Kopf, wird mit einem Würfel geworfen, fällt Zahl so wird nochmal mit der Münze geworfen.
$\Omega = \{K1, K2, K3, K4, K5, K6, ZZ, ZK\}$

b) Rotierender Zeiger
$\Theta = 2\pi x$, $0 \leq x < 1$ sei der Winkel beim Stillstand
$\Omega = [0, 1)$
$A = (0, \frac{1}{4})$ ist das Ereignis: "Zeiger stoppt im I. Quadranten"

Verknüpfungen von Ereignissen werden durch mengentheoretische Operationen beschrieben.

Definition 1.1

a) *Seien $A, B \subset \Omega$ Ereignisse. So heißt*
$A \cap B = AB = \{\omega \in \Omega | \omega \in A \underline{\text{ und }} \omega \in B\} = \{\omega \in \Omega | \omega \in A, \omega \in B\}$ ***Durchschnitt von*** A ***und*** B.
$A \cup B = \{\omega \in \Omega | \omega \in A \underline{\text{ oder }} \omega \in B\}$ ***Vereinigung von*** A ***und*** B.
Sind A und B disjunkt, dh. $A \cap B = \emptyset$ dann schreiben wir auch $A + B$
$A \backslash B = \{\omega \in \Omega | \omega \in A, \omega \notin B\}$
Gesprochen: A ohne B. Spezialfall $A = \Omega$ Dann ist $\Omega \backslash B = B^c$
B^c heißt ***Komplement von*** B.

b) *Sind* $\Omega_1, \Omega_2, \ldots, \Omega_n$ *Ergebnisräume, so ist* $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n = \{(a_1, \ldots, a_n) | a_i \in \Omega_i, i = 1, \ldots, n\}$ *das* ***Kartesische Produkt****.*

Beispiel 1.2 2x würfeln
$\Omega = \{(i,j) \in \{1\ldots 6\}\}$
A = Erster Würfel ist eine 6 = $\{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} = \{(6,j)|j \in \{1\ldots 6\}\}$
B = Augensumme ist max 4 = $\{(i,j)|i+j \leq 4\}$ = {(1,1),(1,2),(1,3),(2,2),(3,1)}
$A \cap B = \emptyset$
B^c = Augensumme ist mindestens 5

Mit $\mathcal{P}(\Omega)$ bezeichnen wir die **Potenzmenge von** Ω, d.h. die Menge aller Teilmengen von Ω. (dazu gehören auch Ω und $\emptyset$). Wir wollen nun Ereignissen Wahrscheinlichkeiten zuordnen.
Es sei $\mathcal{A} \subset \mathcal{P}(\Omega)$ die Menge aller Mengen (Ereignisse) denen wir Wahrscheinlichkeiten zuordnen wollen. Um eine sinnvolle math. Theorie zu bekommen, können wir im Allgemeinen nicht $\mathcal{A} = B \backslash \Omega$ wählen. Jedoch sollte das Mengensystem $\mathcal{A}$ gewisse Eigenschaften haben.

Definition 1.2

a) $\mathcal{A} \subset \mathcal{P}(\Omega)$ *heißt* ***Algebra über*** Ω*, falls gilt:*

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

   (ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

   (iii) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \cup A_2 \in \mathcal{A}$

b) $\mathcal{A} \subset \mathcal{P}(\Omega)$ *heißt* ***σ-Algebra über*** Ω*, falls* $\mathcal{A}$ *eine Algebra ist und*

   (iv) $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Bemerkung 1.1

a) Das Paar $(\Omega, \mathcal{A})$ mit $\mathcal{A}$ σ-Algebra über Ω heißt **Messraum**

b) $\mathcal{P}(\Omega)$ ist stets eine σ-Algebra. Ist Ω endlich oder abzählbar unendlich, so kann $\mathcal{A} = \mathcal{P}(\Omega)$ gewählt werden. Ist Ω nicht abzählbar (siehe Beispiel 1.1), so muss eine kleinere σ-Algebra betrachtet werden. (Kapitel 4).

Wir wollen noch die folgenden Mengenverknüpfungen betrachten.

Definition 1.3 *Seien* $A_1, A_2, \ldots \subset \Omega$ *Dann heißt*

$$\limsup_{n\to\infty} A_n = \limsup_{n\to\infty} A_n := \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$$

*der **Limes Superior** der Folge $\{A_n\}$ und*

$$\liminf_{n\to\infty} A_n = \liminf_{n\to\infty} A_n := \bigcup_{k=1}^{\infty} \bigcap_{n=k}^{\infty} A_n$$

*heißt der **Limes Inferior** der Folge $\{A_n\}$*

Bemerkung 1.2 Es gilt:

$$\omega \in \limsup_{n\to\infty} A_n \Leftrightarrow \forall\, k \in \mathbb{N}\ \exists n \geq k : w \in A_n \Leftrightarrow |\{n \in \mathbb{N} | \omega \in A_n\}| = \infty$$

$\limsup_{n\to\infty} A_n$ist das Ereignis "unendlich viele A_n's treten ein"

$$\omega \in \liminf_{n\to\infty} A_n \Leftrightarrow \exists\, k \in \mathbb{N}\text{ ,so dass } \forall\, n \geq k\ \omega \in A_n$$

$\liminf_{n\to\infty} A_n$ist also das Ereignis "alle bis auf endlich viele der A_n's treffen ein¡'

Lemma 1.1 *Seien $A_1, A_2, \ldots, \subset \Omega$.*

a) Falls $\{A_n\}$ wachsend ist, d.h. $A_1 \subset A_2 \subset \ldots$, dann gilt:

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n$$

b) Falls $\{A_n\}$ fallend ist, d.h. $A_1 \supset A_2 \supset \ldots$, dann gilt:

$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n$$

Bemerkung 1.3

a) Für $A_1 \subset A_2 \subset A_3 \subset \ldots$ schreiben wir $A_n \uparrow$, für $A_1 \supset A_2 \supset \ldots$ schreiben wir $A_n \downarrow$

b) Falls
$$\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n \text{ schreiben wir kurz: } \lim_{n\to\infty} A_n$$

Beweis a) Sei

$$A := \bigcup_{n=1}^{\infty} A_n$$

Es gilt

$$\bigcup_{n=k}^{\infty} A_n = A \quad \forall\, k \in \mathbb{N} \Rightarrow \limsup_{n\to\infty} A_n = A$$

Andererseits:

$$\bigcap_{n=k}^{\infty} A_n = A_k \text{ , d.h. } \liminf_{n\to\infty} A_n = A$$

b) analog ■

Ereignissen ordnen wir jetzt Zahlen zwischen 0 und 1 zu, die wir als Wahrscheinlichkeiten interpretieren. Damit dies sinnvoll ist, soll die Zuordnung gewissen AXIOMEN genügen.

Definition 1.4 (Axiomensystem von Kolmogorov 1933)
Gegeben sei ein Messraum $(\Omega, \mathcal{A})$. Eine Abbildung $P : A \rightarrow [0,1]$ heißt Wahrscheinlichkeitsmaß auf $\mathcal{A}$, falls

(i) $P(\Omega) = 1$ *"Normiertheit"*

(ii)

$$P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \quad \forall \text{paarweise disjunkten}\, A_1, A_2, \ldots \in \mathcal{A}$$

(d.h. $A_i \cap A_j = \emptyset \quad \forall i \neq j$*) "$\sigma -$ Additivität"*

$(\Omega, \mathcal{A}, P)$ heißt ***Wahrscheinlichkeitsraum****.*

Beispiel 1.3
Ist $\Omega \neq \emptyset$ eine endliche Menge und $\mathcal{A} = P(\Omega)$, so wird durch $P(A) = \frac{|A|}{|\Omega|} \, \forall A \subset \Omega$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ definiert.
$(\Omega, \mathcal{A}, P)$ nennt man Laplace'schen Wahrscheinlichkeitsraum. Jedes Elementarereignis hat hier die gleiche Wahrscheinlichkeit $\frac{1}{|\Omega|}$.
Wir betrachten den gleichzeitigen Wurf zweier Würfel. Wie groß ist die Wahrscheinlichkeit, dass die Augensumme 11 bzw. 12 ist?
2 Würfel
$\Omega = \{(i,j)|i,j = \{1 \ldots 6\}\}$
$|\Omega| = 36$
A = Augensumme 11
B = Augensumme 12
P(A) = P({(5,6),(6,5)}) = $\frac{2}{36}$
P(B) = P({(6,6)}) = $\frac{1}{36}$

Satz 1.2
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $A, B, A_1, A_2, \ldots \in \mathcal{A}$. Dann gilt:

a) $P(A^c) = 1 - P(A)$

b) Monotonie: $A \subset B \Rightarrow P(A) \leq P(B)$

c)

$$\text{Endliche Additivität} P(\sum_{k=1}^{n} A_k) = \sum_{k=1}^{n} (P(A_k)) \,\text{für paarweise disjunkte}\, A_1 \ldots A_n$$

d) $P(A \cup B) = P(A) + P(B) - P(\overbrace{A \cap B}^{=AB})$

e) *Boole'sche Ungleichung:*

$$P(\bigcup_{k=1}^{n} A_k) \leq \sum_{k=1}^{n} P(A_k) \qquad \forall\, n \in \mathbb{N}$$

Beweis a) Es gilt:

$1 = P(\Omega) = P(A{+}A^c) = P(A{+}A^c{+}\emptyset{+}\emptyset \ldots) \overset{(ii)}{=} P(A){+}P(A^c){+}P(\emptyset){+}P(\emptyset) \ldots$
$\underline{\Rightarrow P(\emptyset) = 0}$ und $P(A^c) = 1 - P(A)$

b) $P(B) = P(A + B\backslash A) = P(A) + \underbrace{P(\overbrace{B\backslash A}^{=B\cap A^c \in \mathcal{A}})}_{\geq 0} \geq P(A)$

c) Setze $A_{n+1} = A_{n+2} = \cdots = \emptyset$ und verwende die $\sigma - Additivität.$

d) Es gilt: $A \cup B = A + B\backslash A \Rightarrow P(A \cup B) = P(A) + P(B\backslash A)$

$B = B\backslash A + A \cap B \Rightarrow P(B) = P(B\backslash A) + P(\overbrace{A \cap B}^{=AB})$
Es folgt: $P(A \cup B) = P(A) + P(B) - P(AB)$

e) Für n=2 folgt die Aussage aus Teil d), da $P(AB) \geq 0$
Induktion: $n \to n + 1$:

$$P(\bigcup_{k=1}^{n+1} A_k) = P(\bigcup_{k=1}^{n} A_k \cup A_{n+1}) \leq P(\bigcup_{k=1}^{n} A_k) + P(A_{n+1}) \leq \sum_{k=1}^{n} P(A_k) + P(A_{n+1})$$

■

Satz 1.3 (Siebformel)
Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, dann gilt für $A_1 \ldots A_n \in \mathcal{A}$:

$$P(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} (-1)^{k-1} \cdot \sum_{1 \leq i_1 \leq i_2 \cdots \leq i_k \leq n} P(A_{i1} \cap \cdots \cap A_{ik})$$

Bemerkung 1.4

a) Die Formel ist auch unter dem Namen: Formel von Poincare-Sylvester oder Formel des Ein- und Ausschließens bekannt.

b) n=2 $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$

Die σ-Additivität ist äquivalent zu einer gewissen Stetigkeit des Wahrscheinlichkeitsmaßes.

Satz 1.4 *Sei $(\Omega, \mathcal{A})$ ein Messraum und sei $P : \mathcal{A} \to [0, 1]$ eine beliebige additive Mengenfunktion, d.h. $P(A + B) = P(A) + P(B)$ gelte für disjunkte $A, B \in \mathcal{A}$. Außerdem sei $P(\Omega) = 1$. Dann sind die folgenden Aussagen äquivalent:*

a) P ist σ-additiv (und damit ein Wahrscheinlichkeitsmaß)

b) P ist stetig von unten, d.h. für $A_n \in \mathcal{A}$ mit $A_n \uparrow$ gilt:

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$$

c) P ist stetig von oben, d.h. für $A_n \in \mathcal{A}$ mit $A_n \downarrow$ gilt:

$$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$$

d) P ist stetig in $\emptyset$, d.h. für $A_n \in \mathcal{A}$ mit $A_n \downarrow \emptyset$ gilt:

$$\lim_{n\to\infty} P(A_n) = 0$$

Beweis *a)* $\Rightarrow$ *b)* Es sei $A_0 := \emptyset$. Dann gilt:

$$P(\lim_{n\to\infty} A_n) \overset{L.1.1}{=} P(\bigcup_{k=1}^{\infty} A_k) = P(\sum_{k=1}^{\infty}(A_k \backslash A_{k-1})) = \sum_{k=1}^{\infty} P(A_k \backslash A_{k-1}) =$$

$$= \lim_{n\to\infty} \sum_{k=1}^{\infty} P(A_k \backslash A_{k-1}) = \lim_{n\to\infty} P(A_n)$$

b) $\Rightarrow$ *c)* $A_n \downarrow \Rightarrow A_n^c \uparrow$ und

$$\lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} 1 - P(A_n^c) = 1 - \lim_{n\to\infty} P(A_n^c) \overset{b)}{=} 1 - P(\bigcup_{n=1}^{\infty} A_n^c) \overset{\text{d'Morgan}}{=} P(\bigcap_{n=1}^{\infty} A_n)$$

Mit Lemma 1.1 folgt die Behauptung

c) $\Rightarrow$ *d)* klar (d) Spezialfall von c))

d) $\Rightarrow$ *a)* Seien $A_1, A_2, \ldots \in \mathcal{A}$ paarweise disjunkt. Dann gilt

$$\bigcup_{k=n+1}^{\infty} A_k \downarrow \emptyset \qquad \text{für}\, n \to \infty$$

Also

$$P(\sum_{k=1}^{\infty} A_k) = P(\sum_{k=1}^{n} A_k + \sum_{k=n+1}^{\infty} A_k) \overset{\text{P endl. Add.}}{=} \sum_{k=1}^{n} P(A_k) + P(\sum_{k=n+1}^{\infty} A_k)$$

Für $n \to \infty$ gilt: $P(\underbrace{\sum_{k=n+1}^{\infty} A_k}_{B_n}) \to 0$ und die Behauptung folgt. ∎

Kapitel 2

Kombinatorik und Urnenmodelle

In diesem Abschnitt nehmen wir an, dass $(\Omega, \mathcal{A}, P)$ ein Laplace'scher Wahrscheinlichkeitsraum ist (vgl. Bsp.1.3), d.h. Ω ist endlich, $\mathcal{A} = P(\Omega)$ und $P(A) = \frac{|A|}{|\Omega|} \forall A \subset \Omega$. Für reale Vorgänge muss zunächst der "richtige" Wahrscheinlichkeitsraum gefunden werden.

Beispiel 2.1 Ein Ehepaar hat zwei Kinder. Wie groß ist die Wahrscheinlichkeit, dass die Kinder unterschiedliches Geschlecht haben.
$\Omega = \{MM, MJ, JM, JJ\}$
$A = \{MJ, JM\} \Rightarrow P(A) = \frac{1}{2}$
Ist der Wahrscheinlichkeitsraum aufgestellt, so müssen zur Bestimmung der Wahrscheinlichkeit P(A) "nur" die Elemente in A (und Ω) gezählt werden.

2.1 Permutationen

Gegeben seien n verschiedene Objekte. Eine Permutation der Objekte ist eine beliebige Anordnung darstellen (in Reihe)

Beispiel 2.2 Gegeben seien a,b,c. Mögliche Permutationen
abc, acb, bac, bca, cab, cba

Lemma 2.1
Die Anzahl der Permutationen von n verschiedenen Objekten ist $n! = n \cdot (n-1) \cdots 1$

Beweis Für den ersten Platz hat man n Möglichkeiten, für den zweiten Platz $(n-1)$ etc. ∎

Bemerkung 2.1 Gegeben seien n Objekte, die nicht alle verschieden sind. Es seien n_1 vom Typ 1, n_2 vom Typ 2, …, n_k vom Typ k
Also: $n_1 + \cdots + n_k = n$

Lemma 2.2 *Die Anzahl der Permutationen von n Objekten mit jeweils* $n_1, n_2, \ldots, n_k$ *gleichen Objekten ist* $\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$

Beispiel 2.3 Eine Schachtel enthält 10 Glühbirnen 5 rote, 2 gelbe, 3 blaue. Die Glühbirnen werden nacheinander zufällig in eine Lichterkette geschraubt. Wie groß ist die Wahrscheinlichkeit, dass zuerst die roten, dann die gelben und zuletzt die blauen Glühbirnen aufgehängt werden?
Antwort: $\frac{5!\cdot 2!\cdot 3!}{10!}$

Beweis (von Lemma 2.2) Zunächst nummerieren wir die gleichen Objekte durch, um sie zu unterscheiden. Dann gibt es nach Lemma 2.1 $n!$ mögliche Permutationen. Zu einer Klasse K_i fassen wir die Permutationen zusammen, bei denen die Elemente vom Typ i die Plätze tauschen.
Da $|K_i| = n_i!$ folgt die Behauptung. ∎

2.2 Urnenmodelle

Gegeben sei eine Urne mit n Objekten, nummeriert mit $1, 2, \ldots, n$. k Objekte werden zufällig gezogen.

A Ziehen mit Zurücklegen, mit Berücksichtigung der Reihenfolge.
Formal können die Ergebnisse dieser Ziehung durch folgende Menge beschrieben werden:
$M_n^k := \{(i_1, \ldots, i_k) | i_\nu \in \{1, \ldots, n\}, \nu = 1, \ldots, k\} = \{1, \ldots, n\}^k$
Die Elemente von M_n^k nennt man k-Permutationen von $\{1, \ldots, n\}$ mit Wiederholung.

Satz 2.3 $|M_n^k| = n^k$

Beispiel 2.4 Wie viele verschiedene 3-stellige Zahlen kann man mit den Ziffern $1, 2, \ldots, 9$ bilden?
Anwort: $9^3 = 729$

B Ziehen ohne Zurücklegen, mit Berücksichtigung der Reihenfolge.
Ergebnisse werden beschrieben durch:
$M_B = \{(i_1, \ldots, i_k) = M_n^k | i_\nu \neq i_\mu \text{ für } \nu \neq \mu\}$
Die Elemente von M_B nennt man die k-Permutationen von $\{1, \ldots, n\}$ ohne Wiederholung.

Satz 2.4 $|M_B| = \frac{n!}{(n-k)!}$

Beweis Beim ersten Zug gibt es n Möglichkeiten.
Beim zweiten Zug gibt es $(n-1)$ Möglichkeiten.
$\vdots$
Beim k-ten Zug gibt es $(n-k+1)$ Möglichkeiten.
Insgesamt also $n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!}$ ∎

Beispiel 2.5 Wieviele 3-stellige Zahlen mit verschiedenen Ziffern $1-9$ gibt es?
Antwort: $9 \cdot 8 \cdot 7 = 504$

C Ziehen ohne Zurücklegen, ohne Berücksichtigung der Reihenfolge.
Ergebnisse werden beschrieben durch:
$M_C = \{(i_1, \dots, i_k) \in M_n^k | i_1 < i_2 < \dots < i_k\}$
Die Elemente von M_C nennt man k-Kombinationen von $\{1, \dots, n\}$ ohne Wiederholung.

Satz 2.5 $|M_C| = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$

Beweis Berücksichtigt man die Reihenfolge, so kann jedes Element aus M_C auf $k!$ verschiedene Arten dargestellt werden. Also gilt der Zusammenhang: $|M_C| \cdot k! = |M_B| \Rightarrow |M_C| = \frac{n!}{k! \cdot (n-k)!}$ ■

Beispiel 2.6 (Lotto: 6 aus 49) Es gibt $\binom{49}{6} = 13983816$ verschiedene Ziehungsergebnisse

D Ziehen mit Zurücklegen ohne Berücksichtigung der Reihenfolge.
Ergebnisse werden beschrieben durch:
$M_D = \{(i_1, \dots, i_k) = M_n^k | i_1 \leq i_2 \leq \dots \leq i_k\}$
Die Elemente von M_D nennt man k-Kombinationen von $\{1, \dots, n\}$ mit Wiederholung.

Satz 2.6 $|M_D| = \binom{n+k-1}{k}$

Beweis Wir betrachten folgende Abbildung f:
Sei $(i_1, \dots, i_k)$ mit $i_1 \leq i_2 \leq \dots \leq i_k$ ein Element aus M_D. Dieses wird abgebildet auf $f((i_1, \dots, i_k)) = (i_1, i_2 + 1, i_3 + 2, \dots, i_k + k - 1) = (j_1, \dots, j_k)$
offenbar gilt $1 \leq j_1 < j_2 < \dots < j_k \leq n + k - 1$
$f : M_D \to \{(j_1, \dots, j_k) \leq M_{n+k-1}^k | j_1 < j_2 < \dots < j_k\} =: M^*$ ist bijektiv da durch $i_\nu = j_\nu - \nu + 1$ die Umkehrabbildung gegeben ist.
Die Anzahl der Elemente in den Mengen M_D und M^* ist also gleich $\Rightarrow$ $|M_D| = \binom{n+k-1}{k}$ ■

Bemerkung 2.2 Für $|M_D|$ gibt es auch eine weitere Interpretation:
$\binom{n+k-1}{k}$ ist die Anzahl der Möglichkeiten k Objekte auf n Fächer aufzuteilen (wobei Mehrfachbelegungen möglich sind).

Beispiel 2.7 Wie viele Möglichkeiten gibt es eine natürliche Zahl k als Summe von n nicht negativen, ganzen Zahlen zu schreiben?
$k = 5, n = 2 \quad \Rightarrow \{(0 + 5), (5 + 0), (1 + 4), (4 + 1), (2 + 3), (3 + 2)\}$
Antwort: $\binom{n+k-1}{k} = \binom{6}{5} = 6$

Zusammenfassung:

Anzahl der Möglichkeiten bei Ziehung vom Umfang k aus $\{1 \dots n\}$	mit Zurücklegen	ohne Zurücklegen
mit Reihenfolge	n^k	$n \cdot (n-1) \cdot \dots \cdot (n-k+1)$ $= \frac{n!}{(n-k)!}$
ohne Reihenfolge	$\binom{n+k-1}{k}$	$\frac{n!}{(n-k)!k!} = \binom{n}{k}$

2.3 Weitere Beispiele

Beispiel 2.8

1. Das Geburtstagsproblem
   Im Hörsaal seien n Studenten. Wie groß ist die Wahrscheinlichkeit, dass mindestens 2 davon am gleichen Tag Geburtstag haben? Wir machen folgende Annahmen:

   - den 29.Februar berücksichtigen wir nicht
   - die Wahrscheinlichkeit an einem bestimmten Tag Geburtstag zu haben ist für alle Tage gleich
   - keine Zwillinge

   Es gilt: $\Omega = \{(i_1, \dots, i_n) | i_\nu \in \{1, \dots, 365\}, \nu = 1, \dots, n\} = \{1, \dots, 365\}^n$
   Also gilt: $|\Omega| = 365^n$
   Sei A das Ereignis, dass mindestens 2 Studenten am gleichen Tag Geburtstag haben. Es gilt $P(A) = \frac{|A|}{|\Omega|} = 1 - \frac{|A^c|}{|\Omega|}$ wobei A^c das Ereignis ist, dass alle Studenten an verschiedenen Tagen Geburtstag haben:
   $A^c = \{(i_1, \dots, i_n) \in \Omega | i_\nu \neq i_\mu \text{ für } \nu \neq \mu\}$
   Somit (Typ B) gilt: $|A^c| = \frac{365!}{(365-n)!}$ und die Wahrscheinlichkeit ist damit $P(A) = 1 - \frac{365!}{(365-n)! \cdot 365^n}$ $(n \leq 365)$
   Für $n = 23$: $P(A) \geq 0,5$
   Für $n = 50$: $P(A) \approx 0,97$
   Offenbar ist die Wahrscheinlichkeit wachsend in n.

2. Das Aufzugsproblem
   Ein Aufzug fährt mit 7 Personen im Erdgeschoss los. Auf der Fahrt zur obersten Etage (5.Stock) steigen alle Fahrgäste aus.

   a) Wie viele Möglichkeiten gibt es die Personen abzusetzen, wenn wir sie nicht unterscheiden wollen?

b) Wie viele Möglichkeiten gibt es die Personen abzusetzen, wenn sie aus 5 Frauen und 2 Männern bestehen und wir Männer und Frauen unterscheiden möchten?

Antwort:

a) Es handelt sich um Typ D. $\Rightarrow$ Wir haben $n = 5$ Stockwerke (=Fächer) auf die wir $k = 7$ Peronen (=Objekte) verteilen.
$\binom{n+k-1}{k}=\binom{5+7-1}{7}=\binom{11}{7}=330$ Möglichkeiten

b) Hier rechnen wir die Möglichkeiten für Männer und Frauen getrennt aus und multiplizieren sie dann.
Frauen: $\binom{5+5-1}{5}=\binom{9}{5}=126$
Männer: $\binom{5+2-1}{2}=\binom{6}{2}=15$
Insgesamt gibt es also $126 \cdot 15 = 1890$ Möglichkeiten.

3. Absolute Permutation
Es sei S_n die Menge aller Permutationen der Zahlen $\{1 \dots n\}$. Eine Permutation heißt absolut, falls sie keine einzige Zahl fest lässt.
Wie groß ist die Wahrscheinlichkeit, dass eine absolute Permutation auftritt, wenn alle Permutationen gleich wahrscheinlich sind?
Es sei Abs die Menge aller absoluten Permutationen und A_k die Menge aller Permutationen, die die Zahl k festhalten, $k = 1 \dots n$.
Dann ist $(Abs)^c = \bigcup_{k=1}^n A_k$.
Es sei

$$\overline{S_k} := \sum_{1 \le i_1 < \dots < i_k \le n} P(A_{i_1} \cap \dots \cap A_{i_k})$$

Mit der Siebformel (Satz 1.3) folgt:

$$P(Abs) = 1 - P(Abs^c) = 1 - \sum_{k=1}^{n} (-1)^{k-1} \cdot \overline{S_n}$$

Die Menge $A_{i_1} \cap \dots \cap A_{i_k}$ ist die Menge aller Permutationen, die die Zahlen $i_1 \dots i_k$ festhalten. Also ist
$P(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(n-k)!}{n!}$
Wichtig: Die letzte Wahrscheinlichkeit hängt nur von k ab, nicht von der konkreten Wahl der i_μ
In diesem Spezialfall gilt mit Typ C (Lotto):

$$\overline{S_k} = \frac{(n-k)!}{n!} \cdot |M_C| = \frac{(n-k)!}{n!} \cdot \binom{n}{k} = \frac{(n-k)!}{n!} \cdot \frac{n!}{k! \cdot (n-k)!} = \frac{1}{k!}$$

Also:

$$P(Abs) = 1 + \sum_{k=1}^{n} (-1)^k \cdot \frac{1}{k!} = 1 + \sum_{k=1}^{n} \frac{(-1)^k}{k!} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} (0! := 1)$$

Insbesondere:

$$\lim_{n \to \infty} P(Abs_{(n)}) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} = e^{-1} = \frac{1}{e}$$

Kapitel 3

Bedingte Wahrscheinlichkeiten und Unabhängigkeiten

Wie können wir Teilinformationen über den Ausgang eines Zufallsexperimentes nutzen?

Beispiel 3.1 Es werden zwei Würfel geworfen. Wir erhalten die Information, dass die Augensumme mindestens 10 ist. Wie groß ist die Wahrscheinlichkeit, dass mind. einer der Würfel 6 zeigt?
Aufgrund der Vorinformation wissen wir, dass das Ergebnis des Experiments in der Menge
$B = \{(4,6), (6,4), (5,5), (5,6), (6,5), (6,6)\}$
liegt. In nur einem Fall (5,5) ist keine 6 dabei.
Wir definieren die folgenden Ereignisse:
A = mindestens einer der Würfel zeigt 6
B = die Augensumme ist mindestens 10
Die gesuchte Wahrscheinlichkeit sollte also wie folgt sein:

$$\frac{|A \cap B|}{|B|} = \frac{5}{6} = \frac{P(A \cap B)}{P(B)}$$

Definition 3.1 *Seien $A, B \in \mathcal{A}$ Ereignisse mit $P(B) > 0$. Dann heißt*

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

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

Bemerkung 3.1 Bezeichnen wir $P_B(A) := P(A|B)$, so ist $(B, \mathcal{A}_B, P_B)$ wieder ein Wahrscheinlichkeitsraum.

Oft ist die bedingte Wahrscheinlichkeit $P(A|B)$ gegeben und man muss $P(A \cap B)$ bestimmen. Also $P(A \cap B) = P(A|B) \cdot P(B)$
Durch vollständige Induktion nach n erhält man:

Satz 3.1 (Multiplikationssatz)
Seien $A_1, \ldots, A_n \in \mathcal{A}$ *Ereignisse mit* $P(A_1 \cap \cdots \cap A_n) > 0$. *Dann gilt*
$P(A_1 \cap \cdots \cap A_n) = \prod_{k=1}^{n} P(A_k|A_1 \cap \cdots \cap A_{k-1})$ *mit* $A_0 = \Omega$
also $P(A_1|A_0) = P(A_1)$.

Beispiel 3.2 Von einem Kartenspiel mit 32 Blatt, wovon 4 Asse sind, bekommt jeder von 3 Spielern 10 Karten. Wie groß ist die Wahrscheinlichkeit, dass jeder der 3 Spieler genau ein As erhält?
Es sei A_i das Ereignis, dass Spieler i genau ein As erhält. $i = 1, 2, 3$
Gesucht ist die Wahrscheinlichkeit $P(A_1 \cap A_2 \cap A_3)$. Es gilt:
$P(A_1 \cdot A_2 \cdot A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_2 \cdot A_1)$
$P(A_1) = \frac{\binom{4}{1} \cdot \binom{28}{9}}{\binom{32}{10}}$, $P(A_2|A_1) = \frac{\binom{3}{1} \cdot \binom{19}{9}}{\binom{22}{10}}$, $P(A_3|A_2 \cdot A_1) = \frac{\binom{2}{1} \cdot \binom{10}{9}}{\binom{12}{10}}$
Satz 3.1 $\Rightarrow P(A_1 \cdot A_2 \cdot A_3) = 0,0556$

Satz 3.2 *Es seien* $B_1, B_2, \ldots \in \mathcal{A}$ *eine* ***Ereignispartition*** *von* Ω, *d.h.*

(i) $B_i \cap B_j = \emptyset$ *für* $i \neq j$

(ii)
$$\sum_{i=1}^{\infty} B_i = \Omega$$

(iii) $P(B_i) > 0 \quad \forall\, i = 1, 2, \ldots$

Dann folgt:

a) ***Satz von der totalen Wahrscheinlichkeit***
Für jedes Ereignis $A \in \mathcal{A}$ *gilt:*
$$P(A) = \sum_{j=1}^{\infty} P(B_j) \cdot P(A|B_j)$$

b) ***Formel von Bayes***
Für jedes Ereignis $A \in \mathcal{A}$ *mit* $P(A) > 0$ *gilt:*
$$P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{j=1}^{\infty} P(B_j) \cdot P(A|B_j)}$$

Beweis a)
$$P(A) = P(A \cap \Omega) \cdot P(A \cap \sum_{j=1}^{\infty} B_j) = P(\sum_{j=1}^{\infty} A \cap B_j) = \sum_{j=1}^{\infty} P(A \cap B_j)$$
$$\overset{\text{Def. bed. W'keit}}{=} \sum_{j=1}^{\infty} P(A|B_j) \cdot P(B_j)$$

b)

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

Einsetzen von a) liefert die Behauptung ■

Beispiel 3.3 Bei einer binären Übertragung von Nachrichten werden durch Störung 5% der gesendeten Nullen zu Einsen und 3% der gesendeten Einsen zu Nullen verfälscht. Das Verhältnis der gesendeten Nullen zu den gesendeten Einsen betrage 3:5. Wie groß ist die Wahrscheinlichkeit, dass eine empfangene Null richtig ist?
Es sei $\Omega = \{(i,j)|i,j \in \{0,1\}\}$
1.Komponente = gesendetes Signal; 2. Komponente = empfangenes Signal
Wir betrachten die folgenden Ereignisse:
$S_0 =$ "eine Null wird gesendet"
$S_1 =$ "eine Eins wird gesendet"
$E_0 =$ "eine Null wird empfangen"
Bekannt sind: $P(E_0|S_0) = 0.95$ $P(E_0|S_1) = 0.03$ $P(S_0) = \frac{3}{8}$ $P(S_1) = \frac{5}{8}$
Außerdem ist $\Omega = S_0 + S_1$. Damit folgt:

$$P(S_0|E_0) = \frac{P(S_0) \cdot P(E_0|S_0)}{P(S_0) \cdot P(E_0|S_0) + P(S_1) \cdot P(E_0|S_1)} = 0.95$$

Falls $P(A|B) = P(A)$, so heißt das, dass "das Eintreten des Ereignisses B keinen Einfluss auf das Eintreten von A hat" Wegen der Definition der bedingten Wahrscheinlichkeit ist dies äquivalent zu $P(A \cap B) = P(A) \cdot P(B)$

Definition 3.2

a) *Zwei Ereignisse $A, B \in \mathcal{A}$ heißen **unabhängig**, falls $P(A \cap B) = P(A) \cdot P(B)$*

b) *Die Ereignisse $A_1, \ldots, A_n \in A$ heißen **unabhängig**, falls für alle $k = 1, \ldots, n$ und für alle k-Tupel, $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ stets gilt:*

$$P(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j})$$

Bemerkung 3.2

a) Teilsysteme von unabhängigen Ereignissen sind unabhänigig

b) Der Begriff der Unabhängigkeit wird auch für unendliche Folgen von Ereignissen benötigt. Man sagt $A_1, A_2, \ldots \in A$ sind unabhängige Ereignisse, falls für jede endliche Teilfolge $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots\}$ die Ereignisse $A_{i_1}, \ldots, A_{i_k}$ unabhängig sind im Sinne von Teil b).

Beispiel 3.4 Eine faire Münze wird $2\times$ geworfen
$\Omega = \{KK, KZ, ZK, ZZ\}$
Es sei $A_1 = \{KK, KZ\}$ "Kopf im ersten Wurf"
$A_2 = \{KK, ZK\}$ "Kopf im zweiten Wurf"
$A_3 = \{KZ, ZK\}$ "Resultate verschieden"
Es gilt: $P(A_1 \cap A_2) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1) \cdot P(A_2)$
$P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$,$P(A_2 \cap A_3) = P(A_2) \cdot P(A_3)$
Aber: $P(A_1 \cap A_2 \cap A_3) = P(\emptyset) = 0 \neq P(A_1) \cdot P(A_2) \cdot P(A_3)$
Damit sind A_1, A_2, A_3 <u>nicht</u> unabhängig.

Kapitel 4

Allgemeine Wahrscheinlichkeitsräume

Ist Ω endlich oder abzählbar unendlich, so kann $\mathcal{A} = \mathcal{P}(\Omega)$ gewählt werden. Was machen wir z.B. bei $\Omega = [0, 1)$ im Beispiel des rotierenden Zeigers 1.1.b)?
$\mathcal{P}([0, 1))$ ist zwar eine σ-Algebra, für eine vernünftige Theorie jedoch zu groß, wie die folgende Überlegung zeigt.
Ist der Zeiger fair, so sollte für $[a, b) \subset [0, 1)$ gelten:
$P([a, b)) = b - a$
Bzw. $\forall A \subset [0, 1)$ und $\forall x \in [0, 1)$:
$P(x + A) = P(A)$ (*)
wobei $x + A = \{x + y \mod 1 | y \in A\}$ ($\Rightarrow P$ ändert sich nicht bei Verschiebung des Intervalls)
P ist also eine Gleichverteilung auf $[0, 1)$. Es gilt jedoch:

Satz 4.1 *Ein Wahrscheinlichkeitsmaß auf (der Potenzmenge) $\mathcal{P}([0, 1))$ mit der Eigenschaft (*) existiert nicht.*

Beweis Betrachte folgende Äquivalenzrelation auf $[0, 1)$: $x \sim y \Leftrightarrow x - y \in \mathbb{Q}$
Die Äquivalenzklassen bilden eine Partition des $[0, 1)$
Auswahlaxiom: Aus jeder Klasse wird ein Element genommen und in eine Menge A gesteckt.
Es gilt nun:

(i) $(x + A) \cap (y + A) = \emptyset \qquad \forall\, x, y \in \mathbb{Q} \cap [0, 1), x \neq y$

(ii) $\bigcup_{x \in \mathbb{Q} \cap [0,1)} (x + A) = [0, 1)$

- zu (i)
  Annahme: $\exists a, b \in A$ und $x, y \in \mathbb{Q} \cap [0, 1), x \neq y$ mit $(a + x) \mod 1 = (b + y) \mod 1$. Da $0 < |x - y| < 1$ folgt $a \neq b$
  Wegen $a - b = y - x (\pm 1) \in \mathbb{Q}$ würde a,b in der gleichen Klasse liegen. Widerspruch.

- zu (ii)
  “ $\subset$ “: ist klar
  “ $\supset$ “: Sei $z \in [0,1) \Rightarrow \exists\, a \in A$ mit $a \sim z$, d.h. $x := z - a \in \mathbb{Q}$ und $-1 < x < 1$
  Falls $x < 0$ ersetze x durch $x + 1$ $(z = (x+1) + a mod 1)$

Sei jetzt P ein Wahrscheinlichkeitsmaß auf $\mathcal{P}([0,1))$ mit (*). Dann gilt:

$$1 \overset{\text{Normiertheit}}{=} P([0,1)) \overset{(i),(ii)}{=} P(\sum_{x\in\mathbb{Q}\cap[0,1)} (x+A)) \overset{\sigma\text{-Add}}{=} \sum_{x\in\mathbb{Q}\cap[0,1)} P(x+A) \overset{(*)}{=} \sum_{x\in\mathbb{Q}\cap[0,1)} P(A)$$

$\Rightarrow$ Widerspruch ■

Definition 4.1
Es sei $\Omega \neq \emptyset$ und $\mathcal{E} \subset \mathcal{P}(\Omega)$. Dann heißt

$$\sigma(\mathcal{E}) := \bigcap_{\mathcal{A}\supset\mathcal{E},\, \mathcal{A}\ \sigma\text{-Algebra}} \mathcal{A}$$

die von $\mathcal{E}$ erzeugte σ-Algebra. $\mathcal{E}$ heißt ***Erzeugendensystem****.*

Bemerkung 4.1

a) $\sigma(\mathcal{E})$ ist die kleinste σ-Algebra, die $\mathcal{E}$ enthält.

b) Der Durchschnitt von beliebig vielen σ-Algebren über Ω ist wieder eine σ-Algebra ($\rightarrow$ *Übung*)

c) $\sigma(\mathcal{E}) \neq \emptyset$, da $\mathcal{P}(\Omega) > \mathcal{E}$ und σ-Algebra ist.

Wir definieren jetzt eine σ-Algebra auf $\mathbb{R}$.

Definition 4.2 *Es sei $\mathcal{E} := \{(a,b], -\infty < a < b < \infty\}$ Dann heißt $\mathfrak{B} = \mathfrak{B}(\mathbb{R}) = \sigma(\mathcal{E})$* ***Borelsche σ-Algebra*** *oder σ-Algebra der Borelschen Mengen von $\mathbb{R}$*

Bemerkung 4.2

a) Es gilt auch
   $\mathfrak{B} = \sigma(\{(-\infty, a], a \in \mathbb{R}\}) = \sigma(\{F \subset \mathbb{R} | F \text{ abgeschlossen}\})$
   $= \sigma(\{U \subset \mathbb{R} | U \text{ offen}\})$
   zur letzten Gleichung: $\mathfrak{B} = \sigma(U \subset \mathbb{R} | U \text{ offen}\})$

   (i) $\mathfrak{B} \supset \sigma(\{U \subset \mathbb{R} | U \text{ offen}\})$, da sei $U \subset \mathbb{R}$ offen $\forall x \in U\ \exists (a,b] \subset U$ mit $x \in (a,b], a, b \in \mathbb{Q}$
   also $U = \bigcup_{\{(a,b)\in\mathbb{Q}^2 | (a,b]\subset U\}} (a,b] \in \mathfrak{B}$

   (ii) “$\subset$“ $(a,b] = \bigcap_{n=1}^{\infty} (a, b + \frac{1}{n}) \in \sigma(\{U \subset \mathbb{R} | U \text{ offen}\})$

b) Sei $A \subset \mathbb{R}, A \neq \emptyset$ Dann ist
$\mathfrak{B}_A := \{B \cap A | B \in \mathfrak{B}\}$ eine σ-Algebra über A

Satz 4.2 *Es gibt ein Wahrscheinlichkeitsmaß P auf $([0,1), \mathfrak{B}_{[0,1)})$ mit der Eigenschaft $P(A) = P(A+x) \forall A \in \mathfrak{B}_{[0,1)}, \forall x \in [0,1)$*
Insbesondere gilt:

$$P([a,b]) = b - a \ \forall 0 \leq a < b < 1$$

Bemerkung 4.3 P heißt Gleichverteilung auf dem Einheitsintervall.
Sei $x \in [0,1)$ Wegen $P(\{x\}) = \lim_{n\to\infty} P([x, x+\frac{1}{n})) = \lim_{n\to\infty} \frac{1}{n} = 0$
gilt $P([a,b]) = P([a,b))$ für $a, b \in [0,1]$
Ist das Wahrscheinlichkeitsmaß aus Satz 4.2 eindeutig?

Definition 4.3 *Sei $\Omega \neq \emptyset$ $\mathcal{D} \subset \mathcal{P}(\Omega)$ heißt **Dynkin-System**, falls gilt:*

(i) $\Omega \in \mathcal{D}$

(ii) $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$

(iii) $A_1, A_2, \ldots \in \mathcal{D}$ *mit* $A_i \cap A_j = \emptyset$ *für* $i \neq j \Rightarrow \sum_{i=1}^{\infty} A_i \in \mathcal{D}$

Bemerkung 4.4 Der Durchschnitt von beliebig vielen Dynkin-Systemen ist wieder ein Dynkin-System. Es sei

$$\delta(\mathcal{E}) := \bigcap_{\mathcal{D} \supset \mathcal{E}, \mathcal{D} Dynkin-System} \mathcal{D}$$

das von $\mathcal{E}$ erzeugte Dynkin-System.

Definition 4.4 *Ein Mengensystem $\mathcal{E}$ heißt **durchschnittsstabil** ($\cap$-stabil), falls $A, B \in \mathcal{E} \Rightarrow A \cap B \in \mathcal{E}$*

Satz 4.3 (Satz über monotone Klassen)
Ist $\mathcal{E}$ ein $\cap$-stabiles Mengensystem, so gilt $\delta(\mathcal{E}) = \sigma(\mathcal{E})$

Satz 4.4 *Sei $\mathcal{A}$ eine σ-Algebra mit $\cap$-stabilem Erzeuger $\mathcal{E}$. Sind P und Q Wahrscheinlichkeitsmaße auf $\mathcal{A}$ mit der Eigenschaft $P(E) = Q(E) \ \forall E \in \mathcal{E}$,*
so gilt: $P(A) = Q(A) \ \forall A \in \mathcal{A}$.

Beweis Sei $\mathcal{D} := \{A \in \mathcal{A} | P(A) = Q(A)\}$ Nach Voraussetzung ist $\mathcal{E} \subset \mathcal{D}$. Wegen den Eigenschaften von Wahrscheinlichkeitsmaßen ist $\mathcal{D}$ ein Dynkin-System.
Satz 4.3 $\mathcal{D} \supset \delta(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}$ ■

Bemerkung 4.5 $\mathcal{E} := \{[a,b) | 0 \leq a < b < 1\}$ ist ein Erzeugendensystem von $\mathfrak{B}_{[0,1)}$.
Offenbar ist $\mathcal{E}$ durchschnittsstabil.
Also ist P aus Satz 4.3 eindeutig.

Kapitel 5

Zufallsvariable, Verteilung, Verteilungsfunktion

5.1 Zufallsvariable

Sei $(\Omega, \mathcal{A}, P)$ ein beliebiger Wahrscheinlichkeitsraum. Häufig interessiert nicht ω selbst, sondern eine Kennzahl $X(\omega)$, d.h. wir betrachten eine Abbildung $\omega \mapsto X(\omega)$

Beispiel 5.1 $2\times$ würfeln
$\Omega = \{(i,j)|i,j \in \{1,2\ldots,6\}\}$
$X(\omega) = X((i,j)) = i+j$ "Augensumme"

Sei $X : \Omega \to \mathbb{R}$ Wir möchten jetzt dem Ereignis $X \in B = \{\omega \in \Omega | X(\omega) \in B\}, B \subset \mathbb{R}$ eine Wahrscheinlichkeit zuordnen.
Also muss $\{\omega | X(\omega) \in B\} \in \mathcal{A}$ sein.

Definition 5.1 *Sei $(\Omega, \mathcal{A})$ ein beliebiger Messraum. Eine Abbildung $X : \Omega \to \mathbb{R}$ heißt Zufallsvariable, falls*
$X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\} \in \mathcal{A}, \quad \forall\, B \in \mathfrak{B}$
Diese Bedingung nennt man auch $(\mathcal{A}, \mathfrak{B})$-Messbarkeit von X.

Satz 5.1
*Eine Abbildung $X : \Omega \to \mathbb{R}$ ist genau dann **Zufallsvariable**, wenn*

$$X^{-1}((-\infty, a]) = \{\omega | X(\omega) \leq a\} =: \{X \in B\} =: (X \in B) \in \mathcal{A}, \qquad \forall\, a \in \mathbb{R}$$

Beweis "$\Rightarrow$" Sei X Zufallsvariable. $(-\infty, a] \in \mathfrak{B} \Rightarrow$ Behauptung

"$\Leftarrow$" $X^{-1}((-\infty, a]) \in \mathcal{A}$
Definiere $\mathcal{A}_0 = \{B \subset \mathbb{R} | X^{-1}(B) \in \mathcal{A}\}$. $\mathcal{A}_0$ ist eine σ-Algebra über $\mathbb{R}$:

(i) $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{A} \Rightarrow \mathbb{R} \in \mathcal{A}_0$

(ii) $X^{-1}(B^c) = \{\omega | X(\omega) \notin B\} = \{\omega | X(\omega) \in B\}^c = (X^{-1}(B))^c$
Also $B \subset \mathcal{A}_0 \Rightarrow B^c \subset \mathcal{A}_0$

(iii)

$$X^{-1}(\bigcup_{n=1}^{\infty} B_n) = \bigcup_{n=1}^{\infty} X^{-1}(B_n)$$

Nach Voraussetzung ist $\mathcal{E} = \{(-\infty, a], a \in \mathbb{R}\} \subset \mathcal{A}_0 \Rightarrow \mathcal{A}_0 \supset \sigma(\mathcal{E}) = \mathfrak{B} \Rightarrow$ Behauptung ■

Bemerkung 5.1

a) Satz 5.1 bleibt richtig, wenn wir $\mathcal{E} = \{(-\infty, a], a \in \mathbb{R}\}$ durch ein anderes Erzeugendensystem von $\mathfrak{B}$ ersetzen.

b) Bei Anwendungen ist oft $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathfrak{B})$

Wann ist eine Abbildung $X : \mathbb{R} \to \mathbb{R}$ messbar (ZV)?

Satz 5.2 *Sei $(\mathbb{R}, \mathfrak{B})$ gegeben, $X : \mathbb{R} \to \mathbb{R}$ ist z.B. messbar, falls X stetig oder (schwach) monoton wachsend oder fallend.*

Beweis Sei X stetig. Dann ist $X^{-1}(U)$ offen, falls U offen.
Sei X wachsend $\Rightarrow \{\omega \in \mathbb{R} | X(\omega) \leq a\}$ ist von der Gestalt $(-\infty, b) \in \mathfrak{B}$ oder $(-\infty, b] \in \mathfrak{B}$ ■

Bemerkung 5.2 Es sei $X(\omega) = c\ \forall \omega \in \Omega$, $c \in \mathbb{R}$. Dann ist X eine Zufallsvariable.

Satz 5.3 *Seien $X : \Omega \to \mathbb{R}$ und $Y : \mathbb{R} \to \mathbb{R}$ Zufallsvariablen, dann ist $Y \circ X : \Omega \to \mathbb{R}$ wieder eine Zufallsvariable.*

Satz 5.4 *Sei $(\Omega, \mathcal{A})$ ein Messraum und X, Y Zufallsvariablen darauf.*

a) $\{X < Y\} = \{\omega \in \Omega | X(\omega) < Y(\omega)\}$,$\{X \leq Y\}, \{X = Y\} \in \mathcal{A}$

b) Sind $\alpha, \beta \in \mathbb{R}$, so sind
$\alpha X + \beta, X + Y, X \cdot Y, X \wedge Y = \min\{X, Y\}, X \vee Y = \max\{X, Y\}$
Zufallsvariablen

c) Ist $(X_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen, so sind auch
$\sup_{n \in \mathbb{N}} X_n, \inf_{n \in \mathbb{N}} X_n, \limsup_{n \to \infty} X_n, \liminf_{n \to \infty} X_n$
Zufallsvariablen, falls sie $\mathbb{R}$-wertig sind.
Gilt $X_n(\omega) \to X(\omega) \forall \omega \in \Omega$, so ist auch X eine Zufallsvariable.

Beweis a) $\{X < Y\} = \bigcup_{q \in \mathbb{Q}} \{\underbrace{X < q}_{\in \mathcal{A}}\} \cap \{\underbrace{Y > q}_{\in \mathcal{A}}\}$

$\{X \leq Y\} = \{X > Y\}^c \in \mathcal{A}$,$\{X = Y\} = \{X \leq Y\} \cap \{X \geq Y\} \in \mathcal{A}$

b) (i) $x \mapsto \alpha x + \beta$ ist stetig

(ii) $\{X+Y \leq a\} = \{X \leq a-Y\} = \{X \leq a-Y\} \in \mathcal{A} \forall a \in \mathbb{R}$, da $a-Y$ Zufallsvariable + Teil a)

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

(iv) $\{X \vee Y \leq a\} = \{X \leq a\} \cap \{Y \leq a\} \in \mathcal{A}$

c) $\{\sup_{n\in\mathbb{N}} X_n \leq a\} = \bigcap_{n=1}^{\infty}\{X_n \leq a\} \in \mathcal{A}$
$\inf_{n\in\mathbb{N}} X_n = -\sup_{n\in\mathbb{N}}(-X_n)$
$\limsup_{n\to\infty} X_n = \inf_{n\in\mathbb{N}} \sup_{m\geq n} X_m$
$\liminf_{n\to\infty} X_n = \sup_{n\in\mathbb{N}} \inf_{m\geq n} X_m$
Im Falle der Konvergenz ist $X = \limsup_{n\to\infty} X_n$ ∎

Bemerkung 5.3 Teil c) ist ohne Einschränkung gültig, wenn man $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$ betrachtet und $\mathfrak{B}$ zu $\mathfrak{B}(\overline{\mathbb{R}}) = \sigma(\mathfrak{B} \cup \{\{-\infty\}, \{+\infty\}\})$ erweitert.

5.2 Verteilungen

Definition 5.2
*Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Die **Verteilung** der Zufallsvariablen ist die Mengenfunktion $P_X : \mathfrak{B} \to [0,1]$ mit*
$P_X(B) = P(\{\omega \in \Omega | X(\omega) \in B\}) \quad \forall B \in \mathfrak{B}$

Bemerkung 5.4

a) P_X ist ein Wahrscheinlichkeitsmaß auf dem Messraum $(\mathbb{R}, \mathfrak{B})$, denn:

- $P_X(\mathbb{R}) = P(\Omega) = 1$ (Normiertheit)
- Für $B_1, B_2, \ldots \in \mathfrak{B}$ paarweise disjunkt gilt: (σ-Additivität)

$$P_X(\sum_{i=1}^{\infty} B_i) = P(X^{-1}(\sum_{i=1}^{\infty} B_i)) = P(\sum_{i=1}^{\infty} X^{-1}(B_i)) = \sum_{i=1}^{\infty} P(X^{-1}(B_i)) = \sum_{i=1}^{\infty} P_X(B_i)$$

b) Die Abbildung $P \to P_X$ nennt man **Maßtransport** vom Messraum $(\Omega, \mathcal{A})$ in den Messraum $(\mathbb{R}, \mathfrak{B})$

5.3 Verteilungsfunktion

Eine Verteilung $P_X : \mathfrak{B} \to [0,1]$ kann durch eine "einfachere" Funktion $F_X : \mathbb{R} \to [0,1]$ beschrieben werden.

Definition 5.3 *Es sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Die Funktion $F_X : \mathbb{R} \to [0,1]$ mit*
*$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega | X(\omega) \leq x\}) = P_X((-\infty, x])$ heißt **Verteilungsfunktion** von X.*

Bemerkung 5.5 Da die Mengen $(-\infty, x], x \in \mathbb{R}$ einen $\cap$-stabilen Erzeuger von $\mathfrak{B}$ bilden, wird P_X durch F_X eindeutig festgelegt (siehe Satz 4.4)

Satz 5.5
Sei $X : \Omega \to \mathbb{R}$ eine Zufallsvariable und $F_X : \mathbb{R} \to [0,1]$ ihre Verteilungsfunktion. Dann gilt:

a)

$$\lim_{x\to-\infty} F_X(x) = 0, \qquad \lim_{x\to\infty} F_X(x) = 1$$

b) F_X ist (schwach) monoton wachsend.

c) F_X ist rechtsseitig stetig.

Beweis b) folgt aus der Monotonie von P_X

a) Sei (x_n) eine reellwertige Folge mit $\lim_{n\to\infty} x_n = -\infty$
Setze $y_n := \sup_{m\geq n} x_m$ Dann gilt $y_n \downarrow -\infty$, also $(-\infty, y_n] \downarrow \emptyset$
Da P_X stetig in $\emptyset$ ist (Satz 1.4) folgt:
$0 \leq F_X(x_n) = P_X((-\infty, x_n]) \leq P_X((-\infty, y_n]) \to 0$ für $n \to \infty$
Andere Grenzwertaussage mit Stetigkeit von unten von P_X

c) Sei $x \in \mathbb{R}, x_n \geq x \,\forall n \in \mathbb{N}$ und $\lim_{x\to\infty} x_n = x$
Setze $y_n = \sup_{m\geq n} x_m$, also $y_n \downarrow x$ und
$F_X(x) = P_X((-\infty, x]) \leq P_X((-\infty, x_n]) = F_X(x_n) \leq$
$\leq P_X((-\infty, y_n]) \overset{n\to\infty}{\longrightarrow} P_X((-\infty, x]) = F_X(x)$
weil P_X stetig von oben.

Umgekehrt gibt es zu jeder Funktion $F : \mathbb{R} \to [0,1]$ mit den Eigenschaften a),b),c) aus Satz 5.5 eine Zufallsvariable X, so dass $F = F_X$. ∎

Definition 5.4
*Es sei $F : \mathbb{R} \to [0,1]$ eine Funktion mit den Eigenschaften a),b),c) aus Satz 5.5. Die **Quantilfunktion** F^{-1} zu F ist:*

$$F^{-1}(0,1) \to \mathbb{R}, \quad F^{-1}(y) = \inf\{x \in \mathbb{R} | F(x) > y\}$$

Bemerkung 5.6

a) Ist F stetig und streng monoton wachsend, so ist F^{-1} die übliche Umkehrfunktion.

b) Für $0 < \alpha < 1$ heißt $F_X^{-1}(x)$ α-Quantil zu X

Lemma 5.6

$$y \leq F(x) \Leftrightarrow F^{-1}(y) \leq x, \quad \forall\, y \in (0,1), x \in \mathbb{R}$$

Beweis "⇒" Definition von F^{-1}

"⇐" $F(x) < y \Rightarrow F(x + \frac{1}{n}) < y$ für ein $n \in \mathbb{N}$ (F ist rechtsseitig stetig)

$\Rightarrow F^{-1}(y) \geq x + \frac{1}{n}$ (F monoton wachsend)

$\Rightarrow F^{-1}(y) > x$ ∎

Satz 5.7
Es sei $F : \mathbb{R} \to [0,1]$ eine Funktion mit den Eigenschaften a),b),c) aus Satz 5.5. Dann gibt es einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ und eine Zufallsvariable $X : \Omega \to \mathbb{R}$ mit Verteilungsfunktion F.

Beweis Wähle $\Omega = [0,1), \mathcal{A} = \mathfrak{B}_{[0,1)}, P = \mathrm{Unif}[0,1)$ (Gleichverteilung).
Definiere $X : \Omega \to \mathbb{R}$ durch $X(\omega) := F^{-1}(\omega)$
Offenbar ist F^{-1} monoton wachsend, also X eine Zufallsvariable und
$F_X(x) = P(X < x) = P(\{\omega \in \Omega | F^{-1}(\omega) \leq x\}) \overset{L.5.6}{=} P(\{\omega \in \Omega | \omega \leq F(x)\}) = P([0, F(x)]) = F(x)$ ∎

Kapitel 6

Einige Verteilungen

Im folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum

Definition 6.1 *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ bzw. ihre Verteilung heißen* ***diskret*** *falls es eine endliche oder abzählbare Menge $C \subset \mathbb{R}$ gibt, so dass $P(X \in C) = 1$. O.B.d.A. sei $C = \{x_1, x_2, \ldots\}$ (x verschieden). Die Folge $\{p_X(k)\}_{k\in\mathbb{N}}$ mit $p_X(k) = P(X = x_k)$ heißt* ***Zähldichte*** *(oder Wahrscheinlichkeitsfunktion) von X*

Bemerkung 6.1

a) Für eine Zähldichte $\{p_X(k)\}_{k\in\mathbb{N}}$ gilt $p_X(k) \geq 0 \quad \forall k \in \mathbb{N}$ und $\sum_{k=1}^{\infty} p_X(k) = 1$

b) Die Verteilung von X wird durch die Zähldichte bestimmt, denn $\forall B \in \mathfrak{B}$ gilt:

$$P_X(B) = P_X(B \cap C) = P_X(\sum_{k|x_k \in B} \{x_k\}) = \sum_{k|x_k \in B} p_X(k)$$

6.1 Wichtige diskrete Verteilungen

6.1.1 Binomialverteilungen

Eine diskrete Zufallsvariable X heißt **binomialverteilt**, mit Parameter $n \in \mathbb{N}$ und $p \in [0, 1]$ (kurz $X \sim B(n, p)$) falls
$p_X(k) = P(X = k)=\binom{n}{k}\, p^k(1-p)^{n-k}$ für $k = 0, 1, \ldots, n$

Beispiel 6.1 Eine Münze wird n-mal geworfen
$\Omega = \{(\omega_1, \ldots, \omega_n)|\omega_i \in \{K, Z\}, i = 1, \ldots, n\} = \{K, Z\}^n$
Es sei $X : \Omega \to \{0, 1, 2, \ldots, n\}$ mit $X(\omega) = \sum_{i=1}^{n} 1_{\{\omega_i = K\}}$ die Anzahl der Kopf-Würfe in der Folge. Weiter seien die Ereignisse $A_i = \{\omega|\omega_i = K\}$={i-ter Wurf

Kopf}, $i = 1, \ldots, n$ unabhängig und $P(A_i) = p, i = 1, \ldots, n$ Dann gilt:

$$\begin{aligned}
P(X=k) &= P(\sum_{1 \le i_1 < \cdots < i_k \le n} A_{i_1} \cap \cdots \cap A_{i_k} \bigcap_{j \notin \{i_1, \ldots, i_k\}} A_j^c) \\
&= \sum_{1 \le i_1 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k} \bigcap_{j \notin \{i_1, \ldots, i_k\}} A_j^c) \\
&= \sum_{1 \le i_1 < \cdots < i_k \le n} \underbrace{\underbrace{P(A_{i_1})}_{=p} \cdot \cdots \cdot \underbrace{P(A_{i_k})}_{=p}}_{=p^k} \cdot \underbrace{\prod_{j \notin \{i_1, \ldots, i_k\}} \underbrace{P(A_j^c)}_{=1-p}}_{=(1-p)^{n-k}} \\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned}$$

Die Verteilungsfunktion F_X ist hier

$$F_X(x) = \sum_{k \le x} \binom{n}{k} p^k (1-p)^{n-k}, x \in \mathbb{R}$$

hier fehlt ein Bild

Es gilt $P(X = x) = F_X(x) - F_X(x^-), x \in \mathbb{R}$

6.1.2 Hypergeometrische Verteilung

Eine Urne enthält r rote und s schwarze Kugeln, $r + s = n$. Es werden m Kugeln ohne Zurücklegen gezogen. $X(\omega)$ sei die Anzahl der gezogenen roten Kugeln. Dann gilt:

$$P(X = k) = \frac{\binom{r}{k}\binom{n-r}{m-k}}{\binom{n}{m}}$$

X nimmt die Werte $k = \max\{0, m - s\}, \ldots, \min\{r, m\}$ an und X heißt **hypergeometrisch verteilt** mit Parameter $r, n, m \in \mathbb{N}$

6.1.3 Geometrische Verteilung

Eine diskrete Zufallsvariable X heißt **geometrisch verteilt** mit Parameter $p \in (0, 1)$, falls
$P(X = k) = (1 - p)^{k-1}p$ für $k = 1, 2, 3, \ldots$

Beispiel 6.2 Wir würfeln bis erstmals eine 6 fällt. $X(\omega)$ sei die Anzahl der benötigten Würfe. Dann gilt:

$$P(X = k) = P(\text{Wurf 1 bis } k-1 \text{ keine 6, dann 6}) = \frac{5^{k-1} \cdot 1}{6^k} = \frac{1}{6}(\frac{5}{6})^{k-1}$$

6.1.4 Poisson-Verteilung

Eine diskrete Zufallsvariable X heißt **Poisson-verteilt** mit Parameter $\lambda > 0$ wenn: $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$ für $k = 0, 1, 2 \dots$
Die Poisson-Verteilung kann man auffassen als Approximation der Binomialverteilung bei großem n und kleinem p. Es gilt:

Satz 6.1 *Sei $\lambda > 0$ und $p_n := \frac{\lambda}{n} < 1$. Dann gilt:*

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

Beweis

$$\begin{aligned}
\binom{n}{k} p_n^k (1-p_n)^{n-k} \quad &= \quad \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \quad \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n!}{(n-k)!n^k}}_{\to 1} \cdot \frac{\overbrace{(1-\frac{\lambda}{n})^n}^{\to e^{-\lambda}}}{\underbrace{(1-\frac{\lambda}{n})^k}_{\to 1}} \\
&\xrightarrow{n\to\infty} \quad \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

■

Wichtiges Beispiel:
Eine Versicherung hat ein Portfolio von n Risiken (n groß). Die Wahrscheinlichkeit, dass ein Risiko in einem bestimmten Zeitraum einen Schaden liefert sei $p_n = \frac{\lambda}{n}$. Dann ist X = Anzahl der Risiken, die einen Schaden liefern $\sim B(n, p_n)$, also X in etwa Poisson-verteilt.

6.1.5 Diskrete Gleichverteilung

Eine diskrete Zuvallsvariable X heißt **gleichverteilt** auf $\{x_1, \dots, x_m\} \subset \mathbb{R}$, falls: $P(X = x_i) = \frac{1}{m}$ für $i = 1, \dots, m$

6.2 Wichtige stetige Verteilungen

Definition 6.2 *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ bzw. ihre Verteilung heißen **absolutstetig**, falls die Verteilungsfunktion F_X von X die folgende Darstellung besitzt:*

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

*wobei $f_X : \mathbb{R} \to [0, \infty)$ die **Dichte** von X ist.*

Bemerkung 6.2

a) Jede integrierbare Funktion $f_X : \mathbb{R} \to [0, \infty)$ mit der Eigenschaft

$$\int_{-\infty}^{\infty} f_X(y)dy = 1 \quad \text{definiert durch} \quad F_X(x) = \int_{-\infty}^{x} f_X(y)dy$$

ist eine Verteilungsfunktion.

b) Die Dichte ist das stetige Analogon zur Zähldichte. Es gilt:
$P_X(B) = P(x \in B) = \int_B f_X(y)dy \qquad \forall\, B \in \mathfrak{B}$
da P_X eine Verteilung ist (nachrechnen!) und auf $\{(-\infty, b], b \in \mathbb{R}\}$ mit F_X übereinstimmt.
f_X kann aber Werte größer als 1 annehmen.

c) Bei einer absolutstetigen Zufallsvariable gilt $P(X = x) = 0 \quad \forall\, x \in \mathbb{R}$ und $F_X(x)$ ist stetig. Ist $f_X(x)$ stetig, so ist F_X differenzierbar und es gilt: $F'_X(x) = f_X(x)$. Im Allgemeinen ist F_X aber <u>nicht</u> differenzierbar.

6.2.1 Gleichverteilung

Eine Zufallsvariable X heißt **gleichverteilt** auf dem Intervall (a, b), $a < b$ (Schreibweise: $X \sim U(a, b)$ bzw. $\mathrm{Unif}(a, b)$), falls die Dichte von X gegeben ist durch:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{für } a < x < b \\ 0, & \text{sonst} \end{cases}$$

Die Verteilungsfunktion ist gegeben durch:

$$F_X(x) = \begin{cases} 0, & \text{falls } x \leq a \\ \int_a^x \frac{1}{b-a} dy = \frac{x-a}{b-a}, & \text{falls } x \leq b \\ 1, & \text{falls } x \geq b \end{cases}$$

6.2.2 Exponentialverteilt

Eine Zufallsvariable X heißt **exponentialverteilt** mit Parameter $\lambda > 0\,(X \sim \exp(\lambda))$, falls die Dichte von X gegeben ist durch:

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

Für die Verteilungsfunktion gilt:

$$F_X(x) = \begin{cases} 0, & x \leq 0 \\ \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

Die Exponentialverteilung wird oft zur Beschreibung von Lebens- oder Zeitdauern verwendet und besitzt die Eigenschaft der „Gedächtnislosigkeit“, d.h. für zwei Zeitpunkte $0 < s < t$ gilt:

$$
\begin{aligned}
P(X \geq t | X \geq s) = \frac{P(X \geq t, X \geq s)}{P(X \geq s)} &= \frac{P(X \geq t)}{P(X \geq s)} \\
&= \frac{1 - F_X(t)}{1 - F_X(s)} = \frac{e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda(t-s)} = P(X \geq t - s)
\end{aligned}
$$

6.2.3 Normalverteilung

Eine Zufallsvariable X heißt **normalverteilt** mit Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ (Schreibweise $X \sim \mathcal{N}(\mu, \sigma^2)$) falls die Dichte von X gegeben ist durch:

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) \quad , x \in \mathbb{R} \\
&=: \varphi_{\mu,\sigma^2}(x)
\end{aligned}
$$

Ist $\mu = 0$ und $\sigma^2 = 1$ so nennt man X **standard normalverteilt**. Die Verteilungsfunktion wird hier häufig mit Φ bezeichnet, also:

$$
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)\, dy
$$

Lemma 6.2 *Es gilt:*

a) $\lim_{x\to\infty} \Phi(x) = 1$

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

c) $X \sim N(\mu, \sigma^2),\ a \neq 0,\ b \in \mathbb{R} \to aX + b \sim N(a\mu + b, a^2\sigma^2)$

Beweis a) ($\frac{1}{\sqrt{2\pi}}$ ist Konstante und wird hier weggelassen; Trick: wir quadrieren)

$$
(\int_{-\infty}^{\infty} \exp(-\frac{1}{2}y^2)\, dy)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-\frac{1}{2}(x^2+y^2))\, dxdy = \int_0^{2\pi}\int_0^{\infty} re^{-\frac{r^2}{2}}\, drd\varphi =
$$

$$
= \int_0^{2\pi} -e^{\frac{r^2}{2}}\Big|_0^{\infty} d\varphi = 2\pi
$$

b) es gilt $\varphi_{0,1}(x) = \varphi_{0,1}(-x)$

c) sei $Y = aX + b$. Es gilt:

$$
\begin{aligned}
P(Y \leq y) = P(X \leq \frac{y-b}{a}) &= \int_{-\infty}^{\frac{y-b}{a}} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})\, dx \\
&= \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(x'-b-a\mu)^2}{a^2\sigma^2})\frac{1}{a}\, dx' \\
&\Rightarrow \ Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)
\end{aligned}
$$

■

Kapitel 7

Erwartungswert und Varianz von Zufallsvariablen

Im Folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum. Der Erwartungswert von X ist ein Lebesgue-Integral (allerdings allgemeiner als in Analysis II). Zunächst wird der Erwartungswert für sogenannte Elementare Zufallsvariablen definiert.

Definition 7.1 *Eine Zufallsvariable $X : \Omega \to \mathbb{R}$ heißt **elementar**, falls sie eine Darstellung*

$$X(\omega) = \sum_{i=1}^{m} \alpha_i \cdot 1_{A_i}(\omega)$$

besitzt, mit $A_i \in \mathcal{A}$, $\alpha_i \in \mathbb{R}_+$,$m \in \mathbb{N}$. M^E sei die Menge aller elementaren Zufallsvariablen auf dem Warscheinlichkeitsraum.
Für $X \in M^E$ sei das Integral von X bezüglich P definiert durch $\int XdP := \int_{i=1}^{n} \alpha_i P(A_i)$

Bemerkung 7.1 a) $\int XdP$ ist unabhängig von der gewählten Darstellung von X (vgl. Analysis II)

b) Sei X eine diskrete Zufallsvariable. Wir führen das Zufallsexperiment n-mal druch (n groß). Welchen Wert erhält man im Mittel für X? Der Wert x_k tritt bei dem Experiment n_k-mal auf ($\sum_{k=0}^{\infty} n_k = n$). Mittelwert: $\frac{1}{n} \sum_{k=0}^{\infty} n_k x_k$

Jetzt wird der Integralbegriff erweitert. Sei $M^+ := \{X : \Omega- > \mathbb{R}_+ | X \text{ ist Zufallsvariable}\}$. Für $X \in M^+$ betrachte die Folge $(X_n)_{n\in\mathbb{N}}$ mit

$$X_n := \sum_{i=0}^{n-2^n} \frac{i}{2^n} 1_{A_i^n} \text{ mit } A_i^n = \begin{cases} \{\frac{i}{2^n} \leq X \leq \frac{i+1}{2^n}\} & \text{, falls } i = 0, 1, \ldots, n \cdot 2^n - 1 \\ \{X \geq 1\} & \text{, falls } i = n \cdot 2^n \end{cases}$$

Offenbar ist $X_n \in M^E$ und $x_n(\omega) \leq x_{n+1}(\omega)\ \forall \omega \in \Omega$. Außerdem gilt $X_n(\omega) \uparrow X(\omega)$ punktweise $\forall \omega \in \Omega$.

$$\int XdP := \lim_{n\to\infty} \int X_n dP\,.$$

Bemerkung 7.2 a) Der Grenzwert existiert wegen der Monotonie

b) Der Grenzwert ist unabhängig von der gewählten Folge $(X_n)_{x\in\mathbb{N}}$: Sei $(Y_n)_{n\in\mathbb{N}}$ eine weitere Folge elementarer Zufallsvariablen, die monoton wachsend gegen X konvergiert, so gilt $\lim_{n\to\infty}\int X_n dP = \lim_{n\to\infty} Y_n dP$ (vergleiche Analysis II).

Für eine beliebige Zufallsvariable $X:\Omega\to\mathbb{R}$ gilt: $X = X^+ - X^-$ wobei $X^+ = \max\{X,0\}$ und $X^- = -\min\{X,0\}$, also $X^+, X^- \in M^+$.
Wir definieren durch

$$\int X dP = \int X^+ dP - \int X^- dP =: \int_\Omega X(\omega)dP(\omega) =: EX$$

den **Erwartungswert** von X. X heißt integrierbar, falls $\int X^+ dP < \infty$ und $\int X^- dP < \infty$, d.h. wenn $\int |X| dP < \infty$

Bemerkung 7.3 a) Für $A\in\mathcal{A}$ sei $\int X dP := \int_\Omega X 1_A dP$

b) In Stochastik II wird das Thema weiter vertieft.

Satz 7.1
Es seien X, Y Zufallsvariablen mit existierendem Erwartungswert und $a, b\in\mathbb{R}$

a) Dann existiert auch $E(aX+bY)$ und es gilt:

$$E(aX+bY) = aEX + bEY \qquad \text{„Linearität“}$$

b) Gilt $X\le Y$, d.h. $X(\omega)\le Y(\omega) \quad \forall\,\omega\in\Omega$, so folgt:

$$EX \le EY \qquad \text{„Monotonie“}$$

Beweis vgl. Analysis II ■

Satz 7.2 *Sei $(\Omega,\mathcal{A},P)$ ein Wahrscheinlichkeitsraum und $X:\Omega\to\mathbb{R}$ eine Zufallsvariable mit Verteiltung P_X. $g:\mathbb{R}\to\mathbb{R}$ sei messbar (Zufallsvariable). Dann ist (im Falle der Existenz):*

$$Eg(X) = \int_\Omega g(X)dP = \int_\mathbb{R} g dP_X$$

Beweis Sei nunächst $g\in M^E$, also $g(\omega) = \sum_{i=0}^m \alpha_i 1_{B_i}(\omega)$ für $m\in\mathbb{N}, \alpha\in\mathbb{R}_+, B_i\in\mathfrak{B}$ somit $g(X) = \sum_{i=1}^m \alpha_i\cdot 1_{A_i}$, $A_i = X^{-1}(B_i)$ und $\int_\Omega g(x)dP = \sum_{i=1}^m \alpha_i\cdot P(A_i) = \sum_{i=1}^m \alpha_i P_X(B_i) = \int_\mathbb{R} g dP_X$.
Falls $g\ge 0$, wähle $\{g_n\}\subset M^E$ mit $g_n\uparrow g$. Die Gleichung gilt für jedes g_n, Grenzübergang liefert die Gleichheit für g. Falls g beliebig, betrachte $g = g^+ - g^- \Rightarrow$ Behauptung. ■

Wir unterscheiden jetzt die beiden Fälle dass X diskret bzw. absolutstetig ist. Hier ergeben sich relativ einfache Formeln.

Satz 7.3 *Sei X eine diskrete Zufallsvariable mit Werten $x_0, x_1, x_2, \dots$ und Zähldichte $\{P_X^{(k)}\}_{k\in\mathbb{N}_0}$. $g : \mathbb{R} \to \mathbb{R}$ sei messbar. Dann existiert $Eg(X)$, falls $\sum_{k=0}^{\infty} |g(x_k)|P_X(k) < \infty$ und es gilt:*

$$Eg(X) = \sum_{k=0}^{\infty} g(x_k)P_X(k)$$

Beweis Sei zunächst $g \in M^E$, also $g = \sum_{i=1}^{m} \alpha_i 1_{B_i}$ für $m \in \mathbb{N}, \alpha_i \in \mathbb{R}_+, B_i \in \mathfrak{B}$. Es gilt (vgl. Beweis vorher): $Eg(X) = \sum_{i=0}^{m} \alpha_i P_X(B_i) = \sum_{i=1}^{m} \alpha_i \left(\sum_{x_k \in B_i} P_X(k)\right) = \sum_{i=1}^{m} \alpha_i \sum_{k=0}^{\infty} 1_{B_i}(x_k)P_X(k) = \sum_{k=0}^{\infty} \underbrace{\sum_{i=1}^{m} \alpha_i 1_{B_i}(x_k)P_X(k)}_{=g(x_k)} = \sum_{k=0}^{\infty} g(x_k)P_X(k)$. Allgemeines g wie im Beweis von Satz 7.2 ∎

Beispiel 7.1 Sei $X \sim B(n,p)$ (binomialverteilt). Dann gilt:

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

Also folgt:

$$EX = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} =$$

$$= np \underbrace{\sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}}_{=(p+(1-p))^{n-1}=1} = np$$

Satz 7.4 *Sei nun $X : \Omega \to \mathbb{R}$ eine absolut stetige Zufallsvariable mit Dichte f_X. $g : \mathbb{R} \to \mathbb{R}$ sei messbar. Dann existiert $Eg(X)$, falls $\int_{-\infty}^{\infty} |g(x)| f_X(x) dx < \infty$ und es gilt:*

$$Eg(X) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Beweis ähnlich wie in Satz 7.3 ∎

Beispiel 7.2 Sei $X \sim N(\mu, \sigma^2)$ (X normalverteilt). Also ist

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$$

Es folgt:

$$\begin{aligned} EX &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x \exp(-\frac{(x-\mu)^2}{2\sigma^2})\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma u + \mu) \exp(-\frac{1}{2}u^2)\, du \\ &= \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma u \exp(-\frac{1}{2}u^2)\, du}_{=0 \text{ wg. Symmetrie}} + \mu \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-\frac{1}{2}u^2)\, du}_{=1, \text{ da Dichte}} = \mu \end{aligned}$$

Definition 7.2 *Sei X eine Zufallsvariable*

*a) Ist $k \in \mathbb{N}$ und existiert $E|X|^k$, dann heißt EX^k, **k-tes Moment von X** und $E(X-EX)^k$, **k-tes zentriertes Moment von X***

*b) Das zweite zentrierte Moment heißt auch **Varianz** von X.*
Wir schreiben: $\operatorname{Var}(X) = E(X-EX)^2$
$\sigma(X) := \sqrt{\operatorname{Var}(X)}$ *heißt **Standardabweichung***

Bemerkung 7.4 Die Varianz misst die mittlere quadratische Abweichung der Zufallsvariable X von ihrem Mittelwert. $\sigma(X)$ hat die gleiche Dimension wie X.

Satz 7.5 *Sei X eine Zufallsvariable. Falls die entsprechenden Größen existieren, gilt:*

a) $\operatorname{Var}(X) = EX^2 - (EX)^2$

b) $\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(X)$ *für* $a, b \in \mathbb{R}$

c) $\operatorname{Var}(X) \geq 0$ *und* $\operatorname{Var}(X) = 0 \Leftrightarrow P(X=c) = 1$ *für ein* $c \in \mathbb{R}$

Beweis a) $\operatorname{Var}(X) = E(X^2 - 2XEX + (EX)^2) \stackrel{\text{Satz 7.2a}}{=} EX^2 - 2(EX)^2 + (EX)^2 = EX^2 - (EX)^2$

b) Wir verwenden a): $\operatorname{Var}(aX+b) = E(aX+b)^2 - (aEX+b)^2 =$
$= E(a^2X^2 + 2abX + b^2) - a^2(EX)^2 - 2abEX - b^2 =$
$= a^2EX^2 + 2abEX + b^2 - a^2(EX)^2 - 2abEX - b^2 =$
$= a^2(EX^2 - (EX)^2) = a^2 \operatorname{Var}(X)$

c) Da $0 \leq (X-EX)^2 \stackrel{7.2b}{\Rightarrow} \operatorname{Var}(X) \geq 0$
Ist X diskret, so gilt: $\operatorname{Var}(X) = \sum_{k=1}^{\infty} (x_k - EX)^2 P(X = x_k)$
$\operatorname{Var}(X) = 0 \Leftrightarrow X$ nimmt nur den Wert $x_1 = EX$ ($x_k = EX \,\forall\, k \in \mathbb{N}$) an.
Analog im stetigen Fall. ■

Beispiel 7.3 Sei $X \sim N(\mu, \sigma^2)$
Bsp. 7.2 $\Rightarrow EX = \mu$

$$\text{Also:} \quad \operatorname{Var}(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \cdots = \sigma^2$$

($\rightarrow$ Übung)

Die folgende Ungleichung ist wegen ihrer Allgemeinheit nützlich:

Satz 7.6 (Tschebyscheff-Ungleichung)
Sei X eine Zufallsvariable mit $E|X| < \infty$ und $\varepsilon > 0$. Dann gilt:

$$P(|X - EX| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \operatorname{Var}(X)$$

Beweis
Betrachte:

$$g : \mathbb{R} \to \{0, 1\} \qquad g(x) = \begin{cases} 1 & , \quad \text{falls} \quad |x - EX| \geq \varepsilon \\ 0 & , \quad \text{sonst} \end{cases}$$

$$\text{und} \quad h : \mathbb{R} \to \mathbb{R} \qquad h(x) = \frac{1}{\varepsilon^2}(x - EX)^2$$

Offenbar gilt $g(x) \leq h(x) \, \forall \, x \in \mathbb{R}$ Also folgt $g(X) \leq h(X)$ und mit Satz 7.2 b

$$P(|X - EX| \geq \varepsilon) = Eg(X) \leq Eh(X) = \frac{1}{\varepsilon^2} \operatorname{Var}(X)$$

Kapitel 8

Zufallsvektoren

8.1 Mehrstufige Zufallsexperimente

Oft besteht ein Zufallsexperiment aus einer Reihe von Vorgängen. Sei $(\Omega_i, \mathcal{A}_i, P_i)$ ein Wahrscheinlichkeitsraum für die Vorgänge $i = 1, \dots, n$.
Für das Gesamtexperiment wählen wir dann:
$\Omega = \Omega_1 \times \dots \times \Omega_n$
$\mathcal{A} := \mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$
wobei $\mathcal{A}$ die sogenannte **Produkt-σ-Algebra** ist, d.h.
$\mathcal{A} = \sigma(\{A_1 \times \dots \times A_n | A_i \in \mathcal{A}_i, i = 1 \dots n\})$

Bemerkung 8.1

a) $A_1 \times \dots \times A_n$ nennt man **Rechteckmengen**.

b) Ist $\mathcal{A}_1 = \dots = \mathcal{A}_n = \mathfrak{B} = \mathfrak{B}(\mathbb{R})$, so gilt:
$\mathfrak{B}(\mathbb{R}^n) := \underbrace{\mathfrak{B} \otimes \dots \otimes \mathfrak{B}}_{n-mal} = \sigma(\{(-\infty, x_1] \times \dots \times (-\infty, x_n] | x_i \in \mathbb{R}, i = 1, \dots, n\})$

Sei P nun ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$. Für $A_i \in \mathcal{A}_i$ wird durch
$Q_i(A_i) := P(\Omega_1 \times \dots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \dots \times \Omega_n)$
auf (Ω_i, A_i) wieder ein Wahrscheinlichkeitsmaß definiert, die sogenannte **Randverteilung (Marginalverteilung)**, denn:

(i) $Q_i(\Omega_i) = P(\Omega) = 1$

(ii)

$$Q_i(\sum_{j=1}^{\infty} A_i^{(j)}) = P(\Omega_1 \times \dots \times \Omega_{i-1} \times \sum_{j=1}^{\infty} A_i^{(j)} \times \Omega_{i+1} \times \dots \times \Omega_n) =$$

$$= P(\sum_{j=1}^{\infty} (\Omega_1 \times \dots \times \Omega_{i-1} \times A_i^{(j)} \times \Omega_{i+1} \times \dots \times \Omega_n) = \sum_{j=1}^{\infty} Q_i(A_i^{(j)})$$

Damit P sinnvoll ist, sollte gelten $Q_i(A_i) = P_i(A_i) \quad \forall\, A_i \in \mathcal{A}, i = 1 \dots n$

8.2 Zufallsvariablen

Gegeben sei ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Wir betrachten jetzt mehrere Zufallsvariablen.

Beispiel 8.1 n-mal würfeln $\Omega = \{1, \dots, 6\}^n \quad \mathcal{A} = P(\Omega) \quad P(A) = \frac{|A|}{|\Omega|}$

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

$$Y(\omega) = Y((\omega_1, \dots, \omega_n)) = \min_{i=1,\dots,n} \omega_i$$

$P(X = 3, Y = 3) = P(\{3, \dots, 3\}) = \frac{1}{6^n}$

Definition 8.1 *Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X_1, \dots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen.*

*a) $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ heißt **Zufallsvektor***

*b) Die **(gemeinsame) Verteilung** von X ist gegeben durch:*

$$P_X : \mathfrak{B}(\mathbb{R}^n) \to [0,1) \quad P_X(B) := P(\{\omega \in \Omega | X(\omega) \in B\}), B \in \mathfrak{B}(\mathbb{R}^n)$$

*c) Die (gemeinsame) **Verteilungsfunktion** $F_X : \mathbb{R}^n \to [0,1]$ ist definiert durch: $F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$*

Bemerkung 8.2 Wie im Fall $n = 1$ ist P_X durch F_X bestimmt.

Definition 8.2 *Sei $X = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$ ein Zufallsvektor.*

*a) X heißt **diskret**, falls es eine endliche oder abzählbare Menge $C = \{x_1, x_2, \dots\} \subset \mathbb{R}^n$ gibt, so dass $P(X \in C) = 1$. Die Folge $\{p_X(k)\}_{k \in \mathbb{N}}$ mit $p_X(k) = P(X = x_k)$ heißt **(gemeinsame) Zähldichte**.*

*b) X heißt **absolutstetig**, falls es eine integrierbare Funktion $f_X : \mathbb{R}^n \to [0, \infty)$ (die gemeinsame Dichte) gibt mit:*

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

Bemerkung 8.3

a) Ist $X = (X_1, \dots, X_n)$ diskret bzw. stetig, so sind auch $X_1 \dots X_n$ selbst diskret bzw. stetig und wir können die **Rand-(Marginal) Zähldichte (Dichte)** bestimmen:

$$P(X_i = x_i) = P(\{\omega | X(\omega) \in C, X_i(\omega_i) = x_i\}) = \sum_{\substack{y \in C \\ y_i = x_i}} P(X = y)$$

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

$$\text{denn:} \quad F_{X_i}(x_i) = P(X_i \leq x_i) = \lim_{\substack{x_j \to \infty \\ (i \neq j)}} \underbrace{P(X_1 \leq x_1, \ldots, X_n \leq x_n)}_{=F_X(x_1, \ldots, x_n)} = \int_{-\infty}^{x_i} f_{X_i}(u) du$$

b) Ist X absolutstetig mit Dichte f_X, so ist die Verteilung von X gegeben durch:

$$P_X(B) := \int_B f_X(y) dy \qquad \forall\, B \in \mathfrak{B}(\mathbb{R}^n)$$

Beispiel 8.2 (Multinomialverteilung) Ein Experiment (z.B. Würfeln) hat r mögliche Ausgänge $E_1, \ldots, E_r$ mit jeweiliger Wahrscheinlichkeit $p_1, \ldots, p_r$, wobei $p_1 + \cdots + p_r = 1$.

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) | w_i \in \{E_1, \ldots, E_r\}\}, \quad \mathcal{A} = \sigma(\Omega)$$

$X_i(\omega)$ sei die Anzahl der E_i-Ausgänge, $P(\{\omega_1, \ldots, \omega_n\}) := p_1^{X_1(\omega)} \cdot \cdots \cdot p_r^{X_r(\omega)}$
Sei nun $k_1, \ldots, k_n \in \mathbb{N}_0$ mit $k_1 + \cdots + k_n = n$.

$$\begin{aligned} P(X_1 = k_1, \ldots, X_n = k_n) &= p_1^{k_1} \cdots p_r^{k_r} \cdot \text{ Anzahl der } \omega_i \text{, bei denen } k_i \text{ Komponenten} \\ &\ \text{den Wert } E_i \text{ haben, } i = 1 \ldots r \\ &= p_1^{k_1} \cdots p_r^{k_r} \cdot \frac{n!}{k_1! \cdots k_n!} \end{aligned}$$

Dies ist die Zähldichte der **Multinomialverteilung** ($M(n, r, p_1, \ldots, p_r)$.

Bemerkung 8.4 a) Für $r = 2$ erhalten wir die Binomialverteilung, also $M(n, 2, p, 1-p) = B(n, p)$.

b) Die eindimensionalen Randverteilungen einer Multinomialverteilung sind binomialverteilt.

Beispiel 8.3 Der Zufallsvektor $X = (X_1, X_2)$ hat eine **bivariate Normalverteilung**, falls X absolut stetig ist mit Dichte

$$f_X(x_1, x_2) = \frac{1}{2\pi(|\Sigma|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\underline{x} - \underline{\mu})^\top \Sigma^{-1} (\underline{x} - \underline{\mu})\right)$$

$$\text{wobei:} \quad \underline{\mu} = (\mu_1, \mu_2)^\top \in \mathbb{R}^2,\ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix},\ \sigma_1^2 > 0, |\sigma_{12}| < \sigma_1 \sigma_2$$

Schreibweise: $X \sim N(\underline{\mu}, \Sigma)$

Beispiel 8.4 Gegeben sei $X = (X_1, X_2)$ mit Dichte

$$f_{(X_1, X_2)}(x_1, x_2) = \begin{cases} 4x_1 x_2 & \text{für } 0 \leq x_1,\ x_2 \leq 1 \\ 0 & \text{sonst} \end{cases}$$

Test: ($f_{(X_1,X_2)}$ Dichte)

$$\int_0^1 \int_0^1 4x_1x_2 \; dx_1dx_2 = \int_0^1 2x_2 \left[x_1^2\right]_0^1 \; dx_2 = x_2^2 \Big|_0^1 = 1$$

Randdichte von X_1:

$$f_{X_1}(x_1) = \int_0^1 4x_1x_2 \, dx_2 = 2x_1 \left[x_2^2\right]_0^1 = 2x_1 \qquad \text{für } 0 \leq x_1 \leq 1$$

$$\begin{aligned}
P(X_1 \leq 2X_2) &= P\big((X_1, X_2) \in B\big) \\
&= \int_0^{\frac{1}{2}} \int_0^{2t_2} 4t_1t_2 \; dt_1dt_2 + \int_{\frac{1}{2}}^1 \int_0^1 4t_1t_2 \; dt_1dt_2 \\
&= \int_0^{\frac{1}{2}} 2t_2 \left[t_1^2\right]_0^{2t_2} \; dt_2 + \int_{\frac{1}{2}}^1 2t_2 \left[t_1^2\right] \; dt_2 \\
&= 2 \left[t_2^4\right]_0^{\frac{1}{2}} + \left[t_2^2\right]_{\frac{1}{2}}^1 = \frac{7}{8}
\end{aligned}$$

Kapitel 9

Unabhängige Zufallsvariablen

Die Unabhängigkeit von Zufallsvariablen wird auf die Unabhängigkeit von Ereignissen zurückgeführt. Im Folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Definition 9.1

a) *Zufallsvariablen $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ heißen* ***unabhängig****, falls:*

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \ \text{d.h.}$$

$$P(\underbrace{X_1 \leq x_1, \ldots, X_n \leq x_n}_{A_1 \cap \cdots \cap A_n}) = \prod_{i=1}^{n} P(\underbrace{X_i \leq x_i}_{A_i}) \qquad \forall\, (x_1, \ldots, x_n) \in \mathbb{R}^n$$

b) *Sei $X_1, X_2, \ldots$ eine (unendliche) Folge von Zufallsvariablen. $X_1, X_2, \ldots$ heißen unabhängig, falls jede endliche Teilfolge $X_{i_1}, \ldots, X_{i_k}$ aus unabhängigen Zufallsvariablen besteht.*

Satz 9.1 *Die Zufallsvariablen $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ sind genau dann unabhängig, wenn $\forall\, B_1, \ldots, B_n \in \mathfrak{B}$ gilt:*

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} P(X_i \in B_i) \text{ bzw. } P_X(B_1 \times \cdots \times B_n) = \prod_{i=1}^{n} P_{X_i}(B_i)$$

d.h. die gemeinsame Verteilung ist das Produkt der einzelnen Randverteilungen.

Beweis

“$\Leftarrow$“ setze $B_i = (-\infty, x_i], \quad i = 1, \ldots, n$

“$\Rightarrow$“ Folgt aus Satz 4.4 und Übung $(\{(-\infty, x_i] \times \cdots \times (-\infty, x_n], x_i \in \mathbb{R} \cup \{\infty\}\})$ ist ein $\cap$-stabiler Erzeuger von $\mathfrak{B}(\mathbb{R}^n)$)

Beispiel 9.1 Wir betrachten die mehrstufigen Zufallsexperimente aus §8.1. Es sei $\Omega, \mathcal{A}$ wie in §8.1 und

$$P(A_1 \times \cdots \times A_n) = P_1(A_1) \cdots P_n(A_n) \text{ für } A_i \in \mathcal{A}_i \quad i = 1, \ldots, n$$

das sogenannte Produktmaß.
Durch die Festlegung auf den Rechteckmengen ist es auf ganz $\mathcal{A}$ eindeutig bestimmt (Satz 4.4). Weiter sei $X_i(\omega) = X_i\big((\omega_1, \ldots, \omega_n)\big) = \omega_i$ die i-te Projektion, $i = 1, \ldots, n$. Dann sind $X_1, \ldots, X_n$ unabhängig, da

$$\begin{aligned} P(X_1 \in A_1, \ldots, X_n \in A_n) = P(A_1 \times \cdots \times A_n) &= \prod_{i=1}^{n} P_i(A_i) \\ &= \prod_{i=1}^{n} P(\Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \cdots \times \Omega_n) \\ &= \prod_{i=1}^{n} P(X_i \in A) \end{aligned}$$

Satz 9.2 *Sei $X = (X_1, \ldots, X_n)$ ein Zufallsvektor.*

a) Ist X diskret mit $P(X \in C) = 1$, $C \in \mathbb{R}^n$ abzählbar, dann sind $X_1, \ldots, X_n$ unabhängig, genau dann wenn:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) \quad \forall\, (x_1, \ldots, x_n) \in \mathbb{R}^n$$

b) Ist X absolut stetig, so sind $X_1, \ldots, X_n$ unabhängig, genau dann wenn

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) \quad \forall\, (x_1, \ldots, x_n) \in \mathbb{R} \backslash B$$

wobei B "eine kleinere Menge" ist (vom Lebesgue-Maß 0). D.h. die gemeinsame Dichte ist das Produkt der Randdichten.

Beispiel 9.2 Sei $X = (X_1, X_2)$ ein absolutstetiger Zufallsvektor mit Dichte

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2 - 2\varrho x_1 x_2 + x_2^2}{1-\varrho^2}\right) \; x_1, x_2 \in \mathbb{R} \text{ für } \varrho \in [0,1)$$

Dies ist ein Spezialfall von Beispiel 8.4 mit $\sigma_1 = \sigma_2 = 1$, $\sigma_{12} = \varrho$, $\mu_1 = \mu_2 = 0$
Randdichte $f_{X_1}(x_1)$ von X_1:

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)\, dx_2 \\ &= \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2}\frac{x_1^2\left(1-\varrho^2\right)}{1-\varrho^2}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\frac{(x_2 - \varrho x_1)^2}{1-\varrho^2}\right) dx_2 \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}x_1^2\right) \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}\, du}_{=\sqrt{2\pi}} \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \quad \big(\text{also } X_1 \sim N(0,1)\big) \end{aligned}$$

$$\text{Analog:} \quad f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_2^2\right) \quad (\text{also } X_2 \sim N(0,1))$$

das heißt die Randverteilungen sind für alle ϱ Standardverteilungen. X_1, X_2 sind unabhängig $\Leftrightarrow$ $\varrho = 0$. Für $\varrho \neq 0$: $f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$

Bemerkung 9.1 Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen und $n_1 + \cdots + n_k = n \quad n_i \in \mathbb{N}$ $\varphi_i \in \mathbb{R}^{n_i} \to \mathbb{R}$ messbar, $i = 1, \ldots, k$, dann sind auch die Zufallsvariablen $\varphi_1(X_1, \ldots, X_{n_1}), \varphi_2(X_{n_1+1}, \ldots, X_{n_1+n_2}), \ldots, \varphi_k(X_{n_1+\cdots+n_{k-1}+1}, \ldots, X_n)$ unabhängig.

Beispiel 9.3 Es gibt 2 Spieler. Spieler 1 denkt sich zwei Zahlen $a, b \in \mathbb{R}$ aus. Eine faire Münze entscheidet, ob a oder b aufgedeckt wird. Spieler 2 muss raten, ob die verdeckte Zahl größer oder kleiner als die aufgedeckte Zahl ist.
Intuition: Wahrscheinlichkeit richtig zu raten $= \frac{1}{2}$
Es geht besser: Spieler 2 generiert eine Zahl c zufällig $C \sim N(0,1)$
Entscheidung: c und die aufgedeckte Zahl werden verglichen.
Mathematisches Modell:
$(\Omega_1, \mathcal{A}_1, P_1)$ mit $\Omega_1 = \{a,b\}, \mathcal{A}_1 = \mathcal{P}(\Omega), P_1(\{a\}) = P_1(\{b\}) = \frac{1}{2}$
Sei $M : \Omega_1 \to \{a,b\}$, $M(a) = a$, $M(b) = b$ die Zufallsvariablen, die das Ergebnis des Münzwurfes beschreibt.
$(\Omega_2, \mathcal{A}_2, P_2)$ mit $\Omega_2 = \mathbb{R}, \mathcal{A}_2 = \mathfrak{B}, P_2(B) = \int_B \varphi_{0,1}(x)\,dx$
$C : \Omega_2 \to \mathbb{R}$, $C(\omega) = \omega$.
Wichtig: M und C sind unabhängig. Sei o.B.d.A. $a < b$. Wir betrachten den Produktraum mit dem Produktmaß
$P = P_1 \otimes P_2$. Es sei G das Ereignis, dass Spieler 2 richtig rät.

$$P(G|M=a) = \frac{P(G, M=a)}{P(M=a)} = \frac{P(C > a, M=a)}{P(M=a)} \overset{C,M \text{ unabh.}}{=} \frac{P(C>a)P(M=a)}{P(M=a)}$$

$$\text{Analog: } P(G|M=b) = P(C \leq b)$$

$$P(G) = P(G|M=a)\frac{1}{2} + P(G|M=b)\frac{1}{2} = \frac{1}{2}(P(c>a) + P(c \leq b)) =$$

$$= \frac{1}{2}(1 - \Phi(a) + \Phi(b)) = \frac{1}{2}(1 + \underbrace{\Phi(b) - \Phi(a)}_{>0}) > \frac{1}{2}$$

Satz 9.3 *Sei $X = (X_1, X_2) : \Omega \to \mathbb{R}^2$ ein absolutstetiger Zufallsvektor mit gemeinsamer Dichte f_X. Dann ist auch die Zufallsvariable $X_1 + X_2$ absolutstetig und ihre Dichte ist gegeben durch:*

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_X(y, x-y)\,dy \qquad \forall\, x \in \mathbb{R}$$

Falls die Zufallsvariable X_1 und X_2 unabhängig sind, gilt insbesondere die sog. ***Faltungsformel***

$$f_{X_1+X_2}(x) = \int_{-\infty}^{\infty} f_{X_1}(y) \cdot f_{X_2}(x-y)\,dy \qquad \forall\, x \in \mathbb{R}$$

Bemerkung 9.2
Sind X und Y unabhängig, so schreibt man

$$P_{X+Y} = P_X * P_Y$$

und nennt $P_X * P_Y$ **Faltung**

Beweis
Zwischen Verteilungsfunktion und Dichte gibt es eine eindeutige Zuordnung. Also genügt es zu zeigen:

$$P(X_1 + X_2 \leq z) = \int_{-\infty}^{z} \int_{-\infty}^{\infty} f_X(y, x-y)\, dydx$$

$$\begin{aligned}
\int_{-\infty}^{z} \int_{-\infty}^{\infty} f_X(y, x-y)\, dydx &= \int_{-\infty}^{\infty} \int_{-\infty}^{z} f_X(y, x-y)\, dxdy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(y, u)\, dudx \\
&= \int_{\{(u,y)\in\mathbb{R}^2 | u+y \leq z\}} f_X(y, u)\, dudy \\
&= P_X(\{(u, y) \in \mathbb{R}^2 | u + y \leq z\}) = P(X_1 + X_2 \leq z)
\end{aligned}$$

Die Faltungsformel ergibt sich mit Satz 9.1.b).

Beispiel 9.4 Ist $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ und X_1, X_2 unabhängig so gilt:

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

d.h. die Normalverteilung ist **faltungsstabil**.

Satz 9.4 *Seien $X, Y : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen mit existierenden Erwartungswerten, so existiert auch der Erwartungswert von $X \cdot Y$ und es gilt:*

$$EX \cdot Y = EX \cdot EY$$

Beweis
Wir betrachten für den Beweis nur den Fall, dass (X,Y) diskret ist. (Zähldichte)

$$\begin{aligned}
\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} |x_k y_j| P(X = x_k, Y = y_j) &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} |x_k y_j| P(X = x_k) P(Y = y_j) \\
&= \underbrace{(\sum_{k=1}^{\infty} |x_k| P(X = x_k))}_{<\infty \text{ nach Vor.}} \underbrace{(\sum_{j=1}^{\infty} |y_j| P(Y = y_j))}_{<\infty \text{ nach Vor.}} < \infty
\end{aligned}$$

Also existiert EXY. Gleiche Rechnung ohne Betragsstriche ergibt $EXY = EX \cdot EY$. Im Allgemeinen folgt die Existenz von EXY nicht aus der Existenz von EX und EY.

Satz 9.5 (Cauchy-Schwarz-Ungleichung) *Es seien $X, Y : \Omega \to \mathbb{R}$ Zufallsvariablen mit existierendem zweiten Moment, so existiert auch EXY und es gilt:*

$$(EXY)^2 \leq EX^2 EY^2$$

Beweis

Es gilt: $|X \cdot Y(\omega)| = |X(\omega)||Y(\omega)| \overset{da (X-Y)^2 > 0}{\leq} X^2(\omega) + Y^2(\omega) \quad \forall\, \omega \in \Omega$
Aus der Voraussetzung und Satz 7.2.b) folgt die Existenz von EXY. Außerdem folgt mit Satz 7.2 die Existenz von $E(X + aY)^2 \quad \forall\, a \in \mathbb{R}$
Es gilt: $0 \leq E(X + aY)^2 = EX^2 + a^2 EY^2 + 2aEXY \quad \forall\, a \in \mathbb{R}$

1.Fall: $EY^2 = 0 \Rightarrow EXY = 0$, da die rechte Seite eine Gerade in a ist.
$\Rightarrow$ Ungleichung gilt.

2.Fall: $EY^2 \neq 0$: Rechte Seite wird minimal bei $a^* = -\frac{EXY}{EY^2}$
Minimal-Stelle einsetzen ergibt: $\frac{1}{EY^2}(EX^2 EY^2 - (EXY)^2) \geq 0$
$\Rightarrow$ Ungleichung gilt.

Definition 9.2 *Es seien X,Y Zufallsvariablen mit $EX^2 < \infty$, $EY^2 < \infty$.*

a) Dann heißt $\operatorname{Cov}(X, Y) := E[(X - EX)(Y - EY)]$ *die* ***Kovarianz*** *von X und Y. Ist* $\operatorname{Cov}(X, Y) = 0$*, so nennt man X und Y* ***unkorreliert****.*

b) Ist $\operatorname{Var}(X) \cdot \operatorname{Var}(Y) > 0$*, so heißt*

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

der ***Korrelationskoeffizient*** *von X und Y.*

Satz 9.6 *Seien X,Y Zufallsvariablen mit $EX^2 < \infty$, $EY^2 < \infty$. Dann gilt:*

a) $\operatorname{Cov}(X, Y) = EXY - EX \cdot EY$

b) Sind X und Y unabhängig, so auch unkorreliert.

c) Ist $\varrho(X, Y)$ erfüllt, so gilt: $-1 \leq \varrho(X, Y) \leq 1$

Beweis

a) Es gilt: $\operatorname{Cov}(X, Y) = E[XY - XEY - YEX + EX \cdot EY] = EXY - EXEY$

b) folgt aus a) und Satz 9.4

c) Mit Satz 9.5 folgt:

$$\begin{aligned}\operatorname{Var}(X) \operatorname{Var}(Y) \varrho^2(X, Y) = (E[(X - EX)(Y - EY)])^2 &\leq E(X - EX)^2 E(Y - EY)^2 \\ &= \operatorname{Var}(X) \operatorname{Var}(Y)\end{aligned}$$

Bemerkung 9.3

a) Der Korrelationskoeffizient ist ein Maß für die lineare Abhängigkeit der zwei Zufallsvariablen. Betrachte den Extremfall $Y = aX + b$ mit $a \neq 0$, $b \in \mathbb{R}$ gilt $\mathrm{Cov}(X, Y) = E[(X - EX)(aX + b - aEX - b)] = a\,\mathrm{Var}(X)$. und

$$\varrho(X, Y) = \frac{a\,\mathrm{Var}(X)}{\sqrt{a^2\,\mathrm{Var}^2(X)}} = \begin{cases} 1 & \text{falls } a > 0 \\ -1 & \text{falls } a < 0 \end{cases}$$

b) Es gibt Zufallsvariablen, die unkorreliert, aber nicht stochastisch unabhängig sind.
Es sei z.B. $\Omega = \{\omega_1, \omega_2, \omega_3\}$

ω	$P(\{\omega\})$	$X(\omega)$	$Y(\omega)$	$X(\omega) \cdot Y(\omega)$
ω_1	$\frac{1}{3}$	0	-1	0
ω_2	$\frac{1}{3}$	1	0	0
ω_3	$\frac{1}{3}$	0	1	0

Es gilt: $\mathrm{Cov}(X, Y) = 0 - 0 = 0$ unkorreliert, aber $P(X = 0, Y = 1) = \frac{1}{3}$
$P(X = 0)P(Y = 1) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9} \neq \frac{1}{3} = P(X = 0, Y = 1)$ sind <u>nicht</u> unabhängig.

c) Es gilt: $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$, $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$

Satz 9.7
Seien $X_1, \ldots, X_n$ Zufallsvariablen mit existierendem zweiten Moment, so gilt:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_1) + 2 \sum_{1 \leq i,j \leq n} \mathrm{Cov}(X_i, X_j)$$

Sind die Zufallsvariablen unabhängig (oder unkorreliert), so gilt:

$$\mathrm{Var}(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i)$$

Beweis
Mit Satz 7.2 gilt:

$$\mathrm{Var}(X_1 + \cdots + X_n) = E\left(\sum_{i=1}^{n} (X_i - EX_i)\right)^2 = E \sum_{1 \leq i,j \leq n} (X_i - EX_i)(X_j - EX_j)$$

$$= \sum_{1 \leq i,j \leq n}^{n} \mathrm{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \mathrm{Cov}(X_i, X_j)$$

Der zweite Teil aus Satz 9.6 b)

Definition 9.3
Sei $X = (X_1, \ldots, X_n)$ ein Zufallsvektor mit $EX_i^2 < \infty$, $\forall\, i = 1, \ldots, n$

a) *$EX = (EX_1, \ldots, EX_n)$ heißt **Erwartungswertvektor von X**.*

b) *Die $n \times n$-Matrix* $\mathrm{Cov}\, X = (\mathrm{Cov}(X_i, X_j))_{1 \leq i,j \leq n}$
*heißt **Kovarianzmatrix von X**.*

Kapitel 10

Erzeugende Funktionen

In diesem Kapitel sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{N}_0$ eine diskrete Zufallsvariable mit Zähldichte $\{p_X(k)\}_{k\in\mathbb{N}_0}$, also $p_X(k) = P(X = k)\,, k \in \mathbb{N}_0$

Definition 10.1
Sei $X : \Omega \to \mathbb{N}_0$ eine Zufallsvariable mit Zähldichte $\{p_X(k)\}_{k\in\mathbb{N}_0}$. Die Funktion $g_X : [-1, 1] \to \mathbb{R}$ mit

$$g_X(s) = \sum_{k=0}^{\infty} p_X(k)s^k = Es^X$$

*heißt **erzeugende Funktion von X**.*

Bemerkung 10.1

a) $g_X(s)$ ist wohldefiniert für $|s| \leq 1$, da

$$\sum_{k=0}^{\infty} p_X(k)|s|^k \leq \sum_{k=0}^{\infty} p_X(k) = 1$$

Insbesondere: $g_X(1) = 1$

b) $g_X^{(n)}(s)$ ist wohldefiniert für $s \in (-1, 1)$

c) Es gilt:

$$p_X(k) = \frac{g_X^{(k)}(0)}{k!} \quad \text{für } k \in \mathbb{N}_0$$

Es sei $g_X(s^-) = \lim_{z\uparrow s} g_X(z)$ „linksseitiger Grenzwert“

Satz 10.1
Besitzt die $\mathbb{N}_0$-wertige Zufallsvariable X die erzeugende Funktion g_X, so gilt:

a) $EX = g_X'(1^-)$, *falls* EX *existiert*

b) $\mathrm{Var}(X) = g_X''(1^-) + g_X'(1^-) - (g_X'(1^-))^2$ *falls* $\mathrm{Var}(X)$ *existiert.*

Beweis

a)

$$g'_X(s) = \sum_{k=1}^{\infty} p_X(k)k \cdot s^{k-1} \overset{s \to 1}{\to} \sum_{k=1}^{\infty} p_X(k)k = EX$$

b) ähnlich

Beispiel 10.1
Sei $X \sim B(n,p)$, $n \in \mathbb{N}$, $p \in (0,1)$

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

Es gilt:
$g_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot s^k = (sp + 1 - p)^n$
$g'_X(s) = n(sp + 1 - p)^{n-1} p$
$g''_X(s) = n(n-1)(sp + 1 - p)^{n-2} p^2$

Also:
$EX = g'_X(1^-) = np$
$\mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p)$

Satz 10.2 (Eindeutigkeitssatz für erzeugende Funktionen)
Sind X und Y zwei diskrete Zufallsvariablen mit Werten in $\mathbb{N}_0$, Zähldichten $\{p_X(k)\}_{k \in \mathbb{N}_0}$, $\{p_Y(k)\}_{k \in \mathbb{N}_0}$ und erzeugenden Funktionen $g_X(s)$, $g_Y(s)$, so gilt:

$$p_X(k) = p_Y(k) \quad \forall\, k \in \mathbb{N}_0 \Longleftrightarrow g_X(s) - g_Y(s) \quad \forall\, s \in [-1, 1]$$

Beweis
Identitätssatz für Potenzreihen.

Satz 10.3
Sind X und Y zwei unabhängige, diskrete Zufallsvariablen mit Werten in $\mathbb{N}_0$, dann gilt:

$$g_{X+Y}(s) = g_X(s) \cdot g_Y(s) \qquad \forall\, s \in [-1, 1]$$

Beweis
Für $s \in [-1, 1]$ gilt:

$$\begin{aligned}
g_{X+Y}(s) &= \sum_{k=0}^{\infty} s^k \cdot P(X + Y = k) = \sum_{k=0}^{\infty} \overbrace{s^k}^{=s^i s^{k-i}} \sum_{i=0}^{k} \underbrace{P(X = i, Y = k - i)}_{\overset{X,Y \text{ unabh.}}{=} P(X=i) \cdot P(Y=k-i)} \\
&= \sum_{k=0}^{\infty} \sum_{i=0}^{k} s^i P(X = i) s^{k-i} P(Y = k - i) \\
&= (\sum_{i=0}^{\infty} s^i P(X = i)) \cdot (\sum_{j=0}^{\infty} s^j P(Y = j)) \\
&= g_X(s) \cdot g_Y(s)
\end{aligned}$$

Beispiel 10.2
Es sei $X \sim B(n,p)$, $Y \sim B(m,p)$, X und Y unabhängig.
Mit Beispiel 10.1 gilt: $g_X(s) = (sp+1-p)^n$, $g_Y(s) = (sp+1-p)^m$

Also folgt mit Satz 10.3:

$$g_{X+Y}(s) = (sp+1-p)^{n+m} \Rightarrow X+Y \sim B(n+m,p)$$

Insbesondere ist $X = \sum_{i=1}^n X_i$, wobei $X_1,\dots,X_n$ unabhängig und identisch verteilt mit $X_i \sim B(1,p)$

Beispiel 10.3 (Ruinspiel)

- Spieler I besitzt n Euro
- Spieler II besitzt $(N-n)$ Euro
- Pro Runde: Spieler I gewinnt von Spieler II einen Euro mit Wahrscheinlichkeit p, sonst verliert er einen Euro an Spieler II
- Die Runden sind unabhängig
- Gespielt wird bis ein Spieler pleite ist

Wie groß ist die Wahrscheinlichkeit, dass Spieler I gewinnt? Sei dabei $N \in \mathbb{N}$ fest. Wir definieren die Ereignisse $A_n =$ Spieler I gewinnt bei Anfangskapital n und $B =$ Spieler I gewinnt die erste Runde. Mit dem Satz der totalen Wahrscheinlichkeit ergibt sich:

$$P(A_n) = P(A_n|B) \cdot P(B) + P(A_n|B^c) \cdot P(B^c) \text{ für } 0 < n < N$$

Sei $p_n := P(A_n)$: $p_n = p_{n+1} \cdot p + p_{n-1} \cdot (1-p),\ 0 < n < N$ und $p_0 = 0$ und $p_N = 1$. Die ist eine sogenannte Differenzengleichung. Sei $\rho := \frac{1-p}{p}$ und $\rho \neq 1$ (d.h. $p \neq \frac{1}{2}$).

$$\begin{aligned} p_{n+1} &= (1+\rho)p_n - \rho p_{n-1},\ n = 1,2,\dots \\ s^{n+1}p_{n+1} &= (1+\rho)s^{n+1}p_n - \rho s^{n+1}p_{n-1},\ n = 1,2,\dots \end{aligned}$$

Sei $\hat{g}(s) = \sum_{n=0}^\infty p_n s^n$. Dann folgt:

$$\begin{aligned} \hat{g}(s) - p_1 \cdot s &= (1+\rho)s\hat{g}(s) - \rho s^2 \hat{g}(s) \\ \Rightarrow \hat{g}(s) &= \frac{p_1 \cdot s}{1-(1+\rho)s+\rho s^2} = \frac{p_1}{\rho-1}\left(\frac{1}{1-\rho s} - \frac{1}{1-s}\right) \\ &= \frac{p_1}{\rho-1}\left(\sum_{k=0}^\infty (\rho s)^k - \sum_{k=0}^\infty s^k\right) \\ \Rightarrow p_n &= \frac{p_1}{s-1}(\rho^n - 1) \end{aligned}$$

Randbedingung: $p_N = 1$ ergibt $p_1 = \frac{\rho-1}{\rho^N-1}$. Insgesamt:

$$p_n = \frac{\rho^n-1}{\rho^N-1},\ n = 0,1\dots$$

Bei $\rho = 1$:

$$p_n = \frac{n}{N},\ n = 1,2\dots$$

Kapitel 11

Konvergenzbegriffe für Zufallsvariablen

Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $X_1, X_2, \ldots \Omega \to \mathbb{R}$ Zufallsvariablen.
"Konvergenz" der Folge $\{X_n\}$ kann man auf verschiedene Weise definieren.

Definition 11.1

a) X_n konvergiert ***P-fast sicher (P-f.s.)*** *gegen X*
Schreibweise: $X_n \stackrel{fs}{\to} X$, wenn gilt:

$$P(\{\omega \in \Omega | \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1$$

b) X_n konvergiert ***in Wahrscheinlichkeit*** *(stochastisch) gegen X*
Schreibweise: $X_n \stackrel{P}{\to} X$, wenn gilt:

$$\lim_{n\to\infty} P(\{\omega \in \Omega \Big| |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0 \quad \forall\, \varepsilon > 0$$

c) X_n konvergiert ***in Verteilung*** *gegen X*
Schreibweise: $X_n \stackrel{d}{\to} X$, wenn gilt:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \quad \forall\, x \in \mathbb{R} \quad \text{an denen}\, F_X(x)\, \text{stetig ist}$$

Beispiel 11.1
Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten (u.i.v.) Zufallsvariablen mit $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$ $\quad (\Rightarrow EX_n = 0)$
Mit der Ungleichung von Tschebyscheff (Satz 7.4) folgt:

$$P(|\frac{1}{k}\sum_{i=1}^{n} X_i - 0| \geq \varepsilon) \leq \frac{E((\frac{1}{k}\sum_{i=1}^{n} X_i)^2)}{\varepsilon^2} = \frac{1}{n^2} \cdot \frac{n}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \stackrel{n\to\infty}{\to} 0$$

Also:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \stackrel{P}{\to} 0$$

Satz 11.1
Fast sichere Konvergenz impliziert die Konvergenz in Wahrscheinlichkeit.

Beweis
Sei $X_n \overset{fs}{\to} X$ und $N := \{\omega \in \Omega | X_n(\omega) \not\to X(\omega)\}$ also $P(N) = 0$
Sei $\varepsilon > 0$ und für $n \in \mathbb{N}$

$$A_n = \{\omega \in \Omega \Big| \sup_{m \geq n} |X_m(\omega) - X(\omega)| \geq \varepsilon\}$$

Es gilt $A_n \downarrow$ und $\omega \in A_n \, \forall \, n \in \mathbb{N} \Rightarrow \forall \, n \in \mathbb{N}, \exists \, m \geq n : |X_m(\omega) - X(\omega)| \geq \frac{\varepsilon}{2}$
Also:

$$A_n \downarrow \bigcap_{m=1}^{\infty} A_m =: N_\varepsilon \subset \{\limsup_{n\to\infty} |X_n - X| \geq \frac{\varepsilon}{2}\} \subset N$$

Da P stetig von oben folgt:
$0 \leq P(|X_n - X| \geq \varepsilon) \leq P(A_n) \overset{n\to\infty}{\to} P(N_\varepsilon) \leq P(N) = 0$

Bemerkung 11.1
Die Umkehrung von Satz 11.1 gilt nicht.
Sei $(\Omega, \mathcal{A}, P) = ([0,1), \mathfrak{B}_{[0,1)}, \mathrm{Unif}(0,1))$

(Hier fehlen Skizzen, die $X_1, X_2, \ldots$ beschreiben.)
Offenbar $P(\{\omega \in \Omega | \lim_{n\to\infty} X_n(\omega) \, \text{existiert}\}) = 0$
aber $P(\{\omega \in \Omega \Big| |X_n(\omega) - 0| \geq \varepsilon\}) = P(X_n = 1) \to 0$ für $n \to \infty$
Das heißt : $X_n \overset{P}{\to} 0$ aber $X_n \overset{fs}{\not\to} 0$

Satz 11.2
Konvergenz in Wahrscheinlichkeit impliziert Konvergenz in Verteilung.

Beweis
Vorüberlegung: $P(A) = P(AB) + P(AB^c) \leq P(AB) + P(B^c)$
$\Rightarrow P(AB) \geq P(A) - P(B^c) \quad (*)$
Sei $X_n \overset{P}{\to} X$. Für alle $x \in \mathbb{R}$ gilt:
$\{\omega | X_n(\omega) \leq x\} \supset \{\omega | X(\omega) \leq x - \varepsilon, |X_n(\omega) - X(\omega)| < \varepsilon\}, \quad \varepsilon > 0$
da $X_n = X + X_n - X \leq X + |X_n - X| \leq x - \varepsilon + \varepsilon = x$

Also:

$$F_{X_n}(x) = P(X_n \leq x) \geq P(\underbrace{X \leq x - \varepsilon}_{A}, \underbrace{|X_n - X| < \varepsilon}_{B})$$

$$\Rightarrow F_{X_n}(x) \overset{(*)}{\geq} \underbrace{P(X \leq x - \varepsilon)}_{F_X(x-\varepsilon)} - P(|X_n - X| \geq \varepsilon)$$

Andererseits:

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(X_n \leq x, |X_n - X| < \varepsilon) + P(X_n \leq x, |X_n - X| \geq \varepsilon) \\ &\leq \underbrace{P(X \leq x + \varepsilon)}_{=F_X(x+\varepsilon)} + P(|X_n - X| \geq \varepsilon) \end{aligned}$$

Insgesamt:

$$F_X(x-\varepsilon) - P(|X_n - X| \geq \varepsilon) \leq F_{X_n}(x) \leq F_X(x+\varepsilon) + P(|X_n - X| \geq \varepsilon)$$

Mit $n \to \infty$ folgt:

$$F_X(x-\varepsilon) - 0 \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} F_{X_n}(x) \leq F_X(x+\varepsilon) + 0$$

Mit $\varepsilon \to 0$ (x ist Stetigkeitsstelle von F_X) folgt die Behauptung.

Bemerkung 11.2
Die Umkehrung von Satz 11.2 gilt nicht:
$\Omega = \{\omega_1, \omega_2\}, \quad \mathcal{A} = \mathcal{P}(\Omega), \quad P(\{\omega_1\}) = P(\{\omega_2\}) = \frac{1}{2}$
$X(\omega) = 1, \quad X(\omega_2) = -1$. Also:

$$P(X \leq x) = F_X(x) = \left\{ \begin{array}{ll} 0, & x < -1 \\ \frac{1}{2}, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{array} \right\} = F_{-X}(x)$$

Sei $X_n := (-1)^n X \Rightarrow F_{X_n} = F_X \Rightarrow X_n \stackrel{d}{\to} X$
Aber (X_n) konvergiert nicht in Wahrscheinlichkeit.

$$X_{2n} = X \stackrel{P}{\to} X, \quad X_{2n+1} = -X \stackrel{P}{\to} -X \quad \text{und} \quad P(X = -X) = 0$$

Insgesamt:

$$\boxed{X_n \stackrel{fs}{\to} X \;\Rightarrow\; X_n \stackrel{P}{\to} X \;\Rightarrow\; X_n \stackrel{d}{\to} X}$$

Lemma 11.3
Konvergenz in Verteilung gegen eine Konstante $c \in \mathbb{R}$ impliziert Konvergenz in Wahrscheinlichekeit, das heißt:

$$X_n \stackrel{d}{\to} c \;\Rightarrow\; X_n \stackrel{P}{\to} c$$

Beweis
Übung

Kapitel 12

Charakteristische Funktionen

In §10 haben wir für diskrete Zufallsvariablen die erzeugende Funktion betrachtet. Jetzt betrachten wir eine andere Transformierte, die für beliebige Zufallsvariablen X definiert ist. Im Folgenden sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum.

Definition 12.1
Es sei X eine Zufallsvariable. Dann heißt $\varphi_X : \mathbb{R} \to \mathbb{C}$

$$\varphi_X(t) := Ee^{itX}$$

*die **charakteristische Funktion zu X***

Bemerkung 12.1

a) Man kann im Reellen rechnen.
   $Ee^{itX} = E\cos(tX) + iE\sin(tX)$
   Insbesondere existieren die Erwartungswerte ohne weitere Bedingung.

b) $\varphi_X(t)$ hängt nur von der Verteilung von X ab

c) Ist X diskret, so gilt: $\varphi_X(t) = g_X(e^{it})$

d) Ist X absolutstetig, so gilt: $\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx$ (Fourier-Transformierte von f_X)

Beispiel 12.1
Es sei $X \equiv \mu \in \mathbb{R}$ Dann ist $\varphi_X(t) = Ee^{it\mu} = e^{it\mu}$

Beispiel 12.2
Es sei $X \sim N(0,1)$ Also:

$$\begin{aligned}
\varphi_X(t) &= Ee^{itX} = Ecos(tX) + iEsin(tX) \\
&= \int_{-\infty}^{\infty} e^{itx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\
&= \int_{-\infty}^{\infty} cos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx + i \underbrace{\int_{-\infty}^{\infty} sin(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx}_{=0} \\
&= \int_{-\infty}^{\infty} -xsin(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \underbrace{= \cdots =}_{\text{part. Integration}} - \int_{-\infty}^{\infty} tcos(tx) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx = -t\varphi_X(t)
\end{aligned}$$

und $\varphi_X(0) = 1$ Lösung der Dgl. $\varphi_X(t) = e^{-\frac{t^2}{2}}$

Satz 12.1
Es sei φ_X die charakteristische Funktion einer Zufallsvariablen X. Dann gilt:

a) $\varphi_X(0) = 1$

b) $|\varphi_X(t)| \le 1 \qquad \forall\, t \in \mathbb{R}$

c) Für $a, b \in \mathbb{R}$ gilt: $\qquad \varphi_{aX+b}(t) = e^{ibt}\varphi_X(at)$

Beweis

a) $\varphi_X(0) = Ee^{i0X} = 1$

b) $|\varphi_X(t)| \le E|e^{itX}| = 1$

c) $\varphi_{aX+b}(t) = Ee^{it(aX+b)} = e^{itb} \overbrace{Ee^{itaX}}^{=\varphi_X(at)}$

Beispiel 12.3
Es sei $X \sim N(\mu, \sigma^2)$
Es gilt: $\mu + \sigma Z \sim N(\mu, \sigma^2)$, falls $Z \sim N(0, 1)$ (Lemma 6.1)

$$\text{Also:} \qquad \varphi_X(t) = \varphi_{\mu+\sigma Z}(t) \overset{\text{Satz 12.1 c)}}{=} e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}}$$

Satz 12.2
Sind $X_1, \ldots, X_n$ unabhängige Zufallsvariablen mit charakteristischen Funktionen $\varphi_{X_1}, \ldots, \varphi_{X_n}$ so gilt für die charakteristische Funktion $\varphi_{\sum_{i=1}^n X_i}$ von $\sum_{i=1}^n X_i$:

$$\varphi_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^n \varphi_{X_i}(t), \qquad \forall\, t \in \mathbb{R}$$

Beweis

$$\varphi_{X+Y}(t) = Ee^{it(X+Y)} = E(e^{itX}e^{itY}) \overset{X,Y \text{ unabh.}}{=} Ee^{itX} \cdot Ee^{itY} = \varphi_X(t) \cdot \varphi_Y(t)$$

Satz 12.3
Falls $E|X|^n < \infty$, $n \in \mathbb{N}$, so ist φ_X n-mal differenzierbar und es gilt:

$$\varphi_X^{(n)}(0) = i^n EX^n \qquad \text{(n-te Moment)}$$

Beweis
Man darf E (= Integral) und Differentiation vertauschen.
($\to$ Majorisierte Konvergenz Stochastik II)

$$\varphi_X^{(n)}(t) = \frac{d^n}{(dt)^n} Ee^{itX} = E\left(\frac{d^n}{(dt)^n} e^{itX}\right) = E((iX)^n e^{itX}) = i^n EX^n e^{itX}$$

$\Rightarrow \varphi_X^{(n)}(0) = i^n EX^n$

Beispiel 12.4
Sei $X \sim N(\mu, \sigma^2)$ $\quad E|X|^n < \infty \quad \forall\, n \in \mathbb{N}$

$$\text{Beispiel 12.3} \quad \Rightarrow \varphi_X(t) = e^{i\mu t} e^{-\frac{\sigma^2 t^2}{2}}$$

$$\overset{\text{Satz 12.2}}{\underset{(n=1)}{\Longrightarrow}} \; EX = \frac{1}{i}(\varphi_X^i(0)) = \frac{1}{i}((i\mu - \sigma^2 t)e^{i\mu t - \frac{\sigma^2 t^2}{2}}\Big|_{t=0} = \mu$$

Satz 12.4 (Eindeutigkeitssatz für charakteristische Funktionen)
Sind X und Y Zufallsvariablen mit derselben charakteristischen Funktion, so haben X und Y dieselbe Verteilung.

Beweis
Siehe zum Beispiel: Hesse Seite 94

Beispiel 12.5
Es seien $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, X_1, X_2 unabhängig.

$$\text{Beispiel 12.3} \quad \Rightarrow \varphi_{X_1}(t) = e^{it\mu_1} e^{-\frac{\sigma_1^2 t^2}{2}} \qquad \text{(entsprechend für } X_2\text{)}$$

$$\text{Satz 12.2:} \quad \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t) \cdot \varphi_{X_2}(t) = e^{it(\mu_1+\mu_2)} \cdot e^{-\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

$\overset{Satz 12.4}{\Longrightarrow} X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (vgl. Beispiel 9.4 bzw. Übung)

Satz 12.5 (Stetigkeitssatz bei charakteristischen Funktionen)
Es sei (X_n) eine Folge von Zufallsvariablen mit zugehörigen Verteilungsfunktionen $F_{X_n}(x)$ und charakteristischen Funktionen $\varphi_{X_n}(t)$. Folgende Aussagen sind äquivalent:

a) $X_n \xrightarrow{d} X$

b) $\varphi_{X_n}(t) \to \varphi(t) \qquad \forall\, t \in \mathbb{R}$ *und φ ist stetig in 0.*

In diesem Fall ist φ die charakteristische Funktion von X.

ohne Beweis

Kapitel 13

Grenzwertsätze

13.1 Schwache Gesetze der großen Zahlen

Satz 13.1 (Tschebyscheffs schwaches Gesetz der großen Zahlen)
Es seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit $EX_i = \mu$ und $\operatorname{Var}(X_i) < \infty$. Dann gilt:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

Beweis
Mit der Ungleichung von Tschebyscheff (Satz 7.4) folgt:

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\operatorname{Var}\left(\frac{X_1+\cdots+X_n}{n}\right)}{\varepsilon^2} \overset{Satz 9.6}{=} \frac{\frac{1}{n^2}\operatorname{Var}(X_1 + \cdots + X_n)}{\varepsilon^2} =$$

$$= \frac{\operatorname{Var}(X_1)}{n \cdot \varepsilon^2} \to 0 \quad \text{für} \quad n \to \infty$$

$\Rightarrow$ Behauptung

Die Bedingung $\operatorname{Var}(X_i) < \infty$ soll weg. Zur Vorbereitung brauchen wir einige Ergebnisse aus der Analysis.

Lemma 13.2 *Für alle $m \in \mathbb{N}$ und $t \in \mathbb{R}$ gilt:*

$$\left|e^{it} - \sum_{k=0}^{m-1} \frac{(it)^k}{k!}\right| \leq \frac{|t|^m}{m!}$$

Beweis
Es genügt $t \geq 0$ zu betrachten, da $|z| = |\overline{z}|$
Für e^{it} gilt die folgende Taylorentwicklung mit Integraldarstellung des Restgliedes (Beweis durch Induktion nach m):

$$e^{it} = \sum_{k=0}^{m-1} \frac{(it)^k}{k!} + i^m \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} e^{iu} \, du$$

$$\Rightarrow \left| e^{it} - \sum_{k=0}^{m-1} \frac{(it)^k}{k!} \right| = \Big| \underbrace{i^m}_{|\cdot|=1} \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} \underbrace{e^{iu}}_{|\cdot|=1} du \Big| \leq \int_0^t \frac{(t-u)^{m-1}}{(m-1)!} du =$$

$$= -\frac{(t-u)^m}{m!}\Big|_0^t = \frac{t^m}{m!}$$

Lemma 13.3
Sei X eine Zufallsvariable mit existierendem Erwartungswert $EX = \mu$ und $\varphi_X(t)$ die zugehörige charakteristische Funktion. Dann gilt $\forall\, t \in \mathbb{R}$:

$$\varphi_X\left(\frac{t}{n}\right) = 1 + i\mu\left(\frac{t}{n}\right) + o_t\left(\frac{1}{n}\right)$$

Beweis

$$\text{Zu zeigen:} \quad \forall\, t \in \mathbb{R}\,\text{gilt:} \left[\varphi_X\left(\frac{t}{n}\right) - \left(1 + i\mu\frac{t}{n}\right)\right] \cdot n \to 0 \quad \text{für}\, n \to \infty$$

$$n \cdot \left[\varphi_X\left(\frac{t}{n}\right) - \left(1 + i\mu\frac{t}{n}\right)\right] = E\underbrace{\left[n \cdot \left(e^{i\frac{t}{n}X} - 1 - \frac{itX}{n}\right)\right]}_{=:Y_n}$$

Lemma 13.2 mit $m = 2$:

$$|Y_n| \leq n \cdot \frac{|\frac{t}{n}X|^2}{2!} = \frac{t^2X^2}{2!n} \overset{f.s.}{\to} 0 \quad \text{für}\, n \to \infty$$

Limes und E kann man hier vertauschen ($\to$ majorisierte Konvergenz)
$\Rightarrow$ Behauptung

Lemma 13.4
Sei $z \in \mathbb{C}$ fest $\{\eta_n\}_{n\in\mathbb{N}}$ $\quad \eta_n \in \mathbb{C}$ und $\eta_n \to 0$. Dann gilt:

$$\lim_{n\to\infty}\left(1 + \frac{z}{n} + \frac{\eta_n}{n}\right)^n = e^z$$

Beweis

1.Fall: $z = 0$

Zu zeigen: $(1 + \frac{\eta_n}{n})^n \to 1 \quad$ für $n \to \infty$

$$\left|(1 + \frac{\eta_n}{n})^n - 1\right| = \left|\sum_{k=1}^n \binom{n}{k}(\frac{\eta_n}{n})^k\right| \leq \sum_{k=1}^n \binom{n}{k}\left(\frac{|\eta_n|}{n}\right)^k = \left(1 + \frac{|\eta_n|}{n}\right)^n - 1$$

$$\underset{f(x)=(1+x)^n}{\overset{\text{MWS}}{\leq}} \frac{|\eta_n|}{n} n(1+\vartheta)^{n-1} \leq |\eta_n|\left(1 + \frac{|\eta_n|}{n}\right)^n \leq |\eta_n|\left(1 + \frac{1}{n}\right)^n \leq \underbrace{|\eta_n| \cdot e}_{\to 0\,\text{für}\, n\to\infty}$$

2.Fall: $z \neq 0$

$$\left(1 + \frac{z}{n} + \frac{\eta_n}{n}\right)^n = \left(1 + \frac{z}{n}\right)^n \left(\frac{1 + \frac{z}{n} + \frac{\eta_n}{n}}{1 + \frac{z}{n}}\right)^n = \underbrace{\left(1 + \frac{z}{n}\right)^n}_{\to e^z \text{ für } n\to\infty} \cdot \underbrace{\left(1 + \frac{\varepsilon_n}{n}\right)^n}_{\substack{\to 1 \text{ für } n\to\infty \\ \text{(Fall 1)}}}$$

mit $\varepsilon_n \to 0$ für $n \to \infty$

Satz 13.5 (Khinchins schwaches Gesetz der großen Zahlen, 1929)
Es seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit $E|X_i| < \infty$, $EX_i = \mu$. Dann gilt:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

Beweis
Da $\mu \in \mathbb{R}$ konstant, genügt wegen Lemma 11.3 zu zeigen $\frac{X_1+\cdots+X_n}{n} \xrightarrow{d} \mu$

$$\text{Satz 12.5} \Rightarrow \text{zu zeigen ist:} \quad \varphi_{\frac{X_1+\cdots+X_n}{n}}(t) \to \varphi_\mu(t) \overset{\text{Bsp. 12.1}}{=} e^{it\mu}$$

Also:

$$\varphi_{\frac{X_1+\cdots+X_n}{n}}(t) \overset{\text{Satz 12.1c)}}{=} \varphi_{X_1+\cdots+X_n}\left(\frac{t}{n}\right) \overset{\text{Satz 12.2}}{=} \prod_{k=1}^n \varphi_{X_k}(\frac{t}{n}) = \left(\varphi_{X_1}(\frac{t}{n})\right)^n$$

$$\overset{\text{Lem. 13.3}}{=} \left(1 + \frac{i\mu t}{n} + o_t(\frac{1}{n})\right)^n \overset{\text{Lem. 13.4}}{\longrightarrow} e^{it\mu}$$

13.2 Das starke Gesetz der großen Zahlen

Satz 13.6 (Kolmogorovs starkes Gesetz der großen Zahlen)
Es seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit $E|X_i| < \infty$, $EX_i = \mu$. Dann gilt:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{f.s.} \mu$$

ohne Beweis

Beispiel 13.1 (Monte-Carlo-Simulation)
Es sei $f : [0,1] \to \mathbb{R}_+$ eine stetige Funktion. $M := \max_{x\in[0,1]} f(x)$. Wir wollen $\int_0^1 f(x)\,dx$ numerisch (näherungsweise) berechnen.
Dazu $U_1, V_1, U_2, V_2 \ldots$ eine Folge von unabhängigen Zufallsvariablen, wobei
die (U_i) identisch verteilt sind mit $U_i \sim U(0,1)$ und
die (V_i) identisch verteilt sind mit $V_i \sim U(0,M)$. Wir setzen

$$I_k = \begin{cases} 1 & , \quad \text{falls} \quad f(U_k) > V_k \\ 0 & , \quad \text{falls} \quad f(U_k) \leq V_k \end{cases}$$

Dann gilt: $I_1, I_2, \ldots$ sind unabhängig und identisch verteilte Zufallsvariablen mit

$$EI_k = P(f(U_k) > V_k) = P\left((U_k, V_k) \in G\right) = \int_0^1 \int_0^{f(x)} \frac{1}{M} \cdot 1\,dydx = \frac{1}{M}\int_0^1 f(x)\,dx$$

$$\text{Satz 13.6:} \qquad \frac{1}{n}\sum_{k=1}^n I_k \xrightarrow{f.s.} \frac{1}{M}\int_0^1 f(x)\,dx \qquad \text{für } n \to \infty$$

Beispiel 13.2 (Normale Zahlen)
Es sei $\Omega = [0,1)$ und $\omega \in \Omega$. Wir betrachten die Dualbruchzerlegung von ω: das heißt

$$\omega = \sum_{k=1}^{\infty} a_k \cdot \frac{1}{2^k} \qquad a_k \in \{0,1\}$$

ω heißt normal, falls die Werte 0 und 1 asymptotisch gleich häufig auftreten.
Wie viele $\omega \in \Omega$ sind normal?

Sei $\mathcal{A} = \mathfrak{B}_{[0,1)}$ und definiere die Zufallsvariablen $X_1, X_2. \ \Omega \to \{0,1\}$ durch
$X_n(\omega) = a_n$ für $n \in \mathbb{N}$ (Beachte X_n ist Zufallsvariable)
Weiter sei:
$A_n(x_1,\dots,x_n) := \{\omega \in \Omega | X_1(\omega) = x_1, \dots, X_n(\omega) = x_n\}$
$= \{\omega \in \Omega | \frac{x_1}{2} + \frac{x_2}{2^2} + \dots + \frac{x_n}{2^n} \le \omega < \frac{x_1}{2} + \dots + \frac{x_n}{2^n} + \frac{1}{2^n}\}$ für $x_1,\dots,x_n \in \{0,1\}$ fest.
Wir nehmen an, dass P=unif(0,1), das heißt $P([a,b)) = b - a$ für $0 \le a < b < 1$.
Also $P(A_n(x_1,\dots,x_n)) = \frac{1}{2^n}$
$\Rightarrow P(X_1 = 0) = P(A_1(0)) = \frac{1}{2} = P(X_1 = 1)$
$P(X_2 = 0) = P(X_2 = 0, X_1 = 0) + P(X_2 = 0, X_1 = 1) =$
$= P(A_2(0,0)) + P(A_2(1,0)) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$
und $P(X_2 = 1) = \frac{1}{2}$
$P(X_1 = x_1, X_2 = x_2) = \frac{1}{4} = P(X_1 = x_1)P(X_2 = x_2)$ für $x_1, x_2 \in \{0,1\}$

Also sind die Zufallsvariablen $X_1, X_2, \dots$ unabhängig und identisch verteilt mit $P(X_k = 0) = P(X_k = 1) = \frac{1}{2}$ und $EX_k = \frac{1}{2}$

Nach Satz 13.6 gilt:

$$\frac{1}{n}\sum_{k=1}^{n} X_k \overset{f.s.}{\to} \frac{1}{2} \quad \text{für} \quad n \to \infty$$

das heißt fast alle Zahlen des Intervalls $[0,1)$ bis auf eine Menge $A \subset [0,1)$ mit $P(A) = 0$ sind "normal"

13.3 Der zentrale Grenzwertsatz

Betrachte die Situation aus Satz 13.1:
Seien $X_1, X_2, \dots$ unabhängig und identisch verteilte Zufallsvariablen,
$EX_i = \mu, \operatorname{Var}(X_i) < \infty$
Dann gilt:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \ge \varepsilon\right) \le \frac{\operatorname{Var}(X_1)}{n \cdot \varepsilon^2}$$

Setze $\varepsilon = \hat{\varepsilon} \cdot n^{-\frac{1}{2}+\delta}$ mit $\delta > 0$

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \ge \frac{\hat{\varepsilon}}{n^{\frac{1}{2}-\delta}}\right) \le \frac{\operatorname{Var}(X_1)}{n \cdot \hat{\varepsilon}^2 n^{-1+2\delta}} = \frac{\operatorname{Var}(X_1)}{\hat{\varepsilon}^2 n^{2\delta}}$$

$$\text{Also für} \quad \delta > 0 : n^{\frac{1}{2}-\delta}\left(\frac{X_1 + \dots + X_n}{n} - \mu\right) \overset{P}{\to} 0$$

Was ist mit $\delta = 0$?

Satz 13.7 (Zentraler Grenzwertsatz)
Es seien $X_1, X_2, \dots$ unabhängig und identisch verteilte Zufallsvariablen mit $EX_i = \mu$ und $0 < \operatorname{Var}(X_i) = \sigma^2 < \infty$. Dann gilt:

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} X \sim N(0,1)\,,\text{also}$$

$$P\left(\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} \leq x\right) \to \Phi(x) \quad \text{für}\, n \to \infty \quad ,\forall\, x \in \mathbb{R}$$

Beweis
Betrachte zunächst den Fall $\mu = 0, \sigma = 1$

$$\text{zu zeigen:} \quad \frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{d} X \sim N(0,1)$$

$$\varphi_{\frac{X_1+\dots+X_n}{\sqrt{n}}}(t) \overset{\text{Satz 12.1c)}}{=} \varphi_{X_1+\dots+X_n}(\frac{t}{\sqrt{n}}) \overset{\text{Satz 12.2}}{=} \left(\varphi_{X_1}(\frac{t}{\sqrt{n}})\right)^n =$$

$$\overset{\text{vgl.L.13.3}}{=} \left(1 + \underbrace{\frac{it\mu}{\sqrt{n}}}_{=0} + \frac{(it)^2}{2n}\underbrace{EX^2}_{=1} + o(\frac{1}{n})\right)^n = \left(1 - \frac{t^2}{2n} + o(\frac{1}{n})\right)^n \xrightarrow{\text{nach Lem. 13.4}} e^{-\frac{t^2}{2}}$$

Allgemeiner Fall:
Setze $Y_n = \frac{X_n - \mu}{\sigma}$
Also $Y_1, Y_2, \dots$ unabhängig und identisch verteilt mit $EY_1 = 0, \operatorname{Var}(Y_n) = 1$
Wende jetzt Spezialfall an:

$$\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}} \xrightarrow{d} X \sim N(0,1)$$

Korollar 13.8
Unter den Voraussetzungen des ZGWS gilt für feste $\alpha, \beta \in \mathbb{R}, \alpha < \beta$:

$$P\left(\alpha \leq \underbrace{\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n}\sigma}}_{=:T_n} \leq \beta\right) \to \Phi(\beta) - \Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-\frac{t^2}{2}}\, dt$$

Beweis Sei $\varepsilon > 0$.

$$\underbrace{F_{T_n}(\beta) - F_{T_n}(\alpha)}_{\overset{n\to\infty}{\to} \Phi(\beta)-\Phi(\alpha)} \leq P(\alpha \leq T_n \leq \beta) \leq \underbrace{F_{T_n}(\beta) - F_{T_n}(\alpha - \varepsilon)}_{\overset{n\to\infty}{\to} \Phi(\beta)-\Phi(\alpha-\varepsilon)}$$

Mit $\varepsilon \to 0$ folgt daraus die Behauptung. ∎

Satz 13.9 *Sind die Zufallsvariablen* X_1, X_2 *unabhängig und identisch verteilt mit* $EX_1 = \mu$ *sowie* $|\mu|_3 = E|X_1 - \mu|^3$*, so gilt:*

$$\sup_{x \in \mathbb{R}} |F_{T_n}(x) - \Phi(x)| \leq \frac{|\mu|_3}{\sigma^3 \sqrt{n}}$$

Beispiel 13.3
Es seien $X_1, X_2, \ldots$ unabhängig und identisch verteilte Zufallsvariablen mit
$X_n \sim B(1,p)$. Also $EX_n = p, \mathrm{Var}(X_n) = p(1-p)$, $0 < p < 1$
Sei $S_n = X_1 + \cdots + X_n$, dann gilt nach dem ZGWS:

$$\lim_{n \to \infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x)$$

Umgeformt ergibt das:

$$\lim_{n \to \infty} P(S_n \leq np + x\sqrt{np(1-p)}) = \Phi(x)$$

Da $S_n \sim B(n,p)$ heißt das:

$$\lim_{n \to \infty} \sum_{k \leq np + x\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k} = \Phi(x)$$

Diese Version nennt man auch **Grenzwertsatz von DeMoivre Laplace**.

Beispiel 13.4 (Wahlumfrage)
Wir wollen den Anteil p der Anhänger der Partei A unter den Wahlberechtigten ermitteln. Dazu nehmen wir eine Stichprobe $X_1, \ldots, X_n$, wobei:

$$X_i = \begin{cases} 1 & , \text{ falls Person } i \text{ Partei A wählt} \\ 0 & , \text{ falls Person } i \text{ Partei A nicht wählt} \end{cases} \qquad i = 1, \ldots, n$$

Sei $S_n := X_1 + \cdots + X_n$. Wir schätzen $\hat{p} = \frac{S_n}{n}$
Wie groß muss der Stichprobenumfang n mindestens gewählt werden, damit der Schätzfehler $|\hat{p} - p|$ mit einer Wahrscheinlichkeit von mindestens 0.95 nicht größer als 0.02 ist?
Also gesucht ist die kleinste Zahl $n \in \mathbb{N}$ mit

$$P(|\frac{S_n}{n} - p| \leq 0.02) \geq 0.95 \Leftrightarrow P\left(|\frac{S_n - np}{\sqrt{np(1-p)}}| \leq \frac{0.02\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq 0.95$$

Wegen $p(1-p) = \frac{1}{4} - (p - \frac{1}{2})^2 \leq \frac{1}{4} \quad \forall\, p \in [0,1]$ folgt:

$$P\left(|\frac{S_n - np}{\sqrt{np(1-p)}}| \leq \frac{0.02\sqrt{n}}{\sqrt{p(1-p)}}\right) \geq P\left(|\frac{S_n - np}{\sqrt{np(1-p)}}| \leq 0.04\sqrt{n}\right)$$

Für n groß ist etwa $\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0,1)$

$$\Rightarrow P\left(|\frac{S_n - np}{\sqrt{np(1-p)}}| \leq 0.04\sqrt{n}\right) \approx \Phi(0.04\sqrt{n}) - \Phi(-0.04\sqrt{n}) = 2\Phi(0.04\sqrt{n}) - 1 = 0.95$$

$$\Rightarrow n = (25\Phi^{-1}(0.975))^2 = 2401$$

Kapitel 14

Parameterschätzung

Modell
Es sei $\{P_\theta | \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^m$ eine Familie von Verteilungen auf χ (sog. Stichprobenraum), $x = (x_1, \ldots, x_n)$ sei eine Realisierung der Zufallsstichprobe $X = (X_1, \ldots, X_n)$ zu einer Verteilung $P \in \{P_\theta | \theta \in \Theta\}$, wobei der wahre Parameter θ unbekannt ist.

Problem
Schätze unbekanntes θ aus der konkreten Realisierung von X.

Definition 14.1
Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen mit Verteilung P.
Dann heißt der Zufallsvektor $(X_1, \ldots, X_n)$ ***Zufallsstichprobe zur Verteilung*** *P.*
Eine Realisierung $(x_1, \ldots, x_n)$ von $(X_1, \ldots, X_n)$ nennt man ***Stichprobe****.*

Beispiel 14.1
Gegeben sei ein Würfel, den wir n-mal werfen dürfen. Aus der Beobachtung ist die Wahrscheinlichkeit zu schätzen eine 6 zu würfeln.
Modell: $\chi = \{0, 1\}$: $0 \mathrel{\hat{=}}$ keine 6 gewürftelt, $1 \mathrel{\hat{=}} 6$ gewürfelt.
$P_\theta = B(1, \theta)$, $\Theta = [0, 1]$

Definition 14.2 *Eine messbare Abbildug $T : \chi^n \to \tilde{\Theta}$, $\tilde{\Theta} \supset \Theta$, heißt* ***Schätzer*** *für θ.*

Beispiel 14.2

a) Die Statistik $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ heißt **Stichprobenmittel**

b) Die Statistik $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ heißt **Stichprobenvarianz**

Beispiel 14.3 (Schätzung eines Fischbestandes)
Teich enthält unbekannte Zahl $N = \theta$ von Fischen
r Fische werden gefangen, (rot) markiert und wieder ausgesetzt.
In einem zweiten Zug werden m Fische gefangen x davon seien markiert.
Wie groß ist N?

Modell

Urne mit $N = \theta$ Kugeln, r rot, $N - r =: s$ schwarz
m Kugeln werden ohne Zurücklegen gezogen. Die Zufallsvariable X beschreibe die Anzalh der roten unter den gezogenen Kugeln.
Also
$\chi = \mathbb{N}_0$, $P_\theta \sim$ Hypergeom.(θ, r, m), $\Theta = \mathbb{N}$. Beachte: $n = 1$

14.1 Maximum-Likelihood-Methode

Idee:
Wir wählen für θ den Wert, unter dem die Wahrscheinlichkeit, dass die konkrete Stichprobe vorliegt maximiert wird.
Im folgenden sei P_θ diskret mit Zähldichte $p(x;\theta)$ oder stetig mit Dichte $f(x;\theta)$

Definition 14.3
Gegeben sei eine Stichprobe $x = (x_1, \ldots, x_n)$. Dann heißt die Funktion

$$L_x(\theta) := f(x_1;\theta) \cdot \cdots \cdot f(x_n;\theta) \quad \text{bzw.} \quad L_x(\theta) := \underbrace{p(x_1;\theta) \cdot \cdots \cdot p(x_n;\theta)}_{P_\theta(X_1,\ldots,X_n)=(x_1,\ldots,x_n)}$$

*die **Likelihood-Funktion** der Stichprobe.*
*Eine Funktion $\hat{\theta}_{\mathrm{ML}}(x)$ heißt **Maximum-Likelihood Schätzer (MLS)**, falls*

$$L_x(\hat{\theta}_{ML}(x)) = \sup_{\theta \in \Theta} L_x(\theta)$$

Bemerkung 14.1

a) Ist P_θ diskret, so gilt:

$$L_X(\theta) = p(x_1;\theta) \cdots p(x_n;\theta) = P_\theta(X_1 = x_1) \cdots P_\theta(X_n = x_n) \overset{X \text{ unabhängig}}{=} P_\theta(X = x)$$

b) Der MLS $\hat{\theta}_{\mathrm{ML}}(x)$ ist nicht immer eindeutig.

Beispiel 14.4 (vgl. Beispiel 14.3)
$n = 1$

$$\text{Likelihood-Funktion:} \quad L_x(\theta) = \frac{\binom{r}{x}\binom{\theta-r}{k-x}}{\binom{\theta}{k}}$$

Für welches θ ist $L_x(\theta)$ maximal?
Betrachte:

$$\frac{L_x(\theta)}{L_x(\theta-1)} = \frac{\binom{r}{x}\binom{\theta-r}{m-x}\binom{\theta-1}{m}}{\binom{\theta}{m}\binom{r}{x}\binom{\theta-1-r}{m-x}} = \frac{(\theta-r)(\theta-m)}{\theta(\theta-r-m+x)}$$

$$L_x(\theta) > L_x(\theta-1) \Leftrightarrow (\theta-r)(\theta-m) > \theta(\theta-r-m+x) \Leftrightarrow mr > \theta x \Leftrightarrow \theta < \frac{mr}{x}$$

Also ist $L_x(\theta)$ maximal für $\hat{\theta}(x) = \lfloor \frac{mr}{x} \rfloor$

$\hat{\theta}(x)$ ist eindeutig, falls $\frac{mr}{x} \notin \mathbb{N}$
Falls $\frac{mr}{x} \in \mathbb{N}$ sind $\hat{\theta}_1(x) = \frac{mr}{x}$ und $\hat{\theta}_2(x) = \frac{mr}{x} - 1$ MLS

Beispiel 14.5 (Schätzung einer Erfolgswahrscheinlichkeit)
Aus n Bernoulli-Experimenten liegen x Erfolge vor, gesucht ist die Erfolgswahrscheinlichkeit:
Modell: $\chi = \mathbb{N}$, $n = 1$, $P_\theta = B(m, \theta)$, $\Theta = (0, 1)$.

$$\text{Likelihood-Funktion:} \quad L_x(\theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x} \quad , \theta \in [0,1]$$

Statt $L_x(\theta)$ ist es oft einfacher, $\log(L_x(\theta))$ zu maximieren, die sogenannte **Log-Likelihoodfunktion**

$$\log(L_x(\theta)) = \log \binom{m}{x} + x \log \theta + (m-x) \log(1-\theta)$$

$$\frac{\partial}{\partial \theta} \log(L_x(\theta)) = \frac{x}{\theta} - \frac{(m-x)}{1-\theta} = 0 \quad \Leftrightarrow \quad \theta = \frac{x}{m}$$

θ ist tatsächlich Maximum-Stelle, dass heißt $\hat{\theta}_{\text{ML}}(x) = \frac{x}{m}$

14.2 Momentenmethode

Idee:
Die ersten Momente von P_θ sollten mit den empirischen Momenten übereinstimmen.
Aus diesen Gleichungssystemen bestimmen wir den Schätzer.
Es sei $X \sim P_\theta$. Dann ist das k-te Moment

$$\mu_k = \mu_k(\theta) = E_\theta X^k \quad , k = 1, 2, \ldots$$

Wir betrachten nun die **empirischen Momente** zur Stichprobe $x = (x_1, \ldots, x_n)$.

$$\overline{x_k} := \frac{1}{n} \sum_{i=1}^{n} x_i^k$$

Es soll nun gelten: $\mu_k(\theta) = \overline{x_k}$ $k = 1, 2, \ldots, m$. Aufgelöst nach θ ergibt sich dann der Momentenschätzer $\hat{\theta}_{\text{MM}}(x) \in \Theta$.

Beispiel 14.6
$P_\theta = N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, $m = 2$

(1) $\mu_1 = \mu = E_\theta X = \overline{x} = \frac{1}{n} \sum_{i=1}^n x_i$

(2) $\mu_2 = E_\theta X^2 = \text{Var}_\theta(X) + (E_\theta X)^2 = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$

Aus (1) folgt: $\hat{\mu}(x) = \frac{1}{n} \sum_{i=1}^n x_i = \overline{x}$
Aus (2) folgt: $\hat{\sigma}^2(x) = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\overline{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^2$

14.3 Wünschenswerte Eigenschaften von Punktschätzern

Im folgenden sei $X = (X_1, \ldots, X_n)$ eine Zufallsstichprobe zur Verteilung P_θ und $T : \chi \to \tilde{\Theta}$ ein Schätzer von θ. Mit E_θ bezeichnen wir den Erwartungswert bezüglich P_θ.

Definition 14.4

a) *Der Schätzer T heißt **erwartungstreu** (unbiased), falls*

$$E_\theta T(X_1, \ldots, X_n) = \theta \qquad \forall\, \theta \in \Theta$$

b) *$b_T(\theta){:=}\, E_\theta T(X_1, \ldots, X_n) - \theta$ heißt **Verzerrung** (Bias) des Schätzers T. Ein erwartungstreuer Schätzer ist unverzerrt.*

Beispiel 14.7 (vgl. Bsp.14.6)

- $T(x) = \overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$ ist ein erwartungstreuer Schätzer für $\theta = E_\theta X_i$, denn $E_\theta(T(X)) = E_\theta\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E_\theta X_i = \theta$
- Ein erwartungstreuer Schätzer für $\theta = \mathrm{Var}_\theta(X_i)$ ist $[S^2(x) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$

Definition 14.5
Sei T ein Schätzer für θ Dann heißt

$$\mathbf{MSE(T)} := E_\theta[(T(X_1, \ldots, X_n) - \theta)^2]$$

*(mittlerer) **quadratischer Fehler** ("mean-squared-error")*

Beispiel 14.8
Sei $P_\theta = U(0, \theta)$, $\Theta = \mathbb{R}$, $\chi = \mathbb{R}_+$ und $X = (X_1, \ldots, X_n)$ eine Zufallsstichprobe zur Verteilung $U(0, \theta)$
Momentenmethode: $\overline{x} = E_\theta X_i = \frac{\theta}{2} \Rightarrow \hat{\theta}_{\mathrm{MM}} = 2 \cdot \overline{x}$
Maximum-Likelihood-Methode:

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & \text{sonst} \end{cases}$$

$$\Rightarrow L_x(\theta) = L_{(x_1,\ldots,x_n)}(\theta) = f(x_1;\theta) \cdot \cdots \cdot f(x_n;\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 \le x_1, \ldots, x_n \le \theta \\ 0, & \text{sonst} \end{cases}$$

Maximiere $L_x(\theta)$ in θ: $\hat{\theta}_{\mathrm{ML}}(x) = \max(x_1, \ldots, x_n)$
Welcher Schätzer ist besser?

$$E_\theta[\hat{\theta}_{\mathrm{MM}}(X)] = 2E_\theta\overline{X} = \theta \quad \text{, also ist } \hat{\theta}_{\mathrm{MM}} \text{ erwartungstreu}$$

Verteilungsfunktion von $\hat{\theta}_{\mathrm{ML}}(X)$ ist

$$F_\theta(x) = P_\theta(\max(X_1, \ldots, X_n) \le x) = P_\theta(X_1 \le x, \ldots, X_n \le x)$$

$$= P_\theta(X_1 \le x) \cdot \dots \cdot P_\theta(X_n \le x) = \left(\frac{x}{\theta}\right)^n \qquad \text{,falls} \quad 0 \le x \le \theta$$

Also Dichte von $\hat{\theta}_{\mathrm{ML}}(X)$:

$$f_{\hat{\theta}_{\mathrm{ML}}}(x) = \begin{cases} \frac{n}{\theta}(\frac{x}{\theta})^{n-1} & , \text{ falls } \quad 0 \le x \le \theta \\ 0 & , \text{ sonst} \end{cases}$$

und

$$E_\theta[\hat{\theta}_{\mathrm{ML}}(X)] = \int_0^\theta x f_{\hat{\theta}_{\mathrm{ML}}}(x)\,dx = \int_0^\theta n\left(\frac{x}{\theta}\right)^n dx = \frac{n}{n+1}\theta$$

also <u>nicht</u> erwartungstreu.

Aber:

$$\mathrm{MSE}(\hat{\theta}_{\mathrm{MM}}(X)) = E_\theta\left([2\frac{1}{n}\sum_{i=1}^n X_i - \theta]^2\right) = \frac{\theta^2}{3n}$$

$$\mathrm{MSE}(\hat{\theta}_{\mathrm{ML}}(X)) = E_\theta([\max(X_1, \dots, X_n) - \theta]^2) = \frac{2\theta^2}{(n+2)(n+1)}$$

$$\frac{\mathrm{MSE}(\hat{\theta}_{\mathrm{MM}}(X))}{\mathrm{MSE}(\hat{\theta}_{\mathrm{ML}}(X))} = \frac{2}{3}\frac{n}{(n+2)(n+1)} \qquad \textbf{relative Effizienz}$$

Bei großem n ist $\mathrm{MSE}(\hat{\theta}_{\mathrm{ML}}(X))$ kleiner als $\mathrm{MSE}(\hat{\theta}_{\mathrm{MM}}(X))$.

Bemerkung 14.2
Falls T erwartungstreu ist, gilt $MSE(T) = \mathrm{Var}_\theta(T)$

Für $\mathrm{Var}_\theta(T)$ kann man die folgende untere Abschätzung angeben.

Satz 14.1 (Ungleichung von Rao-Cramér)
Sei $X = (X_1, \dots, X_n)$ eine Zufallsstichprobe zur Verteilung P_θ. T sei ein Schätzer für θ. Dann gilt:

$$\mathrm{Var}_\theta(T(X)) \ge \frac{(1 + \frac{\partial}{\partial\theta} b_T(\theta))^2}{E_\theta[(\frac{\partial}{\partial\theta}\log L_X(\theta))^2]}$$

Bemerkung 14.3

(i) $I(\theta) := E_\theta[(\frac{\partial}{\partial\theta}\log L_X(\theta))^2]$ heißt **Fisher-Information**

(ii) Ist T erwartungstreu, so ist $b_T(\theta) = 0$ und $\mathrm{Var}_\theta(T(X)) \ge \frac{1}{I(\theta)}$

Beweis
Wir nehmen an, dass $L_x(\theta) > 0 \;\forall x \in \chi^n \;\forall \theta \in \Theta$, Θ sei ein offenes Intervall in $\mathbb{R}$ und P_θ sei diskret. Es gilt:

$$\frac{\partial}{\partial\theta}\log L_x(\theta) = \frac{L'_x(\theta)}{L_x(\theta)}$$

Weiter gilt:

(1) $\sum_{x \in \chi^n} L_x(\theta) = \sum_{x \in \chi^n} P_\theta(X = x) = 1$

(2) $\theta + b_T(\theta) = E_\theta T(X) = \sum_{x \in \chi^n} T(x) \cdot L_x(\theta)$

Wir differenzieren (1) und (2) nach θ, und nehmen an, dass wir $\frac{\partial}{\partial\theta}$ und $\sum$ vertauschen könne.

(1')

$$0 = \sum_{x \in \chi^n} L'_x(\theta) = \sum_{x \in \chi^n} \frac{\partial}{\partial\theta} \log(L_x(\theta)) \cdot L_x(\theta) = E_\theta \left[\frac{\partial}{\partial\theta} \log L_X(\theta) \right]$$

(2')

$$1 + b'_T(\theta) = \sum_{x \in \chi^n} T(x) L'_x(\theta)$$

$$\sum_{x \in \chi^n} T(x) \frac{\partial}{\partial\theta} \log(L_x(\theta)) \cdot L_x(\theta) = E_\theta \left[T(X) \cdot \frac{\partial}{\partial\theta} \log L_X(\theta) \right]$$

(2') - (1')$E_\theta T(X)$

$$1 + b'_T(\theta) = E_\theta[(T(X) - E_\theta T(X)) \frac{\partial}{\partial\theta} \log L_X(\theta)]$$

Mit der Ungleichung von Cauchy-Schwarz folgt:

$$\left(1 + b'_T(\theta)\right)^2 = \left(E_\theta[(T(X) - E_\theta T(X)) \frac{\partial}{\partial\theta} \log L_X(\theta)] \right)^2$$

$$\leq E_\theta[(T(X) - E_\theta T(X))^2] \cdot E_\theta \left[\frac{\partial}{\partial\theta} \log L_X(\theta)^2 \right] = \mathrm{Var}_\theta T \cdot I(\theta)$$

Kapitel 15

Konfidenzintervalle

Definition 15.1
*Sei $\alpha \in (0,1)$ fest vorgegeben. Ein Intervall der Form $[L(x), U(x)]$ mit messbaren Funktionen $L, U : \chi^n \to \Theta \subset \mathbb{R}$ heißt $(1-\alpha)$-**Konfidenzintervall**, falls $L(x) \leq U(x)\ \forall x \in \chi^n$ und $P_\theta(L \leq U) = 1$ mit $P_\theta(L \leq \theta \leq U) = 1 - \alpha$*

Bemerkung 15.1

(i) Sowohl Lage als auch Länge des Konfidenzintervalls hängen von der konkreten Stichprobe ab.

(ii) Sei zum Beispiel $\alpha = 0.05$, dann enthält das Konfidenzintervall in 95% der Fälle den wahren Parameter.

Beispiel 15.1
Es sei $P_\theta = N(\theta, \sigma^2)$, σ^2 bekannt. Stichprobe X vom Umfang n.
Bestimme $(1-\alpha)$-Konfidenzintervall für θ.
Sei

$$z := \frac{\overline{X} - \theta}{\sigma}.$$

Unter P_θ gilt: $z \sim N(0,1)$, (wegen Lemma 6.2, Bsp. 9.4). Wichtig: die Verteilung von z unter P_θ hängt nicht mehr von θ ab.

$$\Phi(c) - \Phi(-c) = P_\theta(-c \leq z \leq c) = P_\theta(\underbrace{X - \frac{c}{\sqrt{n}}\sigma}_{L(X)} \leq \theta \leq \underbrace{\overline{X} + \frac{c}{\sqrt{n}}\sigma}_{U(X)}) \overset{!}{=} 1 - \alpha$$

$$1 - \alpha = \Phi(c) - \Phi(-c) = \Phi(c) - (1 - \Phi(c)) = 2\Phi(c) - 1 \Rightarrow \Phi(c) = 1 - \frac{\alpha}{2}$$

Mit z_α bezeichnen wir im Folgenden das α-Quantil der Standardnormalverteilung. Also: $\Phi(z_\alpha) = \alpha$. Wegen Symmetrie gilt: $z_\alpha = -z_{1-\alpha}$
$\Rightarrow = z_{1-\frac{\alpha}{2}}$. Ein $(1-\alpha)$-Konfidenzintervall ist also:

$$\left[\overline{x} - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sigma, \overline{x} + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sigma\right]$$

Die ist ein Sonderfall. Die Länge des Konfidenzintervalls: $2 \cdot \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}\sigma$ und hängt nicht vom Zufall ab.

Beispiel 15.2 Es sei $P_\theta = B(m,\theta)$, $\Theta = [0,1]$, $\chi = \mathbb{N}_0$, $n = 1$

$$\text{Beispiel 13.4:} \quad P_\theta\left(\frac{X - m\theta}{\sqrt{m\theta(1-\theta)}} \le x\right) \approx \Phi(x)$$

Dann gilt:

$$P_\theta\left(-c \le \frac{X - m\theta}{\sqrt{m\theta(1-\theta)}} \le c\right) \approx \Phi(c) - \Phi(-c) \overset{!}{=} 1 - \alpha \Rightarrow c = z_{1-\frac{\alpha}{2}}$$

Wir versuchen auf die Darstellung $L(x) \le \theta \le U(x)$ zu kommen:

$$\begin{aligned} -c \le \frac{X - m\theta}{\sqrt{m\theta(1-\theta)}} \le c &\Leftrightarrow |X - m\theta| \le c\sqrt{m\theta(1-\theta)} \\ &\Leftrightarrow (X - m\theta)^2 \le c^2 m\theta(1-\theta) \\ &\Leftrightarrow \theta^2(c^2+m) - \theta(2x+c^2) + \frac{x^2}{m} \le 0 \end{aligned}$$

Nullstellen der Parabel in θ:

$$\theta_{1/2} = \frac{1}{m+c^2}\left(x + \frac{c^2}{2} \pm c\sqrt{\frac{X(m-X)}{m} + \frac{c^2}{4}}\right)$$

Das heißt:

$$L(x) = \frac{1}{m + (z_{\frac{\alpha}{2}})^2}\left(x + \frac{(z_{\frac{\alpha}{2}})^2}{2} + z_{\frac{\alpha}{2}}\sqrt{\frac{x(m-x)}{m} + \frac{(z_{\frac{\alpha}{2}})^2}{4}}\right)$$

$$U(X) = \frac{1}{m + (z_{\frac{\alpha}{2}})^2}\left(x + \frac{(z_{\frac{\alpha}{2}})^2}{2} - z_{\frac{\alpha}{2}}\sqrt{\frac{x(m-x)}{m} + \frac{(z_{\frac{\alpha}{2}})^2}{4}}\right)$$

In einer Klinik gab es im letzten Jahr 87827 Geburten, davon 45195 Jungen.
Gesucht: 0.99 - Konfidenzintervall für die Wahrscheinlichkeit, dass ein Neugeborenes männlich ist.
Hier: $m = 87827$, $\quad x = 45195$, $\quad \alpha = 0,01$

$$z_{\frac{\alpha}{2}} = z_{0,005} = -z_{0,995} = -z_{1-\frac{\alpha}{2}} = -2,576$$

$$\text{Einsetzen: } [0,51091, 0,51961]$$

Kapitel 16

Testtheorie

16.1 Einführung

$\{P_\theta | \theta \in \Theta\}$,
$x = (x_1, \ldots, x_n)$ ist Realisierung von der Zufallsstichprobe $X = (X_1, \ldots, X_n)$ zur Verteilung P_θ.
$\Theta = \Theta_0 + \Theta_1$, θ ist nicht bekannt.
Wir müssen entscheiden, ob eher $\theta \in \Theta_0$ oder $\theta \in \Theta_1$.

D.h. wir wägen die Hypothese $H_0 : \theta \in \Theta_0$
gegen die Alternative $H_1 : \theta \in \Theta_1$
ab.

Falls $\Theta \subset \mathbb{R}$, können folgende typische Fragestellungen auftreten. (sei $\theta_0 \in \Theta$)

$$H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0$$

ein sogenannter **einseitiges Testproblem**.

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

ein sogenanntes **zweiseitiges Testproblem**.

Beispiel 16.1

a) Einseitiger Test:
   ISt der Schadstoffgehalt im Nahrungsmittel der $\not\leq$ der zulässigen Grenze θ_0?

b) Zweiseitiger Test:
   Hält die Abfüllanlage das Sollgewicht θ_0 ein?

Aufgabe:
Bestimme $R \subset \mathbb{R}^n = \chi^n$, so dass H_0 verworfen wird, falls die Stichprobe $x \in R$. R heißt **kritischer Bereich**.

Folgende Entscheidungen sin möglich:
$0 = H_0$ wird nicht verworfen.
$1 = H_0$ wird verworfen.

Definition 16.1
Gegeben sei ein Testproblem H_0 vs H_1.

a) *Sei $x \in \chi^n$ eine Stichprobe. Eine Funktion $\varphi : \chi^n \to \{0,1\}$ heißt* ***Test*** *oder* ***Testverfahren****. Es gilt: $R = \{x \in \chi^n | \varphi(x) = 1\}$*

b) *Einen* ***Fehler erster Art*** *macht man, wenn man zu Unrecht H_0 ablehnt. Einen* ***Fehler zweiter Art*** *macht man, wenn man zu Unrecht H_0 annimmt.*

c) *Sei φ ein Test. Die Funktion $\beta : \Theta \to [0,1]$ definiert durch:*

$$\beta(\theta) = P_\theta(X \in R) = P_\theta(\varphi(x) = 1)$$

heißt ***Gütefunktion****. Für $\theta \in \Theta_1$ heißt $\beta(\theta)$* ***Macht des Test****.*

Bemerkung 16.1

Entscheidung “wahr“	H_0	H_1
H_0	ok	Fehler 1.Art
H_1	Fehler 2.Art	ok

Definition 16.2
Gegeben sei ein Test.
Wir sagen, dass der Test ***Niveau (Signifikanzniveau)*** *α hat, falls $\forall\ \theta \in \Theta_0$ gilt: $\beta(\theta) \leq \alpha$, d.h. die Wahrscheinlichkeit für einen Fehler erster Art ist maximal α.*

16.2 Tests unter Normalverteilungsannahme

Test auf den Mittelwert bei bekannter Varianz

$X = (X_1, ..., X_n), X_i \sim P_\mu = N(\mu, \sigma_0^2), \sigma_0$ sei bekannt, $\mu \in \Theta = \mathbb{R}$.

Testproblem:

$$H_0 : \mu \leq \mu_0 \quad \text{vs} \quad H_1 : \mu > \mu_0 = 1000$$

mit $\mu_0 \in \mathbb{R}$ gegeben.

Sinnvoller Test:

$$\varphi(x) = \varphi(x_1, \dots, x_n) = \begin{cases} 1 & , \text{ falls } \overline{x} \leq c \\ 0 & , \text{ falls } \overline{x}\, c \end{cases} , \overline{x} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$c \in \mathbb{R}$ ist jetzt noch zu bestimmen und zwar so, dass der Test das Niceau α erhält.

Bestimmung der Gütefunktion:
betrachte $\sqrt{n}\ \frac{\overline{x}-\mu}{\sigma_0}$

$$P_\mu \left(\sqrt{n} \frac{\overline{X} - \mu}{\sigma_0} \leq x \right) = \Phi(x), (\text{siehe Lemma 6.2}$$

$$\beta_c(\mu) = P_\mu(\overline{X} > c) = P_\mu \left(\sqrt{n} \frac{\overline{X} - \mu}{\sigma_0} > \sqrt{n} \frac{c - \mu}{\sigma_0} \right) = 1 - \Phi \left(\sqrt{n} \frac{c - \mu}{\sigma_0} \right)$$

Einstellen des Testniveaus: $\beta_c(\mu) \leq \alpha \ \forall \ \mu \leq \mu_0$
Es gilt:

$$\begin{array}{rcll} \mu \mapsto \beta_c(\mu) & , & \text{wachsend für festes} & c \in \mathbb{R} \\ c \mapsto \beta_c(\mu) & , & \text{fallend für festes} & \mu \in \mathbb{R} \end{array}$$

deswegen genügt: $\beta_c(\mu_0) \leq \alpha$

$$\beta_c(\mu_0) \leq \alpha \Leftrightarrow 1 - \Phi \underbrace{\left(\sqrt{n} \frac{c - \mu_0}{\sigma_0} \right)}_{= z_{1-\alpha}} \leq \alpha \Leftrightarrow c \geq \frac{z_{1-\alpha}\, \sigma_0}{\sqrt{n}} + \mu_0 =: c^*$$

Definition 16.3
Gegeben sei ein Testproblem zum Niveau und D_α eine Menge von Tests zum Niveau α.
*$\varphi^* \in D_\alpha$ heißt **gleichmäßig bester Test** in D_α, falls*

$$\forall \, \theta \in \Theta_1 : \beta^*(\theta) = P_\theta(\varphi^*(x) = 1) = \max_{\varphi \in D_\alpha} P_\theta(\varphi(x) = 1)$$

Zurück zum Beispiel:
sei D_α die Menge aller Tests der Form

$$\varphi(x) = \begin{cases} 0 & , \quad \text{falls} \quad \overline{x} \leq c \\ 1 & , \quad \text{falls} \quad \overline{x}\ c \end{cases} \quad \text{und } c \geq c^*$$

I.A. findet man keinen gleichmäßig besten Test.

$$\varphi(x) = \begin{cases} 0 & , \quad \text{falls} \quad \overline{x} \leq c^* \\ 1 & , \quad \text{falls} \quad \overline{x}\ c^* \end{cases}$$

ist gleichmäßig bester Test in D_α, denn:

$$\beta^*_c(\mu_0) = \alpha\beta_{c^*+h}(\mu) = \beta^*_c(\mu - h) \leq \beta^*_c(\mu)\ \forall\ \mu \in \mathbb{R}, h \geq 0$$

Bemerkung 16.2 (Wahl der Nullhypothese)
Wird H_0 verworfen, so hat man eine Sicherheitswahrscheinlichkeit von $1-\alpha$ für die Alternative. Möchte man sich zum Beispiel für $\theta < \theta_0$ entscheiden, sollte man die Nullhypothese $H_0 : \theta \geq \theta_0$ wählen.

Definition 16.4
Seien $X_0, X_1, \ldots, X_r$ unabhängig und identisch verteilte ZV mit $X_i \sim N(0,1), r \in \mathbb{N}$. Dann heißt die Verteilung von

a)

$$\sum_{i=1}^{r} X_i^2$$

*eine χ^2-**Verteilung** mit r Freiheitsgraden (Schreibweise: χ^2_r)*

b)

$$\frac{X_0}{\sqrt{\frac{1}{r}\sum_{i=1}^{r} X_i^2}}$$

*eine **t-Verteilung** mit r Freiheitsgraden (Schreibweise: t_r)*

Satz 16.1
Es sei $X = (X_1, \ldots, X_n)$ eine Zufallsstichprobe zur $N(\mu, \sigma^2)$-Verteilung. Dann gilt:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{ und } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ unabhängig und}$$

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n}) \text{ und} \frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

Beweis Es sei $Y_i := \frac{X_i - \mu}{\sigma}$ $i = 1, \ldots, n$
Vor. $\Rightarrow Y_1, \ldots, Y_n$ unabhängig und identisch verteilt mit $Y_i \sim N(0,1)$
Sei $Y = (Y_1, \ldots, Y_n)$.

$$f_y(y_1, \ldots, y_n) = (\frac{1}{\sqrt{2\pi}})^n e^{-\frac{1}{2}\sum_{i=1}^n Y_i^2}$$

Sei jetzt $a \in \mathbb{R}^{n \times n}$ eine orthogonale Matrix (d.h. $A^{-1} = A^T$) mit

$$a_{nj} = \frac{1}{\sqrt{n}} \text{ , } j = 1, \ldots, n$$

Sei $Z := A \cdot Y$ und $a, b \in \mathbb{R}^n$ mit $a_i \leq b_i$

$$\begin{aligned}
P(Z \in \underbrace{(a,b]}_{=(a_1,b_1]\times\cdots\times(a_n,b_n]}) &= P(Y \in A^{-1}(a,b]) \\
&= \int_{A^{-1}(a,b]} f_y(y_1, \ldots, y_n) dy_1 \cdots dy_n \\
&\stackrel{\text{Substitutionsregel}}{=} \int_{(a,b]} f_y(z_1, \ldots, z_n) dz_1 \cdots dz_n \\
&= P(Y \in (a,b]) \text{ (wegen } \|y\|^2 = \|Ay\|^2, detA = 1)
\end{aligned}$$

$\Rightarrow Z \stackrel{\mathrm{d}}{=} Y$ und $Z_1, \ldots, Z_n$ unabhängig.

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i = \sqrt{n} \cdot \overline{Y} = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu)$$

$$\begin{aligned}
(n-1)S^2(x) &= \sum_{i=1}^n (X_i - \overline{X})^2 = \sigma^2(n-1)S^2(Y) = \sigma^2 \sum_{i=1}^n (Y_i - \overline{Y})^2 \\
&= \sigma^2 \sum_{i=i}^n (Y_i^2 - 2Y_i\overline{Y} + \overline{Y}^2) = \sigma^2 \left(\sum_{i=1}^n Y_i^2 - n\overline{Y}^2 \right) \\
&= \sigma^2 (\underbrace{\|Y\|^2}_{=\|Z\|^2} - Z_n^2) = \sigma^2 (Z_1^2 + \cdots + z_{n+1}^2)
\end{aligned}$$

$\overline{X}$ und $S^2(X)$ sind unbhängig, da $\overline{X}$ nur von Z_n und $S^2(x)$ nur von $Z_1, \ldots, Z_{n-1}$ abhängt. $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$ klar und $\frac{(n-1)S^(X)}{\sigma^2} = (n-1)S^2(Y) = Z_1^2 + \cdot + Z_{n-1}^2 \sim \chi^2_{n-1}$ nach Definition der χ^2-Verteilun. ■

Korollar 16.2 *Unter der Voraussetzung von Satz 16.1 gilt:*

$$\sqrt{n} \cdot \frac{(\overline{X} - \mu)}{S} \sim t_{n-1}$$

Beweis Weil $\overline{X}$ und $S^2(X)$ unabhängig sind, sind auch die Zufallsvariablen

$$\sqrt{n} \left(\frac{\overline{X} - \mu}{\sigma} \right) \text{ und } \sqrt{\frac{1}{(n-1)} \frac{(n-1)S^2(X)}{\sigma^2}} \text{ unabhängig.}$$

Also gilt:

$$\frac{\sqrt{n}(X-\mu)}{S(X)} = \frac{\overbrace{\sqrt{n}\cdot\frac{\overline{X}-\mu}{\sigma}}^{\sim N(0,1)}}{\sqrt{\frac{1}{n-1}\underbrace{\frac{(n-1)S^2(X)}{\sigma^2}}_{\sim\chi^2_{n-1}}}} \sim t_{n-1}$$

16.3 Test auf den Mittelwert bei unbekannter Varianz

$P_{\mu,\sigma^2} = N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2) \in \Theta = \mathbb{R}\times\mathbb{R}_+$. σ^2 ist hier nicht bekannt!
Einseitiges Testproblem: $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ für ein festes $\mu_0 \in \mathbb{R}$
Sei $X = (X_1,\dots,X_n)$ eine Zufallsstichprobe zu P_{μ,σ^2}. Dann gilt: $\sqrt{n}\,\frac{\overline{X}-\mu}{S} \sim t_{n-1}$
Als Test verwenden wir:

$$\varphi(x) = \varphi(x_1,\dots,x_n) = \begin{cases} 1, & \text{falls } \overline{x} \leq \frac{1}{\sqrt{n}} \overbrace{t_{n-1}(1-\alpha)}^{(1-\alpha)\text{-Quantil der } t_{n-1}\text{-Verteilung}} S(x) + \mu_0 \\ 0, & \text{falls } \overline{x} > \cdots \end{cases}$$

Der Test φ hält das Niveau α ein: Sei $\mu \leq \mu_0$

$$\begin{aligned} \beta_\varphi(\mu) &= P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X}-\mu_0}{S(X)} > t_{n-1}(\alpha-1)\right) \\ &\leq P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X}-\mu}{S(X)} > t_{n-1}(\alpha-1)\right) \\ &= 1 - P_{\mu,\sigma^2}\left(\sqrt{n}\cdot\frac{\overline{X}-\mu}{S(X)} \leq t_{n-1}(\alpha-1)\right) = 1-(1-\alpha) = \alpha \end{aligned}$$

Zweiseitiges Testproblem: $H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$
Der zugehörige Test ist:

$$\varphi(x) = \varphi(x_1,\dots,x_n) \begin{cases} 0, & \text{falls } \sqrt{n}\left|\frac{\overline{x}-\mu_0}{S(x)}\right| \leq t_{n-1}(1-\frac{\alpha}{2}) \\ 1, & \text{falls } \sqrt{n}\left|\frac{\overline{x}-\mu_0}{S(x)}\right| > t_{n-1}(1-\frac{\alpha}{2}) \end{cases} \text{hat Niveau } \alpha$$

■

Beispiel 16.2 (Zweiseitiger t-Test) *Dieses Beispiel wurde nur in der Hannoverschen Vorlesung gezeigt.*
Einer früheren Untersuchung zur Folge sind Jungen einer bestimmten Altersgruppe im Mittel $\sigma_0 = 150cm$ groß. Ein Hersteller für Kinderbekleidung möchte feststellen, ob sich seit der letzten Untersuchung eine Veränderung ergeben hat. Dazu wird die Größe von $n = 49$ zufällig ausgesuchten Jungen des entsprechenden Alters gemessen: Es ergibt sich: $\overline{X} = 147$ und $S^2 = \frac{1}{(49-1)}\sum_{i=1}^{49}(X_i - \overline{X})^2 = 36$.
Als Niveau wird $\alpha = 0.05$ gewählt.
Annahme: Körpergröße normalverteilt $N(\theta,\sigma^2)$

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

$$\sqrt{n}\,|\frac{\overline{X}-\theta_0}{S}| = |\frac{147-150}{6}| \cdot 7 = 3,5$$

Aus Tabelle: $t_{48}(1-\frac{\alpha}{2}) = t_{48}(0.975) \approx 2.01$
Ablehnung von H_0 ist auf dem Niveau $\alpha = 0.05$ gesichert.
Bemerkung:
Die angegebenen t-Tests sind unverfälscht und trennscharf in Menge der unverfälschten Tests.

16.4 Test auf die Varianz

$P_{\mu,\sigma^2} = N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. wie gehabt.
Testproblem: $H_0 : \sigma^2 \leq \sigma_0^2$ vs. $H_1 : \sigma^2 > \sigma_0^2$
Nach Satz 16.1 gilt:

$$P_{\mu,\sigma^2}(\frac{(n-1)S(X)^2}{\sigma_0} \leq x) = F_{\chi^2_{n-1}}(x)$$

Als Test verwenden wir:

$$\varphi(x) = \varphi(x_1 \ldots, x_n) = \begin{cases} 0, & \text{falls } \frac{(n-1)S(x)^2}{\sigma_0} \leq \chi^2_{n-1}(1-\alpha) \\ 1, & \text{falls } \frac{(n-1)S(x)^2}{\sigma_0} > \chi^2_{n-1}(1-\alpha) \end{cases}$$

Berechnung des Testniveaus: Sei $\sigma^2 \leq \sigma_0^2$

$$P_{\mu,\sigma^2}\left(\frac{(n-1)S(X)^2}{\sigma_0} > \chi^2_{n-1}(1-\alpha)\right) \leq P_{\mu,\sigma^2}\left(\frac{(n-1)S(X)^2}{\sigma} > \chi^2_{n-1}(1-\alpha)\right) = 1-(1-\alpha) = \alpha$$

Kapitel 17

Randomisierte Tests und das Lemma von Neyman-Pearson

Wir betrachten folgendes Testproblem: Sei $X \sim B(5, \theta)$ mit $\theta \in \Theta = \{\frac{1}{2}, \frac{3}{4}\}$

$$H_0 : \theta = \frac{1}{2} \text{ vs. } H_1 : \theta = \frac{3}{4}$$

Wir wollen einen Test der Form

$$\varphi(x) = \begin{cases} 0, & x \leq c \\ 1, & x > c \end{cases}$$

der das Niveau $\alpha = 0,05$ einhält.

$$\beta(\frac{1}{2}) = P_{\frac{1}{2}}(X > c) \overset{!}{\leq} 5$$

$$\begin{aligned} P_{\frac{1}{2}}(X = 5) &= \left(\frac{1}{2}\right)^5 &&= \frac{1}{32} < 0,05 \\ P_{\frac{1}{2}}(X \in \{4, 5\}) &= \left(\frac{1}{2}\right)^5 + \binom{5}{4} \cdot \left(\frac{1}{2}\right)^5 &&= \frac{6}{32} > 0,05 \end{aligned}$$

Das heißt der Test

$$\varphi(x) = \begin{cases} 0, & \text{falls } x \leq 4 \\ 1, & \text{falls } x = 5 \end{cases}$$

hält das Niveau σ ein. Leider wird das Signifikanzvieau nicht voll ausgeschöpft. $\Rightarrow$ mache bei $x = 4$ ein zusätzliches Experiment.

Definition 17.1
*Eine Funktion $\varphi : \chi^n \to [0, 1]$, die angibt mit welcher Wahrscheinlichkeit $\varphi(x)$ die Hypothese H_0 abgelehnt wird, heißt **randomisierter Test**. Die Gütefunktion wird jetzt definiert durch*

$$\beta(\theta) := E_\theta \, \varphi(X) \, .$$

Bemerkung 17.1
Die bisher betrachteten nicht randomisierten Tests sind ein Spezialfall:

$$\varphi(x) = \begin{cases} 0, & x \notin R \\ 1, & x \in R \end{cases}$$

Die Definition der Gütefunktion stimmt mit der bisherigen überein:

$$\beta(\theta) = E_\theta\, \varphi(X) = P_\theta(X \in R)$$

Im Beispiel: Wir lehnen H_0 jetzt auch im Falle $x = 4$ mit einer Wahrscheinlichekit p ab:

$$\varphi(x) = \begin{cases} 1, & x = 5 \\ \gamma, & x = 4 \\ 0, & x \in \{0,1,2,3\} \end{cases}$$

γ ist so zu bestimmen, dass der Test Niveau $\alpha = 0,05$ hat.

$$0,05 \stackrel{!}{=} \beta\left(\frac{1}{2}\right) = E_{\frac{1}{2}}\varphi(x) = 1 \cdot \frac{1}{32} + \gamma\binom{5}{4}\frac{1}{32} \Rightarrow \gamma = \frac{3}{25}$$

Im Folgenden betrachten wir den Spezialfall $\Theta = \{\theta_0, \theta_1\}$

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1$$

Definition 17.2
*Ein randomisierter Test $\varphi^* : \chi^n \to [0,1]$ heißt **Neyman-Pearson-Test**, wenn eine Konstante $c^* \in [0,\infty)$ und eine Funktioon $\gamma : \chi^n \to [0,1]$ gibt mit*

$$\varphi^*(x) = \begin{cases} 1 & \text{falls } L_x(\theta_1) > c^* L_x(\theta_0) \\ \gamma(x) & \text{falls } L_x(\theta_1) = c^* L_x(\theta_0) \\ 0 & \text{falls } L_x(\theta_1) < c^* L_x(\theta_0) \end{cases}$$

Satz 17.1 (Lemma von Neyman-Pearson)

a) Ist φ^* ein Neyman-Pearson-Test mit $\alpha = \beta_{\varphi^*}(\theta_0)$. Dann ist φ^* trennscharf unter allen Tests zum gleichen Niveau α, dass heißt er hat den kleinsten Fehler zweiter Art.

b) Für jedes $\alpha \in (0,1)$ existiert ein Neyman-Pearson-Test φ^* zum Niveau α. Dabei kann $\gamma(x) \equiv \gamma$ gewählt werden.

Beweis

a) Sei φ ein weiterer Test zum Niveau α. Zu zeigen: $1 - \beta_{\varphi^*}(\theta_1) \leq 1 - \beta_\varphi(\theta_1)$
Sei
$$A := \{x \in \chi^n | \varphi^*(x) > \varphi(x)\} \text{ und } B := \{x \in \chi^n | \varphi^*(x) < \varphi(x)\}$$

Es gilt:

$$\begin{aligned} x \in A \Rightarrow \varphi^*(x) > 0 \qquad & \Rightarrow L_x(\theta_1) \geq c^* L_x(\theta_0) \\ x \in B \Rightarrow \varphi^*(x) < 1 \qquad & \Rightarrow L_x(\theta_1) \leq c^* L_x(\theta_0) \end{aligned}$$

Also (wir betrachten nur den diskreten Fall):

$$\begin{aligned} \beta_{\varphi^*}(\theta_1) - \beta_\varphi(\theta_1) &= \sum_{x \in \chi^n} (\varphi^*(x) - \varphi(x))\, L_x(\theta_1) \\ &= \sum_{x \in A} (\varphi^*(x) - \varphi(x))\, L_x(\theta_1) + \sum_{x \in B} (\varphi^*(x) - \varphi(x))\, L_x(\theta_1) \\ &\geq \sum_{x \in A} (\varphi^*(x) - \varphi(x))\, c^* L_x(\theta_0) + \sum_{x \in B} (\varphi^*(x) - \varphi(x))\, c^* L_x(\theta_0) \\ &= c^* \sum_{x \in \chi^n} (\varphi^*(x) - \varphi(x)) = c^* \left(\underbrace{\beta_{\varphi^*}(\theta_0)}_{=\alpha} - \underbrace{\beta_\varphi(\theta_0)}_{\leq \alpha} \right) \geq 0 \end{aligned}$$

b) Für $c \geq 0$ sei

$$\alpha(c) := P_{\theta_0}\left(\frac{L_X(\theta_1)}{L_X(\theta_0)} > c \right) \text{ sowie } \alpha(c^-) := P_{\theta_0}\left(\frac{L_X(\theta_1)}{L_X(\theta_0)} \geq c \right)$$

Sei $c^* := \inf\{c | \alpha(c) \leq \alpha\}$. Dann gilt: $\alpha(c^*) \geq \alpha \geq \alpha(c^*)$.

Sei außerdem

$$\gamma^* = \begin{cases} 0, & \text{falls } \alpha(c^*-) = \alpha(c^*) \\ \frac{\alpha - \alpha(c^*)}{\alpha(c^*-)\alpha(c^*)}, & \text{falls } \alpha(c^*-) > \alpha(c^*) \end{cases}$$

Dann gilt:

$$\begin{aligned} \beta_{\varphi^*}(\theta_0) &= E_{\theta_0} \varphi^*(X) \\ &= P_{\theta_0}\left(\frac{L_X(\theta_1)}{L_X(\theta_0)} > c^* \right) + \gamma^* \, P_{\theta_0}\left(\frac{L_X(\theta_1)}{L_X(\theta_0)} = c^* \right) + 0 \\ &= \alpha(c^*) + \gamma^*(\alpha(c^{*-}) - \alpha(c^*)) \\ &= \alpha \end{aligned}$$

Beispiel 17.1

Es sei $P_\theta \sim \text{Exp}(\theta)$ und $\Theta = \{\theta_0, \theta_1\}$ mit $\theta 0 < \theta_1$. Es ist

$$L_x(\theta) = \theta^n e^{-\theta \sum_{i=1}^n X_i} = \theta^n e^{-\theta n \overline{X}}$$

Betrachte

$$c^* < q(x) = \frac{L_X(\theta_1)}{L_x(\theta_0)} = \left(\frac{\theta_1}{\theta_0} \right)^n e^{n\overline{x}(\theta_0 - \theta_1)} =: q^*(\overline{x})$$

$q^*(\overline{x})$ ist fallen in $\overline{x}$. Also ist der zughörige Neyman-Pearson-Test äquivalent zu:

$$\varphi(x) = \begin{cases} 1, & \text{falls } \overline{X} < c^* \quad (\Leftrightarrow q(x) > \tilde{c}) \\ \gamma^*, & \text{falls } \overline{X} = c^* \quad (\Leftrightarrow q(x) = \tilde{c}) \\ 0, & \text{falls } \overline{X} > c^* \quad (\Leftrightarrow q(x) < \tilde{c}) \end{cases}$$

$\alpha(c) = P_{\theta_0}(\overline{X} < c) = P_{\theta_0}(\sum_{i=1}^n X_i < nc)$
Es ist $\sum_{i=1}^n X_i \sim \Gamma(n, \theta_0)$ für $\theta \in \Theta$. Offenbar ist $\alpha(c) = P_{\theta_0}(\overline{x} < c)$ stetig in c und damit $\gamma^* = 0$. c ist so zu wählen, dass

$$P_{\theta_0}\left(\sum_{i=1}^n X_i < nc^*\right) \stackrel{!}{=} \alpha$$

Wir betrachten jetzt wieder den Fall:

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0$$

Im allgemeinen können wir hier nicht wie vorher einen trennscharfen Test konstruieren. Es gibt aber Spezielfälle wo das klappt.

Definition 17.3
$\{f(x,\theta), \theta \in \Theta\}$ bzw. $\{p(x,\theta), \theta \in \Theta\}$ heißt ***Familie von (Zähl-)Dichten mit monotonen Dichtequotienten****, falls es eine messbare Funktion $T : \chi^n \to \mathbb{R}$ gibt, so dass*

$$q(x) = \frac{L_x(\theta_1)}{L_x(\theta_0)} = q^*(T(X_1, \ldots, X_n))$$

und q^* eine monotone Funktion in $T(x_1, \ldots, x_n) = T(x)$ ist $\forall \theta_0 < \theta_1$

Beispiel 17.2 (vgl. Beispiel 15.3)
Die Familie der Exponentialverteilungen erfüllt die Bedingung mit $T(x) = \overline{x}$.

Satz 17.2

a) *Sei $x \in \chi^n$ eine Zufallsstichprobe zu einer Verteilung mit monoton nicht fallendem Dichtequotienten in $T(x)$ Jeder Test der Form:*

$$\varphi(x) = \begin{cases} 1, & T(x) > t_0 \\ \gamma, & T(x) = t_0 \\ 0, & T(x) < t_0 \end{cases}$$

ist gleichmäßig bester Test für das Testproblem

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0$$

zum Niveau

$$\alpha = E_{\theta_0}(\varphi(X)) = \sup_{\theta \leq \theta_0} E_\theta(\varphi(X))$$

b) *Für jedes $\alpha \in (0,1)$ und $\theta_0 \in \Theta$ existiert ein Test wie in a) beschrieben.*

Kapitel 18

Likelihood-Quotienten Test

Gegeben sei ein allgemeines Testproblem:

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1$$

Definition 18.1 *Der **Likelihood-Quotient** ist definiert durch:*

$$q(x) = \frac{\sup_{\theta \in \Theta_0} L_x(\theta)}{\sup_{\theta \in \Theta_1} L_x(\theta)}$$

Ein Test der Form:

$$\varphi(x) = \begin{cases} 0, & q(x) > c_0 \\ \gamma, & q(x) = c_0 \\ 1, & q(x) < c_0 \end{cases}$$

*heißt **Likelihood-Quotienten Test**.*

Bemerkung 18.1
Der Neyman-Pearson-Test ist ein spezieller Likelihood-Quotienten-Test.

Beispiel 18.1 $P_{\mu,\sigma^2} = N(\mu, \sigma^2)$, $X = (X_1, \ldots, X_n)$ Zufallsstichprobe zu $N(\mu, \sigma^2)$.
$\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$.
Testproblem:

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0$$
$$\Theta_0 = \{(\mu, \sigma^2) | \mu = \mu_0, \sigma^2 \in \mathbb{R}_+\}$$
$$\Theta_1 = \{(\mu, \sigma^2) | \mu \neq \mu_0, \sigma^2 \in \mathbb{R}_+\}$$

Likelihood-Quotient:

$$\sup_{\theta \in \Theta_0} L_x(\theta) = \sup_{\sigma^2} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_0)^2\right)$$

Der Maximum-Likelihood-Schätzer für σ^2 (bei bekanntem $\mu = \mu_0$) ist:

$$\hat{\sigma}^2_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

$$\Rightarrow \sup_{\theta\in\Theta_0} L_x(\theta) = \frac{n^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} \left(\sum_{i=1}^n (x_i-\mu_0)^2\right)^{\frac{n}{2}}} e^{-\frac{n}{2}}$$

Analog:

$$\begin{aligned}
\sup_{\theta\in\Theta_1} L_x(\theta) &= \frac{n^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}} \left(\sum_{i=1}^n (x_i-\overline{x})^2\right)^{\frac{n}{2}}} e^{-\frac{n}{2}} \\
\Rightarrow q(x) &= \left(\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\sum_{i=1}^n (x_i-\mu-0)^2}\right)^{\frac{n}{2}} \\
&= (1+n\cdot T^*(x))^{-\frac{n}{2}} \text{ mit } T^*(x) = \frac{(\overline{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2}
\end{aligned}$$

$q(x)$ ist fallend in $T^*(x)$, das heißt der kritische Bereich ist:

$$\begin{aligned}
\{x|q(x)<c_0\} = \{x|T^*(x)>c'\} &= \{x|\frac{u(\overline{x}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2} > c'\} \\
&= \{x|\left|\sqrt{n}\cdot\frac{\overline{x}-\mu_0}{S(x)}\right| > \sqrt{n(n-1)c'}\}
\end{aligned}$$

Der Likelihood-Quotiententest ist also äquivalent zu folgendem Test:

$$\varphi(x) = \begin{cases} 0, & \text{falls } \left|\sqrt{n}\cdot\frac{\overline{x}-\mu-0}{S(x)}\right| \le \sqrt{n(n+1)c'} \\ 1, & \text{falls } \left|\sqrt{n}\cdot\frac{\overline{x}-\mu-0}{S(x)}\right| > \sqrt{n(n+1)c'} \end{cases}$$

Beachte:

$$\sqrt{n}\cdot\frac{\overline{X}-\mu-0}{S(x)} \sim t_{n-1}$$

Einstellen des Testniveaus:

$$\begin{aligned}
\beta(\mu_0) = E_{\mu_0}\varphi(X) &= P_{\mu_0}\left(\underbrace{\left|\sqrt{n}\cdot\frac{\overline{X}-\mu_0}{S(X)}\right|}_{=:z} > \underbrace{\sqrt{n(n-1)c'}}_{=:\tilde{c}}\right) \overset{!}{=} \alpha \\
&= P_{\mu_0}(z<-\tilde{c}) + P_{\mu_0}(z>\tilde{c}),\ z\sim t_{n-1}
\end{aligned}$$

Wichtig: Die Dichte der t_{n-1}-Verteilung ist symetisch zu 0

$$\begin{aligned}
&= 2(1-F_{t_{n-1}}(2)) \overset{!}{=} \alpha \\
&\Rightarrow \tilde{c} = t_{n-1}(1-\frac{\alpha}{2})
\end{aligned}$$

Stichwortverzeichnis