Wahrscheinlichkeitstheorie und Inferenz I

Prof. Dr. Volker Schmid
Wintersemester 2016/17

Institut für Statistik
Ludwig–Maximilians–Universität München

Stand: 14. November 2016

Nach einer Vorlage von Prof. Dr. Torsten Hothorn (siehe Lizenz S. 134).

Inhaltsverzeichnis

Kapitel 1

Einleitung

Beispiel 1.1 (Einfacher Münzwurf)**.** *Wir werfen eine Münze. Der Wurf hat zwei mögliche Ergebnisse: "Kopf" und "Zahl" (oder 0 und 1). Sie bilden die Ergebnismenge* $\Omega_1 = \{0, 1\}$*. Wir interessieren uns für die Wahrscheinlichkeit der Ergebnisse.* $P(0) =$*?* $P(1) =$*?*

Beispiel 1.2 (n-facher Münzwurf)**.** *Wir werfen die Münze n-mal. Ergebnisse sind n-Tupel der Form* $\omega = (1, 0, 0, \dots, 1)$*. Die Ergebnismenge* Ω_2 *besteht aus* 2^n *Elementen. Wir interessieren uns für die Wahrscheinlichkeit von Ereignissen, z.B.* $A := \{$*genau 2 Würfe sind 0*$\}$*. A enthält mehrere Ergebnisse, ist also eine Teilmenge von* Ω*. Was ist* $P(A)$*?*

Beispiel 1.3 (Einfacher Münzwurf II)**.** *Wir werfen eine Münze auf eine Gerade der Länge 1. Die Münze bleibt irgendwo zwischen 0 und 1 liegen. Mögliche Ergebnisse sind alle rationalen Zahlen zwischen 0 und 1. Wir interessieren uns für die Wahrscheinlichkeit der Ergebnisse. Z.B.* $P(0.4) =$*? Interessanter sind die Wahrscheinlichkeiten von Ereignissen, z.B.* $B := \{$*Münze liegt links von 0.4*$\}$*. Was ist* $P(A)$*?*

Was ist eigentlich Wahrscheinlichkeit? Lassen sich $P(A)$ in Bsp. 1.2 und Bsp. 1.3 vergleichen? Gibt es eine gemeinsame Definition und Interpretation?

Aus Statistik I/II sind bekannt:

a) $\mathbb{P} : \{A | A \subset \Omega\} \to [0, 1]$ "Wahrscheinlichkeit" mit Axiomen (K1), (K2), (K3) von Kolmogorov.

b) $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$ "Laplace-Wahrscheinlichkeit"

c) $X : \Omega \to \mathbb{R}$ mit $X^{-1}(A') = \{\omega | X(\omega) \in A'\} \in \mathcal{A}$
wobei A' ein "zulässiges" Ereignis und $\mathcal{A}$ eine "zulässige" Menge von Ereignissen.

d) $F(x) = \sum\limits_{i:x_i \leq x} f(x_i)$ diskrete Verteilungsfunktion

$F(x) = \int\limits_{-\infty}^{x} f(t)\,dt$ stetige Verteilungsfunktion

e) Gesetz der großen Zahlen:

$$\mathbb{P}\left(\left|\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - E(X_i)\right| \leq c\right) \to 1 \text{ für } n \to \infty$$

wobei $X_1, \ldots, X_n$ iid

f) Zentraler Grenzwertsatz:

$$\sum_{i=1}^{n} X_i \overset{a}{\sim} \mathrm{N}(n\mu, n\sigma^2), \mu = \mathbb{E}(X_i), \sigma^2 = \mathbb{V}(X_i)$$

wobei $X_1, \ldots, X_n$ iid

g) Zweidimensionale Normalverteilung

Es gibt verschiedene (philosophische) Wahrscheinlichkeitsauffassungen, z.B.

- die klassische Wahrscheinlichkeitsauffassung (De Moivre, Laplace),
- die frequentistische Interpretation (R.A. Fisher),
- die Bayesianische Interpretation,
- die Propensitätsinterpretation (Karl Popper).

Beispiel 1.4 (Bertrand-Paradoxon)**.** *"Wie groß ist die Wahrscheinlichkeit dafür, dass die Länge Y einer zufällig ausgewählten Sehne eines Kreises die Länge X der Seite eines in den Kreis eingeschriebenen Dreiecks übertrifft?"*

- *wähle zwei zufällig ausgewählte Endpunkte für die Sehne* $\to P(Y > X) = 1/3$.
- *wähle einen zufälligen Punkt im Kreis. Dieser bildet den Mittelpunkt der Sehne* $\to P(Y > X) = 1/4$.
- *wähle einen zufälligen Radius, darauf einen zufälligen Punkt, lege Sehne orthogonal zum Radius durch den Punkt* $\to P(Y > X) = 1/2$

Wir betrachten in dieser Veranstaltung nur die mathematische Definition der Wahrscheinlichkeit. Aus Statistik II sind die Axiome von Kolmogorov bekannt:

(K1) Für jedes Ereignis $A \subset \Omega$ ist die Wahrscheinlichkeit von A eine reelle Zahl zwischen 0 und 1: $0 \leq P(A) \leq 1$.

(K2) Das sichere Ereignis Ω hat die Wahrscheinlichkeit 1: $P(\Omega) = 1$.

(K3) Die Wahrscheinlichkeit der Vereinigung abzählbar vieler disjunkter Ereignisse ($A_i \cap A_j = \emptyset \; \forall i \neq j$) ist die Summe der Wahrscheinlichkeiten der einzelnen Ereignisse: $P\left(A_1 \cup A_2 \cup \cdots\right) = \sum P(A_i)$ (σ-Additivität).

Andrej Nikolaevic Kolmogorov (25. April 1903 bis 20. Oktober 1987). Studium der Mathematik, Geschichte und Metallurgie in Moskau, seit 1931 Professor für Mathematik. 1933 "Grundbegriffe der Wahrscheinlichkeitsrechnung", betreute insgesamt 79 Doktoranden. http://de.wikipedia.org/wiki/Andrei_Nikolajewitsch_Kolmogorow

Was passiert, wenn wir unendlich viele Ereignisse haben?

Weiter bekannt sind Zufallsvariablen, die sich als Funktion auf den Ereignisraum definieren lassen:

$$X : \Omega \to \mathbb{R} \text{ mit } X^{-1}(A) = \{\omega | X(\omega) \in A\},$$

wobei A ein "zulässiges" Ereignis ist (Was ist zulässig?). Wir kennen die diskrete Verteilungsfunktion $F(x) = \sum\limits_{i: x_i \leq x} f(x_i)$ und die stetige Verteilungsfunktion $F(x) = \int\limits_{-\infty}^{x} f(t)\, dt$.

Wir werden eine einheitliche Theorie kennen lernen, die die Axiome von Kolmogorov verallgemeinert und die mit einer Definition für stetige und diskrete (und gemischte) Verteilungen bzw. Zufallsvariablen auskommt.

Kapitel 2

Mengen von Ereignissen

Bezeichnungen:

- $\Omega \neq \emptyset$ heißt Basismenge oder Ergebnisraum.
- Eine Menge $A \subseteq \Omega$ heißt Ereignis.
- $\omega_i \in \Omega = \{\omega_1, \omega_2, \ldots\}$ heißt Elementarereignis oder Ergebnis.
- $\mathcal{P} = \{A | A \subset \Omega\}$, also die Menge aller Teilmengen der Basismenge Ω, heißt Potenzmenge.
- $\mathcal{F} \subseteq \mathcal{P}$, also eine Menge von Teilmengen von Ω, heißt Mengensystem.

Bemerkung: Zum Elementarereignis $\omega_i \underbrace{\in}_{\text{Element}} \Omega$ ist $\{\omega_i\} \underbrace{\subset}_{\text{Teilmenge}} \Omega$ ein Ereignis.

2.1 Mengen

Zur Schreibweise: $\mathcal{F}, \mathcal{A}, \mathcal{E}$ Mengensysteme und A, B Mengen.

Definition 2.1 (Mengenoperationen)**.** *Seien* $A, B, A_i \subset \Omega, i \in I$.

Gleichheit:	$A = B \quad :\Leftrightarrow \quad \forall \omega \in \Omega : \quad \omega \in A \Leftrightarrow \omega \in B$
Teilmenge:	$A \subseteq B \quad :\Leftrightarrow \quad \forall \omega \in \Omega : \quad \omega \in A \Rightarrow \omega \in B$
Schnitt:	$A \cap B \quad := \{\omega \in \Omega \mid (\omega \in A) \wedge (\omega \in B)\}$
	$\bigcap_{i \in I} A_i := \{\omega \in \Omega \mid \forall i \in I : \omega \in A_i\}$
Vereinigung:	$A \cup B \quad := \{\omega \in \Omega \mid (\omega \in A) \vee (\omega \in B)\}$
	$\bigcup_{i \in I} A_i := \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i\}$
	$\dot{\bigcup_{i \in I}} A_i := \{\omega \in \Omega \mid \exists i \in I : \omega \in A_i, A_i \cap A_j = \emptyset \text{ für } i \neq j\}$
Differenz:	$A \setminus B := \{\omega \in \Omega \mid (\omega \in A) \wedge (\omega \notin B)\}$
Komplement:	$\bar{A} := \{\omega \in \Omega \mid \omega \notin A\}$
Symmetrische Differenz:	$A \triangle B := (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$
Mächtigkeit:	$\lvert A \rvert :=$ *Anzahl Elemente von A*
Kardinalität:	$\lvert \mathbb{N} \rvert = \aleph_0$
	$\lvert \mathcal{P} \rvert = 2^{\lvert \Omega \rvert}$
Kartesisches Produkt:	$A \times B := \{(a, b) \mid a \in A \wedge b \in B\}$
	$\prod_{i \in I} A_i := \{(a_1, a_2, \ldots, a_{\lvert I \rvert}) \mid a_i \in A_i \forall i \in I\}$
	$A^k = \prod_{i=1,\ldots,k} A = \{(a_1, \ldots, a_k) \mid a_i \in A, i = 1, \ldots, k\}$

Interpretationen

$A \subset \Omega$	...	"A tritt ein", "erscheint", "wird realisiert"
$\bar{A} \subset \Omega$	...	"A tritt nicht ein", "Komplementärereignis"
$A_1 \cup A_2$	...	"A_1 oder A_2 treten ein"
$\bigcup_{i \in I} A_i$	...	"mindestens eines der A_i tritt ein"
$A_1 \cap A_2$	...	"A_1 und A_2 treten ein"
$\bigcap_{i \in I} A_i$	...	"alle A_i treten ein"
$A_1 \cap A_2 = \emptyset$	...	"A_1 und A_2 treten nicht gleichzeitig ein", "sind unvereinbar"
$A_1 \triangle A_2$	...	"Entweder A_1 oder A_2 tritt ein"
$A_1 = A_2$	...	"A_1 und A_2 beschreiben das gleiche Ereignis"
Ω	...	"Das sichere Ereignis"
$\bar{\Omega} = \emptyset$	...	"Das unmögliche Ereignis"

Gesetzmäßigkeiten

$A, B, C \subset \Omega$.

Reflexivität:	$A \subseteq A$
Asymmetrie:	$A \subseteq B$ und $B \subseteq A \Rightarrow A = B$
Transivität:	$A \subseteq B$ und $B \subseteq C \Rightarrow A \subseteq C$
	$\Rightarrow \mathcal{P}$ ist bezüglich $\subseteq$ partiell geordnet.
Kommutativgesetz:	$A \cup B = B \cup A$
	$A \cap B = B \cap A$
Assoziativgesetz:	$(A \cup B) \cup C = A \cup (B \cup C)$
	$(A \cap B) \cap C = A \cap (B \cap C)$
De Morgansche Regeln:	$\overline{A \cup B} = \bar{A} \cap \bar{B}$
	$\overline{A \cap B} = \bar{A} \cup \bar{B}$
Mächtigkeiten:	
Gleichmächtigkeit:	$\lvert A\rvert = \lvert B\rvert :\Longleftrightarrow \exists f : A \to B$ bijektiv
Addition von Mächtigkeiten:	$A \cap B = \emptyset \Longleftrightarrow \lvert A\rvert + \lvert B\rvert = \lvert A \cup B\rvert$
Multiplikation von Mächtigkeiten:	$\lvert A \times B\rvert = \lvert A\rvert \cdot \lvert B\rvert$

Augustus de Morgan (27. Juni 1806 bis 18. März 1871). Geboren in Indien, seit 1828 Professor für Mathematik am University College London. Gemeinsam mit George Boole Begründer der formalen Logik. Lehrer von Ada Lovelace (einzige eheliche Tochter Lord Byrons), der ersten Programmiererin (die Sprache 'Ada' ist nach ihr benannt).
http://de.wikipedia.org/wiki/Augustus_De_Morgan

Definition 2.2. *Wir betrachten Folgen von Teilmengen $A_1, A_2, \ldots, [A_n]$ von Ω.*

a) Die Menge aller $\omega \in \Omega$, die zu unendlich vielen der (A_n) gehört, heißt limes superior der Folge (A_n)

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

b) Die Menge aller $\omega \in \Omega$, die zu fast allen (A_n) gehört (allen bis auf endlich vielen), heißt limes inferior der Folge (A_n)

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m$$

Beispiel 2.1. *Eine Münze wird unendlich oft geworfen. Sei* $A_n :=$ *Menge aller Ergebnisse* ω *von Münzwurfen, die an der* n*-ten Stelle 1 ("Kopf") aufweisen.* A_n *ist "fast so groß" wie* Ω.

$$\bigcap_{m=1}^{\infty} A_m = \{(1,1,\dots)\}$$

$$\bigcup_{m=1}^{\infty} A_m = \Omega \setminus \{(0,0,\dots)\}$$

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \text{ alle Ereignisse, in denen die 1 unendlich oft vorkommt}$$

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \text{ alle Ereignisse, in denen die 0 höchstens endlich oft vorkommt}$$

Bemerkung: Die Folge (A_n) in Bsp. 2.1 konvergiert nicht.

2.2 Kombinatorik

Wir betrachten Indexmengen, also Teilmengen von $\{1,\dots,n\}, n,k \in \mathbb{N}$.

Definition 2.3 (Variationen (Beachtung der Reihenfolge))**.**

- *mit Wiederholung*

$$V^W kn = \{x = (x_1,\dots,x_k) | x \in \{1,\dots,n\}^k\}$$

- *ohne Wiederholung*

$$V^{oW} kn = \{x = (x_1,\dots,x_k) | x \in \{1,\dots,n\}^k ; x_i \neq x_j \forall i \neq j\}$$

Definition 2.4 (Kombinationen (ohne Beachtung der Reihenfolge))**.**

- *mit Wiederholung*

$$K^W kn = \{x = (x_1,\dots,x_k) | x \in \{1,\dots,n\}^k ; 1 \leq x_1 \leq \dots \leq x_k \leq n\}$$

- *ohne Wiederholung*

$$K^{oW} kn = \{x = (x_1,\dots,x_k) | x \in \{1,\dots,n\}^k ; 1 \leq x_1 < \dots < x_k \leq n\}$$

Beispiel 2.2.

Variation:	$(1,3,27,2) \neq (1,27,3,2)$
Kombination:	$(1,3,27,2) = (1,27,3,2)$

Bemerkung: $V^{oW}nn$ ist die Menge aller Permutationen der Zahlen $\{1,\dots,n\}$

Satz 2.1.

a) $|V^W kn| = n^k$

b) $|V^{oW} kn| = \frac{n!}{(n-k)!} \quad (k \leq n)$

c) $|K^W kn| = \binom{n+k-1}{k}$

d) $|K^{oW} kn| = \binom{n}{k} \quad (k \leq n)$

Beweis:

a) **(IA)** $|V^W 1n| = n$

(IV) Sei $|V^W (k-1)n| = n^{k-1}$

(IS) $|V^W kn| = |\{(x_1,\dots,x_{k-1},y) | x \in V^W(k-1)n, y \in \{1,\dots,n\}\}| = n^{k-1}n = n^k$ (Multiplikation von Mächtigkeiten, S. 6)

b) **(IA)** $|V^{oW} 1n| = n = \frac{n!}{(n-1)!}$

(IV) $|V^{oW}(k-1)n| = \frac{n!}{(n-(k-1))!}$

(IS)
$$\begin{aligned} |V^{oW}kn| &= |\{(x_1,\dots,x_{k-1},y) | x \in V^{oW}(k-1)n, y \in \{1,\dots,n\} \setminus \{x_1,\dots,x_{k-1}\}\}| \\ &= \frac{n!}{(n-(k-1))!} \cdot (n-(k-1)) \\ &= \frac{n!}{(n-k+1)!} \cdot (n-k+1) \\ &= \frac{n!}{(n-k)!} \end{aligned}$$

$|V^{oW}nn| = n!$ ($\sharp$ Permutationen von $\{1,\dots,n\}$)

d) Sei $x \in K^{oW}kn$. Sei $\pi \in V^{oW}kk$ eine Permutation von $1,\dots,k$ und $\pi(x)$ eine Permutation von x.

$$
\begin{aligned}
V^{oW}kn &= \bigcup_{x \in K^{oW}kn} \{\pi(x) | \pi \in V^{oW}kk\} \\
|V^{oW}kn| &= |\bigcup_{x \in K^{oW}kn} \{\pi(x) | \pi \in V^{oW}kk\}| \\
&= \sum_{x \in K^{oW}kn} |\{\pi(x) | \pi \in V^{oW}kk\}| \\
&= \sum_{x \in K^{oW}kn} |V^{oW}kk| = c \cdot k! \overset{!}{=} \tfrac{n!}{(n-k)!} \overset{b)}{=} |V^{oW}kn|
\end{aligned}
$$

$$\Leftrightarrow \quad c \cdot k! = \tfrac{n!}{(n-k)!}$$

$$\Leftrightarrow \quad c = \tfrac{n!}{(n-k)!k!} = \tbinom{n}{k}$$

c) $f : K^W kn \to K^{oW} kn + k - 1$ bijektiv
$x \mapsto f(x) = y = (x_1, x_2 + 1, \ldots, x_k + (k-1))$
$f^{-1}(y) = (y_1, y_2 - 1, \ldots, y_k - (k-1)) = x \in K^W kn$
$|K^W kn| = |K^{oW} kn + k - 1| = \binom{n+k-1}{k}$

□

Satz 2.2.

a) *Eine n-elementige Menge $A = \{a_1, \ldots, a_n\}$ besitzt genau $\binom{n}{k}$ k-elementige Teilmengen ohne Wiederholung und ohne Beachtung der Reihenfolge.*

b) *Seien $j_1, \ldots, j_k \in \mathbb{N}_0$; $\sum_{i=1}^{k} j_i = n$.*
Es gibt genau $\binom{n}{j_1, j_2, \ldots, j_k} = \frac{n!}{j_1! j_2! \cdots j_k!}$ Möglichkeiten, $\{1, \ldots, n\}$ in eine Folge von Mengen $M_1, \ldots, M_k$; $M_i \subset \{1, \ldots, n\}$, $|M_i| = j_i$ ohne Wiederholung und ohne Beachtung der Reihenfolge aufzuteilen.

Beweis:

a) $f : K^{oW} kn \to \{A_k | A_k \subset \{a_1, \ldots, a_n\}, |A_k| = k\}$
$x \mapsto f(x) = \{a_{x_1}, a_{x_2}, \ldots, a_{x_k}\}$ bijektiv
$\Rightarrow |\{A_k || A_k| = k\}| = |K^{oW} kn| = \binom{n}{k}$

b) Induktion nach k

(IA) $k = 2$; $j_2 = n - j_1$
$\overset{a)}{\Rightarrow} \binom{n}{j_1} = \frac{n!}{(n-j_1)! j_1!} == \binom{n}{j_1, j_2}$

(IV) Sei $l_1 = j_1, l_2 = j_2, \ldots, l_{k-1} = j_{k-1} + j_k$
Dann gibt es $\binom{n}{l_1, \ldots, l_{k-1}}$ Möglichkeiten einer Aufteilung in $M_1, \ldots, M_{k-1}$ mit $|M_i| = l_i$, $|M_{k-1}| = j_{k-1} + j_k$.

(IS) M_{k-1} kann auf $\binom{l_{k-1}}{j_{k-1}} = \binom{l_{k-1}}{j_{k-1}, j_k}$ Arten aufgeteilt werden.
$\Rightarrow \binom{n}{l_1, \ldots, l_{k-1}} \cdot \binom{l_{k-1}}{j_{k-1}, j_k} = \frac{n!}{j_1! j_2! \cdots j_{k-2}! (j_{k-1}+j_k)!} \cdot \frac{(j_{k-1}+j_k)!}{j_{k-1}! j_k!} = \binom{n}{j_1, \ldots, j_k}$

$\square$

Beispiel 2.3 (Urnenmodell)**.** *N Kugeln, K weiß, $N-K$ rot. Ziehe n Kugeln a) mit b) ohne Zurücklegen. Wie groß ist die Laplace-Wahrscheinlichkeit ($\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\sharp\, günstige}{\sharp\, mögliche}$ Ergebnisse) von Experimenten, bei denen genau k weiße Kugeln in der Stichprobe vom Umfang n enthalten sind?*

a) *Ziehen mit Zurücklegen:*

Nummeriere Kugeln $\underbrace{1, \ldots, K}_{weiß}, \underbrace{K+1, \ldots, N}_{rot}$
$\Omega = V^W nN$, *d.h., die Elementarereignisse sind alle n Nummern der Kugeln, welche potentiell aus den N Kugeln gezogen werden können.*
$A_k = \{x \in V^W nN | \sum_{i=1}^{n} I(x_i \leq K) = k\}$ ... *k weiße Kugeln*
Sei $B \subset \{1, \ldots, n\}$ eine Indexmenge ohne Wiederholungen mit $|B| = k$.
$A_k^{(B)} := \{x \in A_k | x_i \leq K \Leftrightarrow i \in B\}$ ... *genau k Kugeln mit Nummern $i \in B$ sind weiß (Achtung: $i \in \{1, \ldots, n\}$ ohne Wiederholung, aber $x_i \in \{1, \ldots, N\}$ mit Wiederholung).*
$A_k = \bigcup_B A_k^{(B)}$
$|A_k^{(B)}| = |A_k^{(\{1,\ldots,k\})}| = \underbrace{K \cdot K \cdot \ldots \cdot K}_{k\ weiß} \cdot \underbrace{(N-K) \cdot (N-K) \cdot \ldots \cdot (N-K)}_{(n-k)\ rot}$
Es gibt $\binom{n}{k}$ Mengen B ($j_1 = k, j_2 = n-k$)
$\Rightarrow |A_k| = \sum_B |A_k^{(B)}| = \binom{n}{k} K^k (N-K)^{n-k}$
$\mathbb{P}(A_k) = \frac{\binom{n}{k} K^k (N-K)^{n-k}}{N^n (=|\Omega|)} = \binom{n}{k} p^k (1-p)^{n-k}$ *mit* $p = \frac{K}{N}$
vgl. Binomialverteilung

b) *Ziehen ohne Zurücklegen:*
$\Omega = K^{oW} nN; \quad |\Omega| = \binom{N}{n}$
$\underbrace{x_1, \ldots, x_k}_{weiß \in K^{oW} kK}, \underbrace{x_{k+1}, \ldots, x_n}_{rot \in K^{oW} n-kN-K} \in K^{oW} nN$
$\Rightarrow \mathbb{P}(A_k) = \frac{|K^{oW} kK| \cdot |K^{oW} n-kN-K|}{\binom{N}{n} (=|\Omega|)} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$ *mit A_k aus a)*
vgl. Hypergeometrische Verteilung

Beispiel 2.4 (Geburtstagsparadoxon)**.** *Wie groß ist die Wahrscheinlichkeit, daß in einer Gruppe von k Personen mindestens zwei am gleichen Tag Geburtstag haben?*

$\Omega = \{x = (x_1, \ldots, x_k) | x \in \{1, \ldots, 365\}^k\} = V^W k365$

Annahmen:

- *keine Schaltjahre*
- *keine Zwillinge / Mehrlinge*
- *jeder Geburtstag ist gleich wahrscheinlich*

Unterschiedliche Geburtstage
$U = \{x \in \Omega | x_i \neq x_j \forall i \neq j\} = V^{oW} k365$
$|\Omega| = 365^k, \; |U| = \frac{365!}{(365-k)!}$

$\mathbb{P}(\textit{"mindestens zwei gleiche G'tage"}) = \mathbb{P}(\bar{U}) = 1 - \mathbb{P}(U) = 1 - \frac{|U|}{|\Omega|} = 1 - \frac{365!}{(365-k)!365^k} = 1 - \frac{364 \cdot \ldots \cdot (365-k+1)}{365^{k-1}}$

$k = 10 \quad \mathbb{P}(\bar{U}) = 0.12$
$k = 20 \quad \mathbb{P}(\bar{U}) = 0.41$
$k = 23 \quad \mathbb{P}(\bar{U}) = 0.51$
$k = 30 \quad \mathbb{P}(\bar{U}) = 0.71$
$k = 40 \quad \mathbb{P}(\bar{U}) = 0.89$
$k = 50 \quad \mathbb{P}(\bar{U}) = 0.97$

2.3 Mengensysteme

Zur Erinnerung: Mengensysteme nennen wir Mengen von Teilmengen.

Definition 2.5 (π-System)**.** *Ein Mengensystem $\mathcal{F} \subset \mathcal{P}$ heißt durchschnittsstabil, wenn gilt:*

$$A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}.$$

$\mathcal{F}$ heißt dann auch π-System.

Definition 2.6 (σ-Algebra)**.** *Ein Mengensystem $\mathcal{F} \subset \mathcal{P}$ heißt σ-Algebra über Ω oder abgeschlossenes Mengensystem, falls gilt:*

(S1) $\Omega \in \mathcal{F}$ *(Basismenge)*

(S2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$ *(Komplement)*

(S3) $A_i \in \mathcal{F}, i \in \mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ *(abzählbare Vereinigung)*

Beispiel 2.5. $1\times$ *Würfeln* $\quad \Omega_1 = \{1, \ldots, 6\}$
$2\times$ *Würfeln* $\quad \Omega_2 = \{(1,1), (1,2), \ldots, (6,6)\}$
$\vdots$
$n\times$ *Würfeln* $\quad \Omega_n = \Omega_1^n$
$\mathcal{F}_n = \mathcal{P}(\Omega_n)$ *(Potenzmenge) ist σ-Algebra, da* Ω_n *immer endlich*
Von Interesse sind z.B. Ereignisse $A = \{(1,1), (1,2), \ldots, (1,6)\} = \{\text{"1 im ersten Wurf"}\} \in \mathcal{F}_2$

Beispiel 2.6 (kleinste und größte σ-Algebra)**.** *Kleinste σ-Algebra:* $\{\emptyset, \Omega\}$
Größte σ-Algebra: $\mathcal{P}$

Beispiel 2.7. *Sei Ω überabzählbar und* $A \in \mathcal{F}$ *genau dann, wenn* A *oder* $\bar{A}$ *abzählbar sind. Dann ist* $\mathcal{F}$ *eine σ-Algebra.*

Definition 2.7 (erzeugte σ-Algebra)**.** *Ist* $A \subset \Omega$*, so heißt die Menge*

$$\sigma(A) = \{\emptyset, A, \bar{A}, \Omega\}$$

die von A erzeugte σ-Algebra und ist die kleinste σ-Algebra, die A enthält.

Satz 2.3 (Schnitt von σ-Algebren)**.** *Sei* $I \neq \emptyset$ *eine beliebige Indexmenge und* $\mathcal{F}_i$ *eine σ-Algebra über Ω für alle* $i \in I$*. Dann ist auch*

$$\mathcal{F} := \bigcap_{i \in I} \mathcal{F}_i$$

eine σ-Algebra über Ω.

Beweis: Nach Def. 2.6:

(S2) $A \in \mathcal{F} \Rightarrow A \in \mathcal{F}_i \quad \forall i \in I \overset{(S2)}{\Rightarrow} \bar{A} \in \mathcal{F}_i \quad \forall i \in I \Rightarrow \bar{A} \in \mathcal{F}$.

(S1) analog

(S3) analog

□

Definition 2.8 (Erzeuger). *Sei $\mathcal{E} \subset \mathcal{P}$ ein Mengensystem und Σ die Menge aller σ-Algebren über Ω, die $\mathcal{E}$ enthalten. Dann wird die σ-Algebra*

$$\sigma(\mathcal{E}) := \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$$

als die von $\mathcal{E}$ erzeugte σ-Algebra $\sigma(\mathcal{E})$ bezeichnet. Gilt umgekehrt für eine σ-Algebra $\mathcal{A}$

$$\sigma(\mathcal{E}) = \mathcal{A}$$

so heißt $\mathcal{E}$ Erzeuger von $\mathcal{A}$.

Beispiel 2.8. *(σ-Algebren)*

- *Sei $A \subset \Omega$, $\mathcal{E} = \{A\}$ (ein Mengensystem bestehend aus einer Menge). Dann ist*
$$\sigma(\mathcal{E}) = \sigma(A) = \{\emptyset, A, \bar{A}, \Omega\}.$$
- *Sei $\Omega = \{1, 2, \dots, 7\}$ und $\mathcal{E} = \{\{1,2\}, \{6\}\}$. Dann ist*
$$\sigma(\mathcal{E}) = \{\emptyset, \{1,2\}, \underbrace{\{3,4,5,6,7\}}_{\overline{\{1,2\}}}, \{6\}, \underbrace{\{1,2,3,4,5,7\}}_{\overline{\{6\}}}, \underbrace{\{1,2,6\}}_{\{1,2\}\cup\{6\}}, \underbrace{\{3,4,5,7\}}_{\overline{\{1,2\}\cup\{6\}}}, \Omega\}.$$
- *Sei $\mathcal{A}$ eine σ-Algebra über Ω. Dann ist $\sigma(\mathcal{A}) = \mathcal{A}$.*

Definition 2.9 (Borelsche σ-Algebra, Borelsche Mengen). *Sei $\Omega = \mathbb{R}$ und*

$$\mathcal{O} = \{]a,b[| a,b \in \mathbb{R}, a < b\}$$

das Mengensystem der offenen Intervalle von $\mathbb{R}$. Dann heißt

$$\mathcal{B}(\mathbb{R}) := \sigma(\mathcal{O})$$

die Borelsche σ-Algebra über $\mathbb{R}$; ihre Elemente $A \in \mathcal{B}(\mathbb{R})$ heißen Borelsche Mengen. Für $\Omega = \mathbb{R}^n$ setzen wir

$$\mathcal{O}^n = \{U \subset \mathbb{R}^n | U \text{ offen}\} \text{ und } \mathcal{B}^n := \mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{O}^n)$$

Satz 2.4 (Eigenschaften von $\mathcal{B}$).

i) $\emptyset \in \mathcal{B}, \mathbb{R} \in \mathcal{B}$

ii) $\forall a, b \in \mathbb{R}, a < b :$
$[a,b] \in \mathcal{B}, [a,b[\in \mathcal{B},]a,b] \in \mathcal{B}$

iii) $\{c\} \in \mathcal{B} \quad \forall c \in \mathbb{R}$

iv) $\mathbb{N} \in \mathcal{B}, \mathbb{Q} \in \mathcal{B}, \bar{\mathbb{Q}} \in \mathcal{B}$

Beweis:

i) $\mathcal{B}$ ist σ-Algebra $\overset{(S1)}{\Rightarrow} \quad \Omega = \mathbb{R} \in \mathcal{B}$

$\overset{(S2)}{\Rightarrow} \quad \bar{\Omega} = \emptyset \in \mathcal{B}$

ii) $$]a,b] = \underbrace{\bigcap_{m=1}^{\infty} \underbrace{\left]a, b + \frac{1}{m}\right[}_{\in \mathcal{O} \subset \mathcal{B}}}_{\text{(S3) abzählb. Vereinigung und abzählb. Schnitt } \in \mathcal{B}} \in \mathcal{B}$$

$$[a,b[= \underbrace{\bigcap_{m=1}^{\infty} \underbrace{\left]a - \frac{1}{m}, b\right[}_{\in \mathcal{O} \subset \mathcal{B}}}_{\text{(S3) abzählb. Vereinigung und abzählb. Schnitt } \in \mathcal{B}} \in \mathcal{B}$$

$$[a,b] = \underbrace{\overline{\underbrace{\underbrace{]-\infty, a[}_{\in \mathcal{O}} \cup \underbrace{]b, \infty[}_{\in \mathcal{O} \subset \mathcal{B}}}_{\text{abzählb. Vereinigung } \in \mathcal{B}}}}_{\text{Komplement } \in \mathcal{B}} \in \mathcal{B}$$

iii) $$\{c\} = \underbrace{\bigcap_{m=1}^{\infty} \underbrace{\left[c, c + \frac{1}{m}\right]}_{\in \mathcal{B} \text{ nach ii)}}}_{\text{abzählb. Schnitt } \in \mathcal{B}} \in \mathcal{B}$$

iv) $$\mathbb{N} = \underbrace{\bigcup_{i=1}^{\infty} \underbrace{\{i\}}_{\in \mathcal{B} \text{ nach iii)}}}_{\text{abzählb. Vereinigung } \in \mathcal{B}} \in \mathcal{B}$$

ebenso für $\mathbb{Q}$...

□

Beispiel 2.9 (Boule/Boccia)**.** 1 *Wurf* $\quad \Omega_1 = \mathbb{R}^2$
2 *Würfe* $\quad \Omega_2 = \mathbb{R}^2 \times \mathbb{R}^2$
$\vdots$
n *Würfe* $\quad \Omega_n = \prod_{i=1}^{n} \mathbb{R}^2$
$\mathcal{F}_n = (\mathcal{B}^2)^n$ *Borelsche σ-Algebra*
Interessierendes Ereignis: $A = \{(x,y) | x^2 + y^2 \leq r^2\} =$

$= \{1.$ *Wurf nicht mehr als* r *cm von* $(0,0)$ *entfernt*$\} \subset \Omega_1$
$A \in \mathcal{B}^2$.

Kapitel 3

Wahrscheinlichkeit als Maß

3.1 Mathematisches Maß

Definition 3.1 (meßbar)**.** *Sei $\mathcal{F}$ eine σ-Algebra über Ω und $A \in \mathcal{F}$. Dann heißt A meßbar bezüglich $\mathcal{F}$.*

Definition 3.2 (Meßraum)**.** *Sei $\mathcal{F}$ eine σ-Algebra über Ω. Das Tupel $(\Omega, \mathcal{F})$ heißt Meßraum.*

Beispiel 3.1. *Aufrufe einer Webseite pro Stunde*
$\Omega = \mathbb{N}$
$A = \{100, 101, \ldots, 10.000\} = \{$*"zwischen* 100 *und* 10.000 *Aufrufe"*$\} \subset \Omega$
$A \in \sigma(\mathbb{N})$ *ist meßbar.*

Definition 3.3 ((σ-endliches, endliches, normiertes) Maß)**.** *Sei $(\Omega, \mathcal{F})$ ein Meßraum. Eine Funktion $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ heißt Maß auf $\mathcal{F}$, falls die folgenden Bedingungen erfüllt sind*

(M1) $\mu(\emptyset) = 0$

(M2) $\mu(A) \geq 0 \quad \forall A \in \mathcal{F}$

(M3) *für jede Folge disjunkter Mengen $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ gilt*

$$\mu\left(\dot{\bigcup_{n=1}^{\infty}} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \quad (\sigma\text{-Additivität})$$

Gibt es eine Folge (A_n) von Mengen aus $\mathcal{F}$ mit $\bigcup_{n=1}^{\infty} A_n = \Omega$ und $\mu(A_n) < \infty \quad \forall n \in \mathbb{N}$, so heißt μ σ-endlich. Ist $\mu(\Omega) < \infty$, so heißt μ endlich. Ist $\mu(\Omega) = 1$, so heißt μ normiertes Maß.

Beispiel 3.2 (Dirac-Maß).

$$\delta_\omega : \mathcal{F} \to \mathbb{R} \quad mit \quad \delta_\omega(A) = I_A(\omega)$$

Paul Adrien Maurice Dirac (8. August 1902 bis 20. Oktober 1984). Mitbegründer der Quantenphysik und war 1925 zusammen mit Schrödinger Nobelpreisträger für Physik.
http://de.wikipedia.org/wiki/Paul_Dirac

Beispiel 3.3 (Zählmaß).

$$\mu_Z : \mathcal{F} \to \bar{\mathbb{R}}, \quad \mu_Z(A) := \begin{cases} |A| & falls\ A\ endlich \\ \infty & sonst \end{cases}$$

Beispiel 3.4 (Lebesgue-Maß).

$$\lambda : \mathcal{B} \to \mathbb{R}_0^+, \quad \lambda(]a,b[) = b - a$$

Definition 3.4 (Maßraum). *Ist* $(\Omega, \mathcal{F})$ *ein Meßraum und* $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ *ein Maß, so heißt das Tripel* $(\Omega, \mathcal{F}, \mu)$ *Maßraum.*

Beispiel 3.5. *Sei* Ω *überabzählbar und* $A \in \mathcal{F}$ *genau dann, wenn* A *oder* $\bar{A}$ *abzählbar ist*
$\Rightarrow \mathcal{F}$ *ist* σ*-Algebra (Beispiel 2.7)*

$$\mu : \mathcal{F} \to \bar{\mathbb{R}} \qquad \mu(A) := \begin{cases} 0 & A\ abz\ddot{a}hlbar \\ \infty & sonst \end{cases}$$

Dann ist $(\Omega, \mathcal{F}, \mu)$ *Maßraum.*

Bemerkung: Sei $A \in \mathcal{F}$ fest und höchstens abzählbar, dann ist $\mu_Z|_A(B) = |A \cap B|$ ein σ-endliches Maß und heißt reduziertes Zählmaß.

Satz 3.1 (Eigenschaften des Maßes). *Sei* $(\Omega, \mathcal{F}, \mu)$ *ein Maßraum und* $A, B, A_n \in \mathcal{F}, n \in \mathbb{N}$*. Dann gilt*

i) endliche Additivität: $A \cap B = \emptyset \Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$

ii) Subadditivität: $A \subset B$ *und* $\mu(A) < \infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$

iii) Monotonie: $A \subset B \Rightarrow \mu(A) \leq \mu(B)$

iv) Sub-σ*-Additivität:*

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

Beweis:

i) Betrachte Folge $A, B, \emptyset, \emptyset, \ldots \overset{\text{(M3)}}{\Rightarrow}$ i)

ii)/iii) $B = A \dot{\cup} (B \setminus A) \overset{\text{(M3)}}{\Rightarrow}$
$\mu(B) = \underbrace{\mu(A)}_{\geq 0} + \underbrace{\mu(B \setminus A)}_{\geq 0 \text{ nach (M2)}} \geq \mu(A) \Leftrightarrow$ iii) und
$\mu(B \setminus A) = \mu(B) - \mu(A) \Leftrightarrow$ ii) falls $\mu(A) < \infty$

iv) $\cup_{n=1}^{\infty} A_n = \dot{\bigcup}_{n=1}^{\infty} \left(A_n \setminus \cup_{k=1}^{n-1} A_k\right)$

$$\begin{aligned} \overset{\text{(M3)}}{\Rightarrow} \mu\left(\cup_{n=1}^{\infty} A_n\right) &= \textstyle\sum_{n=1}^{\infty} \mu\left(A_n \setminus \cup_{k=1}^{n-1} A_k\right) \\ &\overset{\text{ii)}}{\leq} \textstyle\sum_{n=1}^{\infty} \mu(A_n) \end{aligned}$$

□

Schreibweisen:

$(A_n)_{n \in \mathbb{N}}$

$A_n \uparrow A :\Leftrightarrow A_1 \subset A_2 \subset A_3 \subset \ldots$ und $\bigcup_{i=1}^{\infty} A_i = A$

$A_n \downarrow A :\Leftrightarrow A_1 \supset A_2 \supset A_3 \supset \ldots$ und $\bigcap_{i=1}^{\infty} A_i = A$

Satz 3.2 (Stetigkeit des Maßes μ)**.** *Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A, A_n \in \mathcal{F}, n \in \mathbb{N}$. Dann gilt*

i) Stetigkeit von unten: $A_n \uparrow A \Rightarrow \mu(A_n) \uparrow \mu(A)$

ii) Stetigkeit von oben: $A_n \downarrow A$ und $\mu(A_1) < \infty \Rightarrow \mu(A_n) \downarrow \mu(A)$

Beweis:

i) Sei $A_0 := \emptyset$. $A_n \uparrow A$ meint $A_1 \subset A_2 \subset \ldots$ und $\cup_{n=1}^{\infty} A_n = A$
Also ist $A = \dot{\bigcup}_{n=1}^{\infty} A_n \setminus A_{n-1}$ eine disjunkte Vereinigung

$$\begin{aligned} \Rightarrow \mu(A) &\overset{\text{(M3)}}{=} \textstyle\sum_{k=1}^{\infty} \mu(A_k \setminus A_{k-1}) \\ &= \lim_{n \to \infty} \mu\left(\cup_{k=1}^{n} A_k \setminus A_{k-1}\right) \\ &= \lim_{n \to \infty} \mu(A_n) \end{aligned}$$

ii) $A \subset A_n \subset A_1 \overset{\text{Monotonie}}{\Rightarrow} \mu(A) < \infty$
$\mu(A_n) < \infty \quad \forall n \in \mathbb{N}$
Wegen $A_n \downarrow A$ gilt $A_1 \setminus A_n \uparrow A_1 \setminus A$ und somit

$$\begin{aligned} \mu(A_1 \setminus A_n) &= \mu(A_1) - \mu(A_n) \qquad = \mu(A_1 \setminus A_n) \uparrow \mu(A_1 \setminus A) = \\ &= \mu(A_1) - \mu(A) \Rightarrow \mu(A_n) \downarrow \mu(A) \end{aligned}$$

□

3.2 Wahrscheinlichkeit als normiertes Maß

Jetzt Ω Ergebnisraum. Potentiell interessante Ereignisse sind Teilmengen von Ω, d.h., Elemente einer σ-Algebra $\mathcal{F}$, des Ereignisraumes.

Definition 3.5 (Wahrscheinlichkeitsmaß, Wahrscheinlichkeit)**.** *Sei $(\Omega, \mathcal{F})$ ein Meßraum und $\mathbb{P} : \mathcal{F} \to [0,1]$ ein Maß auf $(\Omega, \mathcal{F})$ mit $\mathbb{P}(\Omega) = 1$. Dann heißt $\mathbb{P}$ Wahrscheinlichkeitsmaß und ordnet jedem Ereignis $A \in \mathcal{F}$ die Wahrscheinlichkeit $\mathbb{P}(A)$ zu.*

Definition 3.6 (Wahrscheinlichkeitsraum, Verteilung)**.** $(\Omega, \mathcal{F}, \mathbb{P})$ *heißt Wahrscheinlichkeitsraum. $\mathbb{P}$ heißt auch Verteilung.*

Andrej Nikolaevic Kolmogorov (25. April 1903 bis 20. Oktober 1987). Studium der Mathematik, Geschichte und Metallurgie in Moskau, seit 1931 Professor für Mathematik. 1933 "Grundbegriffe der Wahrscheinlichkeitsrechnung", betreute insgesamt 79 Doktoranden. http://de.wikipedia.org/wiki/Andrei_Nikolajewitsch_Kolmogorow

Satz 3.3 (Elementare Rechenregeln)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, A_n \in \mathcal{F}$, $n \in \mathbb{N}$. Dann gilt*

i) $\mathbb{P}(\bar{A}) = 1 - \mathbb{P}(A)$

ii) $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

iii) Siebformel:
$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i+1} \sum_{1 \leq j_1 < \ldots < j_i \leq n} \mathbb{P}\left(\bigcap_{k=1}^{i} A_{j_k}\right)$$

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

v) Stetigkeit von unten: $A_n \uparrow A \Rightarrow \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$

vi) Stetigkeit von oben: $A_n \downarrow A \Rightarrow \mathbb{P}(A_n) \downarrow \mathbb{P}(A)$

Beweis: i) ii) iv) $\Leftarrow$ Satz 3.1
v) vi) $\Leftarrow$ Satz 3.2
iii) Übung. □

Beispiel 3.6. *Wir werfen einen (sechsseitigen) Würfel. Eine sinnvolle Ergebnismenge ist*
$$\Omega := \{1, 2, 3, 4, 5, 6\}.$$

Wir interessieren uns für das Ereignis $A :=$"die geworfene Zahl ist gerade". Die davon erzeugte σ-Algebra ist $\mathcal{F} = \{\emptyset, \{1,3,5\}, \{2,4,6\}, \Omega\}$. Wir folgen dem Prinzip vom unzureichenden Grund und nehmen an, dass alle ω gleich wahrscheinlich sind. Daraus lässt sich ein Wahrscheinlichkeitsmaß $P : \mathcal{F} \to [0,1]$ definieren:

- $P(\emptyset) = 0,$
- $P(\{1,3,5\}) = 0.5,$
- $P(\{2,4,6\}) = 0.5,$
- $P(\Omega) = 1.$

Definition 3.7 (Nullmenge)**.** *Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A \in \mathcal{F}$ mit $\mu(A) = 0$, so heißt A (μ-)Nullmenge.*

Definition 3.8 (μ-fast-überall)**.** *Die Eigenschaft E sei für die Elemente $\omega \in \Omega$ eines Maßraumes $(\Omega, \mathcal{F}, \mu)$ sinnvoll. E gilt (μ-)fast-überall, wenn E für alle $\omega \notin N \subset \Omega$ gilt und N eine μ-Nullmenge ist.*

Definition 3.9 ($\mathbb{P}$-fast sicher)**.**

$$\mathbb{P}\text{-fast überall} \Leftrightarrow \mathbb{P}\text{-fast sicher}$$

Beispiel 3.7. *Situation von Bsp. 1.3 (Mnzwurf auf eine Gerade). Das Ereignis $A = \{\text{Mnze liegt genau auf } 0.4\}$ ist eine Nullmenge;*

$$P(A) = 0.$$

A tritt also $\mathbb{P}$-fast sicher nicht ein.

Kapitel 4

Die Verteilungsfunktion

4.1 Verteilungsfunktion

Definition 4.1 (Verteilungsfunktion). *Ist $\mathbb{P} : \mathcal{B} \to [0,1]$ ein Wahrscheinlichkeitsmaß auf $\mathbb{R}$, so heißt*

$$\begin{aligned} F_{\mathbb{P}} : \mathbb{R} &\to [0,1] \\ x &\mapsto F_{\mathbb{P}}(x) := \mathbb{P}(]-\infty, x]) \end{aligned}$$

die Verteilungsfunktion von $\mathbb{P}$.

Satz 4.1 (Eigenschaften der Verteilungsfunktion). *Eine Verteilungsfunktion $F_{\mathbb{P}}$ ist*

i) monoton wachsend,

ii) rechtsstetig,

iii) und es gilt $\lim\limits_{x\to-\infty} F_{\mathbb{P}}(x) = 0; \quad \lim\limits_{x\to\infty} F_{\mathbb{P}}(x) = 1.$

Beweis:

i) $a \leq b \Rightarrow F_{\mathbb{P}}(b) - F_{\mathbb{P}}(a) = \mathbb{P}(]a, b]) \geq 0$

ii) $x_n \downarrow x \Rightarrow$
$\lim\limits_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim\limits_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz } 3.2}{=} \mathbb{P}(]-\infty, x]) = F_{\mathbb{P}}(x)$

iii) $x_n \downarrow -\infty \Rightarrow$
$\lim\limits_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim\limits_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz } 3.2}{=} \mathbb{P}(\emptyset) = 0$
$x_n \uparrow \infty \Rightarrow$
$\lim\limits_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim\limits_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz } 3.2}{=} \mathbb{P}(]-\infty, \infty[) = \mathbb{P}(\mathbb{R}) = 1$

□

Definition 4.2 (Verteilungsfunktion)**.** *Jede monoton wachsende, rechtsstetige Funktion*

$$F : \mathbb{R} \to [0,1] \quad mit$$

$$\lim_{x\to-\infty} F(x) = 0 \;\; und \quad \lim_{x\to\infty} F(x) = 1$$

heißt Verteilungsfunktion.

4.2 Korrespondenzsatz

Sei jetzt $\Omega = \mathbb{R}^n$ und $]a,b[:= \{x \in \mathbb{R} | a < x < b\}$ ein offenes, beschränktes Intervall. Die Borelsche σ-Algebra $\mathcal{B}$ ist die von $\mathcal{O} = \{]a,b[|a,b \in \mathbb{R}, a < b\}$ erzeugte σ-Algebra.

Satz 4.2 (Existenz und Eindeutigkeit des Lebesgue-Maßes)**.** *In jeder Dimension $n \in \mathbb{N}$ gibt es genau ein Maß*

$$\lambda^n : \mathcal{B}^n \to [0,\infty[,$$

sodaß für jedes n-dimensionale Intervall $]a,b] :=]a_1,b_1] \times \ldots \times]a_n,b_n] \subset \mathbb{R}^n$ gilt:

$$\lambda^n(]a,b]) = \prod_{i=1}^{n} (b_i - a_i)$$

λ^n heißt (n-dimensionales) Lebesgue-Maß. Weiterhin gibt es zu jeder rechtsseitig stetigen, monoton wachsenden Funktion $F : \mathbb{R} \to \mathbb{R}$ genau ein Maß

$$\lambda_F : \mathcal{B} \to \bar{\mathbb{R}},$$

sodaß für alle $]a,b] \subset \mathbb{R}$ gilt:

$$\lambda_F(]a,b]) = F(b) - F(a).$$

λ_F heißt Lebesgue-Stieltjes-Maß von F.

Beweis: Siehe Anhang A.1. in Meintrup and Schäffler [2005] □

Thomas Jean Stieltjes (29. Dezember 1856 bis 31. Dezember 1894). Seit 1873 Studium in Delft, dreimal durch Prüfungen gefallen (da er lieber die Arbeiten von Gauss und Jacobi las), danach Arbeit im Observatorium in Leiden, dessen Direktor ihm freie Zeit für mathematische Untersuchungen ließ. 1884 wegen "mangelnder Qualifikation" erfolglose Bewerbung auf eine Professur in Groningen. http://en.wikipedia.org/wiki/Thomas_Joannes_Stieltjes

Bemerkung: $\lambda := \lambda^1$. Ist $x \in \mathbb{R}$, so gilt

$$\lambda(\{x\}) \overset{\text{Satz 3.2}}{=} \lim_{n\to\infty} \lambda\left(\left]x-\frac{1}{n}, x\right]\right) = \lim_{n\to\infty} x - \left(x - \frac{1}{n}\right) = \lim_{n\to\infty} \frac{1}{n} = 0$$

$\lambda(A) = 0$ für jede abzählbare Teilmenge $A \subset \mathbb{R}$, z.B. $\lambda(\mathbb{N}) = 0$ wegen σ-Additivität $\lambda(\mathbb{N}) = \lambda\left(\cup_{i\in\mathbb{N}}\{i\}\right) = \sum_{i\in\mathbb{N}} \underbrace{\lambda(\{i\})}_{=0}$

$\lambda(\mathbb{R}) \geq \lambda(]0,n]) = n \quad \forall n \in \mathbb{N}$
$\Rightarrow \lambda(\mathbb{R}) = \infty$

Satz 4.3 (Translationsinvarianz des Lebesgue-Maßes)**.**

$$\lambda^n(A) = \lambda^n(A+v) \quad \forall v \in \mathbb{R}^n \quad A \in \mathcal{B}^n$$

Beweis:

$$\begin{aligned}\lambda^n(A+v) &= \prod_{i=1}^n (b_i+v_i) - (a_i+v_i) \\ &= \prod_{i=1}^n b_i - a_i = \lambda^n(A)\end{aligned}$$

□

Satz 4.4 (Maßeindeutigkeitssatz)**.** *Es seien μ und ν zwei Maße auf einem Meßraum $(\Omega, \mathcal{F})$ und $\mathcal{E}$ ein durchschnittsstabiler Erzeuger von $\mathcal{F}$ mit folgenden Eigenschaften:*

i) $\mu(E) = \nu(E) \quad \forall E \in \mathcal{E}$

ii) Es gibt eine Folge $(E_n), n \in \mathbb{N}$, disjunkter Mengen aus $\mathcal{E}$ mit $\mu(E_n) = \nu(E_n) < \infty$ und

$$\dot{\bigcup_{n=1}^{\infty}} E_n = \Omega.$$

Dann sind beide Maße identisch: $\mu = \nu$.

Satz 4.5 (Korrespondenzsatz)**.** *Für jede Verteilungsfunktion F ist $\mu := \lambda_F(]a,b]) = F(b) - F(a)$ (siehe Satz 4.2) ein Wahrscheinlichkeitsmaß mit $F_\mu = F$. Umgekehrt ist für jedes reelle Wahrscheinlichkeitsmaß $\mathbb{P}$ die Funktion $G := F_{\mathbb{P}}$ eine Verteilungsfunktion und $\lambda_G = \mathbb{P}$.*

Bemerkung: $\lambda_F(\mathbb{R}) = \lim_{n\to\infty} \lambda_F(]-\infty, n]) = \lim_{n\to\infty} F(n) = 1$
$\Rightarrow \lambda_F$ ist Wahrscheinlichkeitsmaß.

Beweis:

"$\Rightarrow$": F Verteilungsfunktion, $\mu = \lambda_F$
$F_\mu(x) = \lim_{n\to\infty} \mu(]-n, x]) = \lim_{n\to\infty} (F(x) - \underbrace{F(-n)}_{\to 0}) = F(x) \quad \forall x \in \mathbb{R}$

"$\Leftarrow$": $\mathbb{P}$ W'keitsmaß, $G = F_\mathbb{P}$
$\lambda_G(]a,b]) = G(b) - G(a)$
$= \mathbb{P}(]-\infty, b]) - \mathbb{P}(]-\infty, a])$
$= \mathbb{P}(]a,b]) \quad \forall \underbrace{]a,b], a < b \in \mathbb{R}}_{\pi\text{-System Erzeuger von } \mathcal{B}}$
$\Rightarrow \lambda_G$ und $\mathbb{P}$ stimmen auf einem durchschnittsstabilen Erzeuger von $\mathcal{B}$ überein
$\overset{\text{Satz 4.4}}{\Rightarrow} \mathbb{P} = \lambda_G$

□

Bemerkung: Wahrscheinlichkeitsmaße auf $\mathbb{R}$ sind durch Verteilungsfunktionen vollständig definiert.

Lemma 4.1. *Sei $\mathbb{P} : \mathcal{B} \to [0,1]$ ein Wahrscheinlichkeitsmaß und $F_\mathbb{P}$ seine Verteilungsfunktion. Dann gilt*

$$\mathbb{P}(\{x\}) = F_\mathbb{P}(x) - F_\mathbb{P}(x^-) \quad \forall x \in \mathbb{R}$$

mit $F_\mathbb{P}(x^-) = \lim_{t\uparrow x} F(t) \quad \forall x \in \mathbb{R}$ (linksseitiger Grenzwert)

Beweis: $\left]x - \frac{1}{n}, x\right] \downarrow \{x\} \overset{3.2}{\Rightarrow}$
$\mathbb{P}(\{x\}) = \lim_{n\to\infty} \mathbb{P}\left(\left]x - \frac{1}{n}, x\right]\right) =$
$= \lim_{n\to\infty} \left(F_\mathbb{P}(x) - F_\mathbb{P}\left(x - \frac{1}{n}\right)\right)$
$= F_\mathbb{P}(x) - F_\mathbb{P}(x^-)$ □

Korollar 4.1. *Sei $\mathbb{P} : \mathcal{B} \to [0,1]$ ein Wahrscheinlichkeitsmaß und $F_\mathbb{P}$ seine Verteilungsfunktion. Dann sind äquivalent*

i) $F_\mathbb{P}$ *ist stetig*

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

Kapitel 5

Meßbare Abbildungen auf Wahrscheinlichkeitsräumen

Definition 5.1 (Indikatorfunktion)**.** *Sei $A \subset \Omega$. Dann versteht man unter der Indikatorfunktion I_A von A die Abbildung $I_A : \Omega \to \mathbb{R}$,*

$$I_A(\omega) := \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{falls } \omega \notin A \end{cases}$$

Satz 5.1. *Seien A und B zwei Teilmengen von Ω.*

a) Es gilt $I_{\bar{A}} = 1 - I_A$.

b) Für die Vereinigung gilt $I_{A \cap B} = I_A \cdot I_B$.

c) Ist $A \cap B = \emptyset$, dann gilt $I_{A \cup B} = I_A + I_B$.

d) Ist $A \subset B$, dann gilt $I_{B \setminus A} = I_B - I_A$.

Wir definieren im Folgenden allgemein Abbildungen $f : \Omega_1 \to \Omega_2$ ber $f(\omega_1) = \omega_2$, wobei $\omega_1 \in \Omega_1$ und $\omega_2 \in \Omega_2$.

5.1 Meßbare Abbildungen

Sei $f : \Omega_1 \to \Omega_2$ eine Abbildung, so ist das Bild einer Menge $A \subset \Omega_1$

$$f(A) := \{f(\omega) \in \Omega_2 | \omega \in A\}$$

und das Urbild einer Menge $B \subset \Omega_2$

$$f^{-1}(B) := \{\omega \in \Omega_1 | f(\omega) \in B\}.$$

Dabei muss f nicht bijektiv sein! Es gilt für beliebige $B, B_i \in \Omega_2, i \in I$ ("Operationstreue")

$$\begin{aligned}
f^{-1}\left(\bigcup_{i\in I} B_i\right) &= \left\{\omega \in \Omega_1 | f(\omega) \in \bigcup_{i\in I} B_i\right\} \\
&= \bigcup_{i\in I} \{\omega \in \Omega_1 | f(\omega) \in B_i\} \\
&= \bigcup_{i\in I} f^{-1}(B_i) \\
f^{-1}\left(\bigcap_{i\in I} B_i\right) &= \bigcap_{i\in I} f^{-1}(B_i) \\
f^{-1}(\bar{B}) &= \{\omega \in \Omega_1 | f(\omega) \in \bar{B}\} \\
&= \Omega_1 \setminus \{\omega \in \Omega_1 | f(\omega) \in B\} \\
&= \Omega_1 \setminus f^{-1}(B) = \overline{f^{-1}(B)}
\end{aligned}$$

Bemerkung: f^{-1} heißt auch Urbildfunktion. Ist f bijektiv (dies ist genau dann der Fall, wenn $f^{-1}(\omega_2)$ genau ein Element $\omega_1 \in \Omega_1$ enthält für alle Elemente $\omega_2 \in \Omega_2$), so heißt f auch Umkehrfunktion oder inverse Funktion.

Beispiel 5.1 (Zweifachen Würfelwurf)**.** *Sei $f : \Omega \to \mathbb{R}$ die Abbildung $f(\omega) :=$ 'Anzahl der Augen beider Würfe'. Dann ist das Urbild $f^{-1}(\{3\}) = \{(1,2),(2,1)\}$.*

Satz 5.2. *Sei $\mathcal{F}_i$ eine σ-Algebra über $\Omega_i, i = 1,2$. Ist $f : \Omega_1 \to \Omega_2$ eine Abbildung, so ist $f^{-1}(\mathcal{F}_2)$ eine σ-Algebra über Ω_1 und $\{B \subset \Omega_2 | f^{-1}(B) \in \mathcal{F}_1\}$ eine σ-Algebra über Ω_2.*

Beweis: (S1), (S2), (S3) folgen aus der Operationstreue, etwa z.B.

$$\begin{aligned}
\text{(S2) } A \in f^{-1}(\mathcal{F}_2) &\Leftrightarrow A = f^{-1}(B), B \in \mathcal{F}_2 \\
&\Rightarrow \bar{B} \in \mathcal{F}_2 \\
&\Rightarrow \bar{A} = \overline{f^{-1}(B)} = f^{-1}(\bar{B}) \in f^{-1}(\mathcal{F}_2)
\end{aligned}$$

□

Definition 5.2 (meßbare Abbildung)**.** *Seien $(\Omega_1, \mathcal{F}_1)$ und $(\Omega_2, \mathcal{F}_2)$ zwei Meßräume. Eine Abbildung $f : \Omega_1 \to \Omega_2$ heißt $\underline{\mathcal{F}_1\text{-}\mathcal{F}_2\text{-meßbar}}$, falls*

$$f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1.$$

Von einer meßbaren Abbildung f spricht man, falls die involvierten σ-Algebren eindeutig bekannt sind.

Satz 5.3. *Seien* $(\Omega_i, \mathcal{F}_i), i = 1,2$ *zwei Meßräume, wobei* $\mathcal{F}_2 = \sigma(\mathcal{E})$ *von einem Mengensystem* $\mathcal{E}$ *erzeugt ist. Die Abbildung* $f : \Omega_1 \to \Omega_2$ *ist genau dann* $\mathcal{F}_1$*-*$\mathcal{F}_2$*-meßbar, wenn*

$$f^{-1}(\mathcal{E}) \subset \mathcal{F}_1.$$

Beweis:

"$\Rightarrow$" f $\mathcal{F}_1$-$\mathcal{F}_2$-meßbar

$$\begin{aligned} &\Leftrightarrow && f^{-1}(\mathcal{F}_2) = \{f^{-1}(B) | B \in \mathcal{F}_2\} \subset \mathcal{F}_1 \\ &\overset{\mathcal{E}\subset\sigma(\mathcal{E})=\mathcal{F}_2}{\Rightarrow} && f^{-1}(\mathcal{E}) = \{f^{-1}(B) | B \in \mathcal{E}\} \subset \mathcal{F}_1 \end{aligned}$$

"$\Leftarrow$" $\{B \subset \Omega_2 | f^{-1}(B) \in \mathcal{F}_1\}$ ist σ-Algebra über Ω_2 (Satz 5.2).
Wegen $f^{-1}(\mathcal{E}) \subset \mathcal{F}_1$ ist $\mathcal{E} \subset \{B \subset \Omega_2 | f^{-1}(B) \in \mathcal{F}_1\}$
und somit auch $\sigma(\mathcal{E}) = \mathcal{F}_2 \subset \{B \subset \Omega_2 | f^{-1}(B) \in \mathcal{F}_1\}$, denn $\sigma(\mathcal{E})$ ist der Schnitt aller σ-Algebren, die $\mathcal{E}$ enthalten.

□

Beispiel 5.2. *Sei* $(\Omega, \mathcal{F})$ *Meßraum, dann ist* $f : \Omega \to \mathbb{R}$ *genau dann meßbar, wenn*

$$f^{-1}(]-\infty, c]) = \{\omega \in \Omega | f(\omega) \leq c\} \in \mathcal{F} \ \forall c \in \mathbb{R},$$

denn $]-\infty, c]$ *bildet ein Erzeugendensystem von* $\mathcal{B}$ *(folgt aus Satz 2.4).*

Beispiel 5.3 (stetige Abbildungen)**.**

$$f : \mathbb{R} \to \mathbb{R} \ \textit{stetig} \ :\Leftrightarrow f^{-1}(O) \in \mathcal{O} \ \forall O \in \mathcal{O}$$

$\mathcal{O}$ *ist Erzeuger von* $\mathcal{B} = \sigma(\mathcal{O})$ *(Def. 2.9) und somit ist* f $\mathcal{B}$*-*$\mathcal{B}$*-meßbar.*

Beispiel 5.4 (Indikatorfunktion)**.** *Sei* $(\Omega, \mathcal{F})$ *Meßraum und* $A \subset \Omega$*.*

$$\begin{aligned} I_A : \Omega &\to \mathbb{R} \\ \omega &\mapsto I_A(\omega) := \begin{cases} 1 & \textit{für } \omega \in A \\ 0 & \textit{sonst} \end{cases} \\ I_A^{-1} \in \{\emptyset, \Omega, A, \bar{A}\} \subset \mathcal{F} &\Leftrightarrow A \in \mathcal{F} \end{aligned}$$

d.h. I_A *ist genau dann* $\mathcal{F}$*-*$\mathcal{B}$*-meßbar, wenn* $A \in \mathcal{F}$*. Deshalb heißt* A *meßbar, wenn* $A \in \mathcal{F}$ *gilt (Def. 3.1)*

Definition 5.3. *Sei* $\mathcal{F} \subset \mathcal{P}$ *ein Mengensystem:*

$$f^{-1}(\mathcal{F}) = \{f^{-1}(B) | B \in \mathcal{F}\}.$$

Beispiel 5.5 (Spur-σ-Algebra). *Sei $\mathcal{F}$ eine σ-Algebra über Ω und $A \subset \Omega$ und sei $f : A \to \Omega,\ f(a) = a$. Die σ-Algebra $f^{-1}(\mathcal{F}) = \{A \cap B | B \in \mathcal{F}\}$ heißt Spur-σ-Algebra von A, wird mit $\mathcal{F}|A$ bezeichnet und ist eine σ-Algebra über A (folgt aus Satz 5.2).*

Definition 5.4 (Bildmaß). *Ist $(\Omega_1, \mathcal{F}_1, \mu)$ ein Maßraum und $(\Omega_2, \mathcal{F}_2)$ Meßraum und $f : \Omega_1 \to \Omega_2$ meßbar, so heißt*

$$\mu_f : \mathcal{F}_2 \to [0, \infty[$$

$$B \mapsto \mu_f(B) = \mu(f^{-1}(B))$$

das Bildmaß μ_f von μ unter f und $(\Omega_2, \mathcal{F}_2, \mu_f)$ ist ein Maßraum.

Beispiel 5.6. $(\mathbb{R}, \mathcal{B}, \lambda)$ *und* $(\mathbb{R}, \mathcal{B})$
$f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + a$
$\lambda_f(A) = \lambda(f^{-1}(A)) = \lambda(A - a) = \lambda(A)\ \forall A \in \mathcal{B}$ *und λ_f ist Maß auf* $(\mathbb{R}, \mathcal{B})$.

Definition 5.5 (Einschränkung, Erweiterung). *Sei $f : \mathcal{F} \to \bar{\mathbb{R}}$ und $\mathcal{A} \subset \mathcal{F}$. Dann heißt die Funktion*

$$f|\mathcal{A} : \mathcal{A} \to \bar{\mathbb{R}} \quad \textit{mit} \quad f|\mathcal{A}(A) = f(A) \quad \forall A \in \mathcal{A}$$

Einschränkung von f auf $\mathcal{A}$ und f eine *Erweiterung von $f|\mathcal{A}$ auf $\mathcal{F}$.*

Bemerkung: $\lambda^n|_A(B) = \lambda^n(A \cap B)$ mit $A \in \mathcal{B}^n$ ist ein Maß auf $(\mathbb{R}^n, \mathcal{B}^n)$ und heißt Spurmaß.

Satz 5.4 (Meßbarkeit der Komposition). *Sind $(\Omega_i, \mathcal{F}_i), i = 1, 2, 3$ Meßräume und $f : \Omega_1 \to \Omega_2$ und $g : \Omega_2 \to \Omega_3$ meßbar. Dann ist auch $g \circ f : \Omega_1 \to \Omega_3$ meßbar.*

Beweis: $f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$ und $g^{-1}(\mathcal{F}_3) \subset \mathcal{F}_2 \Rightarrow$
$f^{-1}(\underbrace{g^{-1}(\mathcal{F}_3)}_{\subset \mathcal{F}_2}) \subset \mathcal{F}_1 \Leftrightarrow (g \circ f)^{-1}(\mathcal{F}_3) \subset \mathcal{F}_1$ $\square$

Bemerkung: Sei $(\Omega, \mathcal{F})$ ein Meßraum und $f = (f_1, \dots, f_n) : \Omega \to \mathbb{R}^n$ und $\mathcal{B}^n$ die Borelsche σ-Algebra über $\mathbb{R}^n$. Dann ist f $\mathcal{F}$-$\mathcal{B}^n$-meßbar, wenn $f_i, i = 1, \dots, n$ $\mathcal{F}$-$\mathcal{B}$-meßbar sind; siehe Satz 1.28 in Meintrup and Schäffler [2005].

Satz 5.5. *Es sei $(\Omega, \mathcal{F})$ ein Meßraum, $f, g : \Omega \to \mathbb{R}$ meßbare Funktionen und $\alpha, \beta \in \mathbb{R}$. Dann sind die Funktionen*

a) $\alpha f + \beta g$

b) $f \cdot g$

c) f/g *falls* $g(\omega) \neq 0 \quad \forall \omega \in \Omega$

ebenfalls meßbar.

Beweis:

$$\begin{aligned} h : \Omega &\rightarrow \mathbb{R}^2 \\ \omega &\mapsto h(\omega) = (f(\omega), g(\omega)) \text{ meßbar (siehe letzte Bemerkung)} \end{aligned}$$

a)

$$\begin{aligned} l : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto l(x, y) = \alpha x + \beta y \text{ stetig und somit meßbar} \end{aligned}$$

also ist $l \circ h = \alpha f + \beta g$ mit Satz 5.4 meßbar

b) $l(x, y) = x \cdot y$ meßbar weil stetig, weiter wie in a)

c) $l(x, y) = x/y$ meßbar weil stetig, weiter wie in a)

□

5.2 Zufallsvariablen

Definition 5.6 (Zufallsvariable)**.** *Ist* $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $(\Omega_2, \mathcal{F}_2)$ *ein Meßraum, so heißt eine* $\mathcal{F}_1$*-*$\mathcal{F}_2$ *meßbare Abbildung*

$$X : \Omega_1 \rightarrow \Omega_2$$

Zufallsvariable (ZV). Ist $\Omega_2 = \mathbb{R}$*, so heißt* X *reelle Zufallsvariable,* $\Omega_2 = \bar{\mathbb{R}}$ *numerische Zufallsvariable,* $\Omega_2 = \mathbb{R}^n$ *n-dimensionale reelle Zufallsvariable.*

Bemerkung: Eine Funktion $f : \Omega \rightarrow \mathbb{R}$ heißt reellwertig. Eine Funktion $f : \Omega \rightarrow \mathbb{R} \cup \{-\infty, \infty\} =: \bar{\mathbb{R}}$ heißt numerisch. Ist $(\Omega, \mathcal{F})$ ein Meßraum, so heißt f $\mathcal{F}$-$\bar{\mathcal{B}}$-meßbar, wenn $\{\omega \in \Omega | f(\omega) \leq c\} \in \mathcal{F} \quad \forall c \in \mathbb{R}$ bei $\bar{\mathcal{B}} = \sigma(\mathcal{B} \cup \{\infty\} \cup \{-\infty\})$
Schreibweise:

$$\begin{aligned} \{f \leq c\} &:= \{\omega \in \Omega | f(\omega) \leq c\} \\ \{f < c\} & \qquad \vdots \\ \{f = c\} & \qquad \vdots \end{aligned}$$

Bemerkung: Das Bildmaß $\mathbb{P}_X$ von $\mathbb{P}$ unter X

$$\mathbb{P}_X(A) = \mathbb{P}(\underbrace{X^{-1}(A)}_{\in \mathcal{F}_1}) \in [0,1] \quad \forall A \in \mathcal{F}_2$$

ist wieder ein Wahrscheinlichkeitsmaß.

Definition 5.7 (Verteilung einer Zufallsvariablen)**.** *Ist $(\Omega_1, \mathcal{F}_1, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\Omega_2, \mathcal{F}_2)$ ein Meßraum und $X : \Omega_1 \to \Omega_2$ eine Zufallsvariable, so heißt das Bildmaß $\mathbb{P}_X$ von $\mathbb{P}$ unter X Verteilung von X.*

Bemerkung: zur Schreibweise:

$$\begin{aligned}
\mathbb{P}_X(A) &= \mathbb{P}(X^{-1}(A)) \\
&= \mathbb{P}(\{\omega \in \Omega_1 | X(\omega) \in A\}) \\
&=: \mathbb{P}(X \in A) \\
\mathbb{P}_X(\{c\}) &= \mathbb{P}(X^{-1}(\{c\})) \\
&= \mathbb{P}(\{\omega \in \Omega_1 | X(\omega) = c\}) \\
&=: \mathbb{P}(X = c) \qquad \leq \geq < > \text{ genauso.}
\end{aligned}$$

Satz 5.6 (Funktionen von Zufallsvariablen)**.** *Sei $X : \Omega \to \mathbb{R}^n$ eine Zufallsvariable und $g : \mathbb{R}^n \to \mathbb{R}^m$ meßbar. Dann ist $g \circ X : \Omega \to \mathbb{R}^m$ ebenfalls eine Zufallsvariable.*

Beweis: $g \circ X$ ist als Komposition von meßbaren Funktionen $\mathcal{F}$-$\mathcal{B}$-meßbar Satz 5.4 und somit Zufallsvariable. □

Bemerkung:

- $aX + b$ ist ZV
- $X_1 + X_2$ ist ZV (wenn $X = (X_1, X_2)$ ZV und $g(X) = X_1 + X_2$))
- $X_1 \cdot X_2$ ist ZV, X_1/X_2 genauso
- X^2 ist ZV

  $\vdots$

  nach Satz 5.5

Bemerkung: zur Schreibweise:

$$\begin{aligned}
\mathbb{P}_{g(X)}(A) &= \mathbb{P}((g \circ X)^{-1}(A)) \\
&= \mathbb{P}(\{\omega \in \Omega | (g \circ X)(\omega) \in A\}) \\
&=: \mathbb{P}(g(X) \in A)
\end{aligned}$$

Beispiel 5.7. $\mathbb{P}(X^2 \leq c) = \mathbb{P}(\{\omega \in \Omega | (X(\omega))^2 \leq c\})$

Satz 5.7. *Sei* $(\Omega, \mathcal{F})$ *ein Meßraum und* $(f_n)_{n\in\mathbb{N}}$ *eine Folge meßbarer numerischer Funktionen*

$$f_n : \Omega \to \bar{\mathbb{R}}, n \in \mathbb{N}.$$

Dann sind $\sup\limits_n f_n$, $\inf\limits_n f_n$, $\limsup\limits_n f_n$ *und* $\liminf\limits_n f_n$ *meßbar.*

Beweis: $\sup_n f_n$:

$$(\sup_n f_n)^{-1}((-\infty, c]) = \{\omega \in \Omega | \sup_n f_n(\omega) \leq c\} = \underbrace{\bigcap_{n=1}^{\infty} \underbrace{\{\omega \in \Omega | f_n(\omega) \leq c\} \in \mathcal{F}}_{\text{nach Voraussetzung}}}_{\text{abzählb. Schnitt in } \mathcal{F} \text{ enthalten, weil } \mathcal{F}\ \sigma\text{-Algebra}}$$

$\inf\limits_n f_n = -\sup\limits_n (-f_n)$ Komposition $\Rightarrow$ meßbar nach Satz 5.4, denn

$-\sup\limits_n (-f_n) = (a \circ \sup\limits_n \circ a)(f_n)$ mit $a(f) = -f \quad \forall f \in M$

$\limsup\limits_n f_n = \inf\limits_{n\geq 1} (\sup\limits_{k\geq n} f_k)$ meßbar nach Satz 5.4

$\liminf\limits_n f_n = \sup\limits_{n\geq 1} (\inf\limits_{k\geq n} f_k)$ meßbar nach Satz 5.4 □

Kapitel 6

Lebesgue-Integral und Radon-Nikodym-Dichte

Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{R}$ eine $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion. Sei weiterhin $M := \{f | f : \Omega \to \bar{\mathbb{R}}, f \text{ meßbar}\}$ die Menge der meßbaren numerischen Funktionen und $M^+ := \{f | f \in M, f \geq 0\}$ die nicht-negativen Funktionen aus M.

6.1 Lebesgue-Integral

6.1.1 Lebesgue-Integral für Treppenfunktionen

Definition 6.1 (Treppenfunktion)**.** *Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{R}$ eine $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion mit endlichem Bild $f(\Omega) = \{y_1, \ldots, y_n\}$, so läßt sich f für ein $n \in \mathbb{N}$ und $A_i \in \mathcal{F}, i = 1, \ldots, n$, darstellen als*

$$f = y_1 I_{A_1} + \ldots + y_n I_{A_n} = \sum_{i=1}^{n} y_i I_{A_i}$$

und heißt Treppenfunktion. Desweiteren sei $T := \{f | f : \Omega \to \mathbb{R},\ f \text{ Treppenfunktion}\}$ und $T^+ := \{f | f \in T, f \geq 0\}$ die Menge der (nicht-negativen) Treppenfunktionen.

Bemerkung: $f \in T^+$ läßt sich darstellen als

$$f = \sum_{i=1}^{n} y_i I_{A_i} \qquad y_i \geq 0 \quad \forall i = 1, \ldots, n.$$

Definition 6.2 (Lebesgue-Integral für Treppenfunktionen)**.** *Das Integral für $f \in T^+$*

$$\int f\, d\mu := y_1\mu(A_1) + \ldots + y_n\mu(A_n)$$

heißt Lebesgue-Integral von f nach μ.

Beispiel 6.1. *Sei $\Omega = 0, 1$. Sei $(\Omega^2, \mathcal{P}(\Omega^2), \mathbb{P})$ ein Wahrscheinlichkeitsraum mit $\mathbb{P}(\omega) = 1/4 \,\forall\, \omega \in \Omega^2$. Sei $X : \Omega^2 \to \mathbb{R}$ eine Zufallsvariable mit $X(\omega) =$ 'Anzahl von 1 in ω'. Darstellung als Treppenfunktion:*

$$\begin{aligned} A_1 &= \{(0,0)\}; y_1 = 0; \\ A_2 &= \{(0,1),(1,0)\}; y_2 = 1; \\ A_3 &= \{(1,1)\}; y_3 = 2. \end{aligned}$$

Alternative Darstellung:

$$\begin{aligned} B_1 &= \{(0,1),(1,1)\}; y_1 = 1; \\ B_2 &= \{(1,0),(1,1)\}; y_2 = 1. \end{aligned}$$

Es gilt

$$\begin{aligned} \int X d\,\mathrm{Pr} &= 0 \cdot \mathbb{P}(A_1) + 1 \cdot \mathbb{P}(A_2) + 2 \cdot \mathbb{P}(A_3) \\ &= 1 \cdot \mathbb{P}(B_1) + 1 \cdot \mathbb{P}(B_2) = 1 \end{aligned}$$

Beispiel 6.2. *Sei $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1$ für $0 \le x \le 1$. Es gilt mit $A =]0, 1[$*

$$\int f d\lambda = 1 \cdot \lambda(A) = 1.$$

Henri Léon Lebesgue (28. Juni 1875 bis 26. Juli 1941). 1899–1902 Gymnasiallehrer in Nancy, dabei Arbeit am Integralbegriff, 1902 Dissertation, seit 1906 Professor.
http://de.wikipedia.org/wiki/Henri_L%C3%A9on_Lebesgue

Bemerkung: Besitzt $f \in T^+$ eine weitere Darstellung der Gestalt

$$f = \sum_{i=1}^{m} z_i I_{B_i} \quad \text{mit } B_i \in \mathcal{F}, z_i \ge 0 \quad \forall i = 1, \ldots, n$$

so gilt:

$$\sum_{i=1}^{n} y_i\mu(A_i) = \sum_{i=1}^{m} z_i\mu(B_i)$$

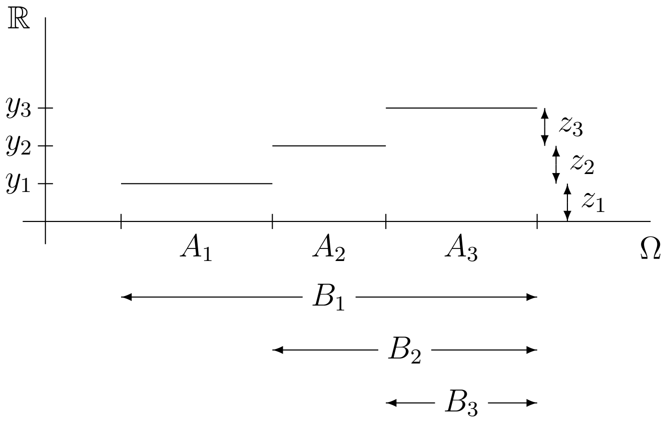


Abbildung 6.1: Zwei Treppenfunktionen stellen dieselbe Funktion dar.

(Übergang auf $C_{ij} = A_i \cap B_j$).

Daher ist $\int f\, d\mu$ wohldefiniert.

Bemerkung: Die Indikatorfunktion I_A ist die einfachste Treppenfunktion mit Integral

$$\int I_A\, d\mu = \mu(A).$$

Satz 6.1 (Eigenschaften des Integrals). *i) Linearität: Für $f, g \in T^+$ und $\alpha, \beta \geq 0$ gilt*

$$\int (\alpha f + \beta g)\, d\mu = \alpha \int f\, d\mu + \beta \int g\, d\mu.$$

ii) Monotonie: Sind $f, g \in T^+$ und $f \leq g$, so folgt

$$\int f\, d\mu \leq \int g\, d\mu.$$

Beweis: O.B.d.A. sei $f = \sum_{i=1}^n y_i I_{A_i}, \quad g = \sum_{i=1}^n z_i I_{A_i}$.

i) $\int (\alpha f + \beta g)\, d\mu = \int \alpha \sum_{i=1}^n y_i I_{A_i} + \beta \sum_{i=1}^n z_i I_{A_i} d\mu$
$= \int \sum_{i=1}^n (\alpha y_i + \beta z_i) I_{A_i} d\mu \overset{\text{Def. 6.2}}{=} \sum_{i=1}^n (\alpha y_i + \beta z_i)\mu(A_i)$
$= \alpha \sum_{i=1}^n y_i \mu(A_i) + \beta \sum_{i=1}^n z_i \mu(A_i) = \alpha \int f\, d\mu + \beta \int g\, d\mu$

ii) $f \leq g \Leftrightarrow f(x) \leq g(x) \quad \forall x \in \Omega$
$\Leftrightarrow y_i \leq z_i \quad \forall i = 1, \ldots, n$
$\Rightarrow \sum_{i=1}^n y_i \mu(A_i) \leq \sum_{i=1}^n z_i \mu(A_i)$
$\Leftrightarrow \int f\, d\mu \leq \int g\, d\mu$

□

Beispiel 6.3. $f \equiv 0 \equiv 0 \cdot I_{\mathbb{R}}$
$\int f d\lambda = \int 0 \cdot I_{\mathbb{R}} d\lambda = 0 \cdot \lambda(\mathbb{R}) = \underbrace{0 \cdot \infty = 0}_{\text{Konvention}}$

6.1.2 Lebesgue-Integral für nicht-negative Funktionen

Lemma 6.1. *Ist $f \in M^+$ eine nicht-negative Funktion, so gibt es eine Folge $(f_n)_{n\in\mathbb{N}} \in T^+$ von nicht-negativen Treppenfunktionen, so daß $f_n \uparrow f$.*

Beweis: Eine solche Funktion ist z. B. $f_n : \Omega \to \mathbb{R}, n \in \mathbb{N}$ mit

$$f_n(x) := \frac{\lfloor 2^n f(x) \rfloor}{2^n} \uparrow f(x) \quad \forall x \in \Omega$$

Etwa $f = \sin : [0, \pi] \to \mathbb{R}$ (siehe Abbildung 6.2).

□

Definition 6.3 (Lebesgue-Integral für nicht-negative Funktionen)**.** *Sei $f \in M^+$ und $(f_n)_{n\in\mathbb{N}} \in T^+$ mit $f_n \uparrow f$. Das Lebesgue-Integral von f nach μ ist dann*

$$\int f\, d\mu := \lim_{n\to\infty} \int f_n\, d\mu.$$

Lemma 6.2. *Sei $f \in M^+$ und $g \in T^+$ mit $g \leq f$. Dann gilt für eine Folge $(f_n)_{n\in\mathbb{N}} \in T^+$:*

$$f_n \uparrow f \Rightarrow \lim_{n\to\infty} \int f_n\, d\mu \geq \int g\, d\mu.$$

Beweis: o.B.d.A. $g = I_A, A \in \mathcal{F}$ (wegen der Linearität)

$\Rightarrow f(x) \geq 1 \quad \forall x \in A$
$\Rightarrow \quad \forall \epsilon > 0 : A_n := \{x \in A | f_n(x) \geq 1 - \epsilon\} \uparrow A$ wegen $f_n \uparrow f$ von unten!
$\Rightarrow f_n \geq (1 - \epsilon) I_{A_n} \qquad \forall n \in \mathbb{N}$

Und damit

$$\begin{aligned} \int f_n\, d\mu &\overset{\text{Monotonie}}{\geq} \int (1-\epsilon) I_{A_n}\, d\mu \overset{\text{Linearität Def.}}{=} (1-\epsilon)\mu(A_n) \overset{\text{Stetigkeit von } \mu}{\uparrow} (1-\epsilon)\mu(A) \\ &= (1-\epsilon) \int g\, d\mu \end{aligned}$$

Für $\epsilon \to 0$ folgt die Behauptung. □

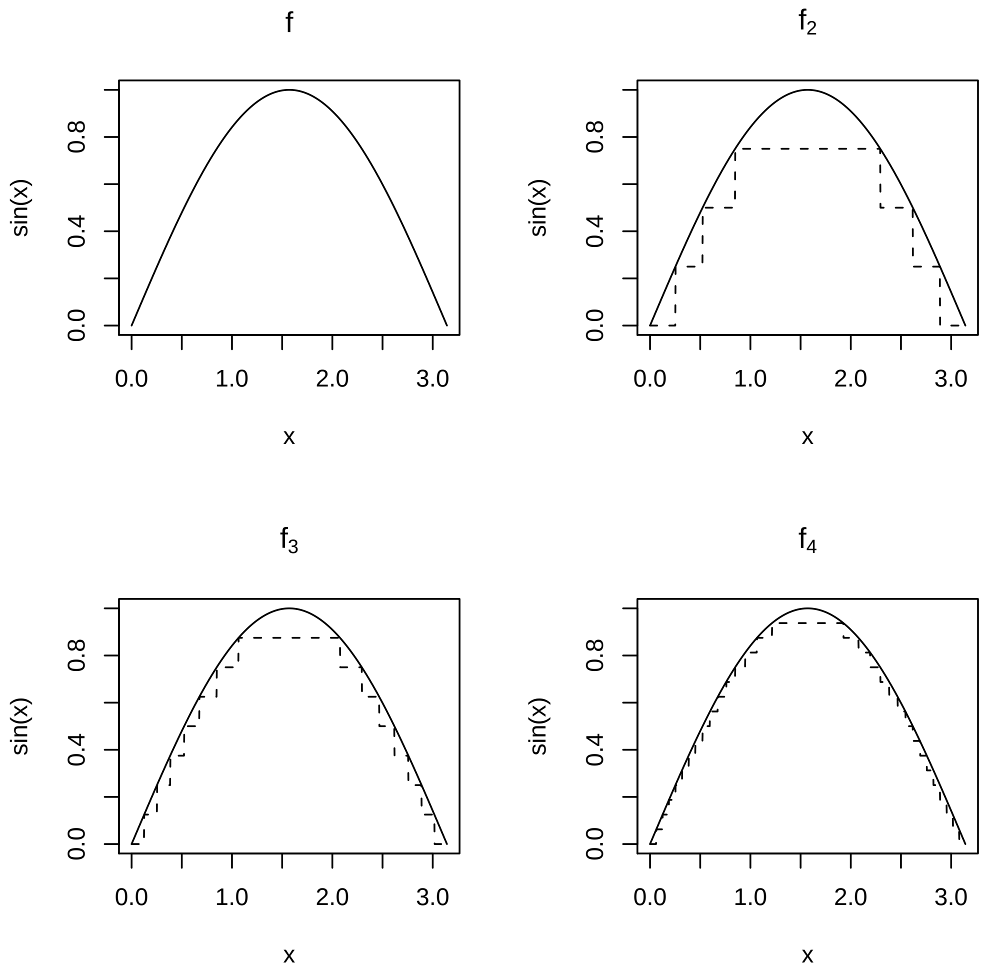


Abbildung 6.2: Approximation von $f = \sin \in M^+$ durch $f_n \in T^+$ für $n = 2, 3, 4$.

Lemma 6.3. *Sind $(f_n)_{n\in\mathbb{N}}$ und $(g_n)_{n\in\mathbb{N}}$ zwei monoton wachsende Folgen von Funktionen aus T^+ mit*

$$\lim_{n\to\infty} f_n = \lim_{n\to\infty} g_n,$$

so gilt:

$$\lim_{n\to\infty} \int f_n \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu.$$

Beweis: Laut Voraussetzung $\underbrace{g_k}_{\in T^+} \leq \lim\limits_{n\to\infty} f_n \quad \forall k \in \mathbb{N}$ (Monotonie)

Also $\int g_k \, d\mu \overset{\text{Lemma 6.2}}{\leq} \lim\limits_{n\to\infty} \int f_n \, d\mu \quad \forall k \in \mathbb{N}$
und somit $\lim\limits_{k\to\infty} \int g_k \, d\mu \leq \lim\limits_{n\to\infty} \int f_n \, d\mu$.
Umgekehrt: $f_k \leq \lim\limits_{n\to\infty} g_n$ und also $\lim\limits_{k\to\infty} \int f_k \, d\mu \leq \lim\limits_{n\to\infty} \int g_n \, d\mu$
Zusammen: $\lim\limits_{n\to\infty} \int f_n \, d\mu = \lim\limits_{n\to\infty} \int g_n \, d\mu$ □

Satz 6.2 (Eigenschaften des Integrals)**.** *Für $f, g \in M^+$ und $\alpha, \beta \geq 0$ gilt:*

i) Linearität:
$$\int (\alpha f + \beta g) \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu$$

ii) Monotonie: Ist $f \leq g$ so folgt
$$\int f \, d\mu \leq \int g \, d\mu.$$

Beweis: Sind $\underbrace{f_n}_{\in T^+} \uparrow f$, $\underbrace{g_n}_{\in T^+} \uparrow g$, so auch $\alpha f_n + \beta g_n \uparrow \alpha f + \beta g$

i) folgt somit aus Satz 6.1 (i) für f_n, g_n

ii) $f_n \leq \max(f_n, g_n) \uparrow g$ für $f \leq g$
und aus Satz 6.1 (ii) folgt ii)

□

6.1.3 Lebesgue-Integral für numerische Funktionen

Jetzt allgemein $f : \Omega \to \bar{\mathbb{R}}$ meßbar.

$$f = \underbrace{f^+}_{\in M^+} - \underbrace{f^-}_{\in M^+}$$

Definition 6.4 (quasi-integrierbar, Lebesgue-Integral)**.** *Sei $f : \Omega \to \bar{\mathbb{R}}$ eine meßbare numerische Funktion. Die Funktion heißt*
<u>(μ-) quasi-integrierbar</u>, falls $\int f^+ \, d\mu < \infty$ oder *$\int f^- \, d\mu < \infty$. Ist f <u>(μ-) quasi-integrierbar</u>, so ist durch*
$$\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$$
das <u>Lebesgue-Integral</u> von f definiert. Gilt $\int f^+ \, d\mu < \infty$ <u>und</u> $\int f^- \, d\mu < \infty$, so heißt f (μ-)integrierbar.

Bemerkung: $|f| = f^+ + f^-$
f integrierbar $\Leftrightarrow$ $|f|$ integrierbar.
Bemerkung:

$$\int_A f\, d\mu := \int f I_A\, d\mu.$$

Satz 6.3. *Sind $f, g : \Omega \to \bar{\mathbb{R}}$ integrierbare numerische Funktionen und $\alpha, \beta \in \mathbb{R}$, so gilt*

i) Linearität: $\alpha f + \beta g$ ist integrierbar und

$$\int (\alpha f + \beta g)\, d\mu = \alpha \int f\, d\mu + \beta \int g\, d\mu$$

ii) Monotonie: Ist $f \leq g$, so folgt

$$\int f\, d\mu \leq \int g\, d\mu$$

iii) Für jedes $A \in \mathcal{F}$ gilt

$$\int f\, d\mu = \int_A f\, d\mu + \int_{\bar{A}} f\, d\mu.$$

Beweis:

i), ii) $\checkmark$

iii) $f = f I_A + f I_{\bar{A}} \quad \checkmark$

□

Satz 6.4. *Für $f \in M^+$ gilt*

$$\int f\, d\mu = 0 \Leftrightarrow \underbrace{\{\omega \in \Omega | f(\omega) > 0\}}_{=: \text{ Träger von } f} \text{ ist } \mu\text{-Nullmenge}$$

Beweis: $A := \{\omega \in \Omega | f(\omega) > 0\}$
$A_n := \{\omega \in \Omega | f(\omega) > \frac{1}{n}\} \quad \forall n \in \mathbb{N}$
$A_n \uparrow A$

"⇒": $\int f\, d\mu = 0$
$\frac{1}{n} I_{A_n} \leq f \quad \forall n \in \mathbb{N}$
$$0 \leq \underbrace{\frac{1}{n}\mu(A_n) =}_{\mu \text{ ist Maß}} \underbrace{\int \frac{1}{n} I_{A_n}\, d\mu}_{\text{Integral über Indikatorfkt.}} \overset{\text{Monotonie}}{\leq} \int f\, d\mu \overset{\text{nach Vor.}}{=} 0$$
$\Rightarrow \mu(A_n) = 0 \quad \forall n \in \mathbb{N}$ und aus $A_n \uparrow A$ folgt $\mu(A_n) \uparrow \mu(A) = 0$ (Stetigkeit von unten)

"⇐": $\mu(A) = 0$
$f \leq \infty \cdot I_A$
$0 \leq \int f\, d\mu \leq \int \infty \cdot I_A\, d\mu = \infty \cdot \mu(A) = 0$

□

Bemerkung: zu Satz 6.4: $f \in M^+$

$$\int f\, d\mu = 0 \Leftrightarrow f = 0 \quad \mu\text{-f.ü.}$$

Satz 6.5 (Riemann & Lebesgue-Integral)**.** *Sei $f : \mathbb{R} \to \mathbb{R}$ meßbar und auf $[a, b] \subset \mathbb{R}$ Riemann-integrierbar (Ober- und Untersumme konvergieren gegen Integralwert). Dann ist f Lebesgue-integrierbar und es gilt*

$$\underbrace{\int_{[a,b]} f\, d\lambda}_{\textit{Lebesgue}} = \underbrace{\int_a^b f(x)\, dx}_{\textit{Riemann}}.$$

Beweis: Satz 2.17 Meintrup and Schäffler [2005]. □
Bemerkung: Etwas allgemeiner gilt auch

$$\int_I f d\lambda = \int_I f(x) dx, \quad I \subset \mathbb{R}$$

Siehe Satz 2.18 Meintrup and Schäffler [2005].

Bemerkung: Es gibt Funktionen f, die Lebesgue-integrierbar aber nicht Riemann-integrierbar sind und vice versa.

Bemerkung: Wenn f Riemann-integrierbar ist gilt:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \text{ und } F_{\mathbb{P}}(x) = \mathbb{P}(]-\infty, x]) = \int_{-\infty}^{x} f(y)\, dy.$$

Berhard Riemann (17. September 1826 bis 20. Juli 1866). Studium in Göttingen (u.a. bei Gauß) und Berlin, während der Revolution im März 1848 im Studentenkorps in Berlin, Promotion 1851 und Habilitation 1854 in Göttingen. Seit 1857 außerordentlicher Professor in Göttingen.
http://de.wikipedia.org/wiki/Bernhard_Riemann

6.2 Integralsätze

Satz 6.6. *Sei $X : \Omega \to \mathbb{R}^n$ eine n-dimensionale reelle Zufallsvariable und $f : \mathbb{R}^n \to \mathbb{R}$ eine meßbare Funktion, so daß $f \circ X : \Omega \to \mathbb{R}$ integrierbar ist. Dann gilt*

$$\int f \, d\mathbb{P}_X = \int f \circ X \, d\mathbb{P}.$$

Beweis: Ist $f = I_A, A \in \mathcal{B}^n$, so ist

$$\begin{aligned}
(f \circ X)(\omega) &= f(X(\omega)) \\
&= I_A(X(\omega)) = \begin{cases} 1 & X(\omega) \in A \\ 0 & \text{sonst} \end{cases} \\
&= I_{\underbrace{X^{-1}(A)}_{\in \mathcal{F}}}(\omega) = \begin{cases} 1 & \omega \in X^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\} \\ & \\ 0 & \text{sonst} \end{cases}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \int f \, d\mathbb{P}_X &= \int I_A \, d\mathbb{P}_X \\
&= \mathbb{P}_X(A) \\
&= \mathbb{P}(X^{-1}(A)) \\
&= \int I_{X^{-1}(A)} \, d\mathbb{P} \\
&= \int f \circ X \, d\mathbb{P}
\end{aligned}$$

Für $f \in T^+$ folgt alles weitere aus der Linearität und $f \in M^+$ über Grenzwertbildung. □

Satz 6.7 (Satz von der monotonen Konvergenz, Satz von Beppo Levi)**.** *Für eine monoton wachsende Folge $(f_n)_{n\in\mathbb{N}}$ von Funktionen aus M^+ gilt:*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int \lim_{n\to\infty} f_n d\mu.$$

Beweis:

"$\leq$": $f := \lim\limits_{n\to\infty} f_n \overset{\text{Monotonie}}{=} \sup\limits_{n\to\infty} f_n$ (meßbar). Nach Vor.: $f_n \leq f \quad \forall n \in \mathbb{N}$ (Monotonie) $\Rightarrow \int f_n \, d\mu \leq \int f \, d\mu$ (Monotonie des Integrals) und somit $\lim\limits_{n\to\infty} \int f_n \, d\mu \leq \int f \, d\mu$

"$\geq$": Sei $(e_n)_{n\in\mathbb{N}} \in T^+$; $e_n \uparrow f$
$A_n := \{\omega \in \Omega | cf_n(\omega) \geq e_k(\omega)\} \uparrow \Omega, \quad c > 1$ beliebig ($c = 1$ würde bei $f \in T^+, e_k = f \quad \forall k \in \mathbb{N}$ und $f_n \leq f = e_k$ nicht reichen), $k \in \mathbb{N}$ fest
$cf_n(\omega) \geq e_k(\omega) \cdot I_{A_n}(\omega) \uparrow e_k(\omega)$
$\Rightarrow$

$$\int e_k \, d\mu = \lim_{n\to\infty} \int e_k I_{A_n} \, d\mu \leq c \lim_{n\to\infty} \int f_n \, d\mu.$$

Weil c beliebig ist, folgt auch

$$\int e_k \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu \quad \forall k \in N$$

(denn sonst wäre ja $\int e_k \, d\mu > \lim_n \int f_n \, d\mu$ und somit gäbe es auch ein $c > 1$ mit $\int e_k \, d\mu > c \lim_n \int f_n \, d\mu$ was ein Widerspruch ist) und

$$\int f \, d\mu = \lim_{k\to\infty} \int e_k \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu.$$

□

Korollar 6.1 (zu Satz 6.7)**.** *Für jede Folge $(f_n)_{n\in\mathbb{N}}$ von Funktionen aus M^+ gilt:*

$$\int \sum_{i=1}^{\infty} f_i d\mu = \sum_{i=1}^{\infty} \int f_i d\mu.$$

Beweis: Wende Satz 6.7 auf $g_n = \sum_{i=1}^{n} f_i$ (Partialsummen) an. □

Satz 6.8 (Lemma von Fatou)**.** *Für jede Folge $(f_n)_{n\in\mathbb{N}} \in M^+$ gilt:*

$$\liminf_{n\to\infty} \int f_n \, d\mu \geq \int \liminf_{n\to\infty} f_n \, d\mu.$$

Beweis: $f := \liminf\limits_{n\to\infty} f_n$ und $g_n := \inf\limits_{k\geq n} f_k, n \in \mathbb{N}$.

$\Rightarrow g_n \uparrow f$ und $\lim\limits_{n\to\infty} \int g_n \, d\mu \overset{\text{Satz 6.7}}{=} \int f \, d\mu$.
Weiterhin: $g_n \leq f_k \quad \forall k \geq n \quad (f_n \in M^+)$
$\Rightarrow \int g_n \, d\mu \leq \inf\limits_{k\geq n} \int f_k \, d\mu$

$$\Rightarrow \underbrace{\lim_{n\to\infty} \int g_n \, d\mu}_{=\int f \, d\mu} \leq \underbrace{\lim_{n\to\infty} \inf_{k\geq n} \int f_k \, d\mu}_{=\liminf\limits_{n\to\infty} \int f_n \, d\mu}$$

□

Pierre Joseph Louis Fatou (28. Februar 1878 bis 10. August 1929). Nach dem Studium 1898–1900 Arbeit im Pariser Observatorium, 1907 Promotion in Mathematik; lieferte Grundlagen der Chaostheorie.
http://de.wikipedia.org/wiki/Pierre_Fatou

Satz 6.9 (Satz von der dominierten Konvergenz)**.** *Seien f, g sowie $(f_n), (g_n)$ meßbare numerische Funktionen auf einem Maßraum $(\Omega, \mathcal{F}, \mu)$ mit $f_n \to f, g_n \to g$ (punktweise) und $|f_n| \leq g_n \quad \forall n \in \mathbb{N}$. Sind g und $g_n \quad \forall n \in \mathbb{N}$ integrierbar und gilt*

$$\lim_{n\to\infty} \int g_n \, d\mu = \int g \, d\mu,$$

so sind f und $f_n \quad \forall n \in \mathbb{N}$ integrierbar und es gilt

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Beweis: $|f_n| \leq g_n$ und $\int g_n^+ \, d\mu < \infty$ und $\int g^- \, d\mu < \infty$
$\Rightarrow \int f_n^+ \, d\mu < \infty$ und $\int f^- \, d\mu < \infty \Leftrightarrow f_n$ integrierbar $\forall n \in \mathbb{N}$.
$|f| \leq g \Rightarrow f$ integrierbar.
Betrachte die Funktionen $g_n + f_n \geq 0$ und $g_n - f_n \geq 0$

$$\begin{aligned}
\int g \, d\mu + \liminf_{n\to\infty} \int f_n \, d\mu &= \liminf_{n\to\infty} \int (g_n + f_n) \, d\mu \\
&\overset{\text{Fatou (6.8)}}{\geq} \int \liminf_{n\to\infty} (g_n + f_n) \, d\mu \\
&= \int g + f \, d\mu \\
&= \int g \, d\mu + \int f \, d\mu
\end{aligned}$$

Durch Subtraktion von $\int g \, d\mu$ auf beiden Seiten erhalten wir

$$\liminf_{n\to\infty} \int f_n \, d\mu \geq \int f \, d\mu$$

Mit $g_n - f_n \geq 0$ gilt

$$\begin{aligned}
\int g\,d\mu + \liminf_{n\to\infty} \int -f_n\,d\mu &\geq \int g\,d\mu - \int f\,d\mu \\
\Leftrightarrow \liminf_{n\to\infty} \int -f_n\,d\mu &\geq -\int f\,d\mu \\
\Leftrightarrow \lim_{n\to\infty} \inf_{k\geq n} \int -f_k\,d\mu &\geq -\int f\,d\mu \\
\Leftrightarrow \lim_{n\to\infty} -\sup_{k\geq n} \int f_k\,d\mu &\geq -\int f\,d\mu \\
\Leftrightarrow -\lim_{n\to\infty} \sup_{k\geq n} \int f_k\,d\mu &\geq -\int f\,d\mu \\
\Leftrightarrow -\limsup_{n\to\infty} \int f_n\,d\mu &\geq -\int f\,d\mu \\
\Leftrightarrow \limsup_{n\to\infty} \int f_n\,d\mu &\leq \int f\,d\mu
\end{aligned}$$

zusammen also

$$\begin{aligned}
&\int f\,d\mu \leq \liminf_{n\to\infty} \int f_n\,d\mu \leq \limsup_{n\to\infty} \int f_n\,d\mu \leq \int f\,d\mu \\
\Leftrightarrow \quad &\lim_{n\to\infty} \int f_n\,d\mu = \int f\,d\mu
\end{aligned}$$

□

Beppo Levi (14. Mai 1875 bis 28. August 1961). Italienischer Mathematiker. Studium in Turin, Professor in Piacenza, Cagliari, Parma und Bologna. Nachdem er seiner jüdischen Herkunft wegen Italien 1939 verlassen mußte, baute er an der Universität Rosario in Argentinien ein mathematisches Institut auf.
http://de.wikipedia.org/wiki/Beppo_Levi

6.3 Maß mit Dichte

Lemma 6.4. *Für jedes $f \in M^+$ ist*

$$\begin{aligned}
&f \odot \mu : \mathcal{F} \to \bar{\mathbb{R}} \\
&(f \odot \mu)(A) := \int_A f\,d\mu = \int f I_A\,d\mu
\end{aligned}$$

ein Maß. Sei $\nu = f \odot \mu$, so gilt für alle $g \in M^+$

$$\int g \, d\nu = \int (gf) \, d\mu.$$

Beweis:

M1) $\surd$

M2) $\surd$

M3) Sei $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}$ mit $\dot{\bigcup}_{n=1}^{\infty} A_n = A$
Dann ist $fI_A = \sum_{n=1}^{\infty} fI_{A_n}$
$(f \odot \mu)(A) = \int fI_A \, d\mu$
$= \int \left(\sum_{n=1}^{\infty} fI_{A_n}\right) d\mu$
$= \sum_{n=1}^{\infty} \int fI_{A_n} \, d\mu$ (da $fI_{A_n} \in M^+ \ \forall \ n \in \mathbb{N}$)
$= \sum_{n=1}^{\infty} (f \odot \mu)(A_n)$

$\square$

Definition 6.5 (Maß mit Dichte, Dichte)**.** *Sei $(\Omega, \mathcal{F}, \mu)$ Maßraum und $f \in M^+$. Dann heißt $f \odot \mu$ Maß mit der Dichte f bezüglich μ. Die Funktion $f : \Omega \to [0, \infty[$ heißt Dichte des Maßes $f \odot \mu$.*

Definition 6.6 (absolute Stetigkeit)**.** *Sind μ und ν zwei Maße auf $(\Omega, \mathcal{F})$, so heißt ν absolut stetig bezüglich μ, wenn $\forall A \in \mathcal{F}$ gilt:*

$$\mu(A) = 0 \Rightarrow \nu(A) = 0$$

Kurz: $\nu \ll \mu$. Man sagt auch: μ dominiert ν.

Satz 6.10 (Satz von Radon-Nikodym)**.** *Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum mit einem σ-endlichen Maß μ und ν ein weiteres Maß auf $(\Omega, \mathcal{F})$, gilt*

$$\nu \ll \mu \Leftrightarrow \exists \text{ Dichte } f \in M^+ : \nu = f \odot \mu.$$

Gibt es eine weitere Funktion $g \in M^+$ mit $\nu = g \odot \mu$ so gilt $g = f$ μ-f.ü.

Beweis:

"$\Leftarrow$": Sei $A \in \mathcal{F}$ μ-Nullmenge, d.h. $\mu(A) = 0$.
Es gilt $Z := \{\omega : (fI_A)(\omega) > 0\} \subset A \Rightarrow \mu(Z) \le \mu(A) = 0 \Rightarrow \mu(Z) = 0$.
Damit: $\nu(A) = (f \odot \mu)(A) = \int_A f \, d\mu = \int fI_A \, d\mu \overset{\text{Satz 6.4}}{=} 0$

"$\Rightarrow$": siehe Meintrup and Schäffler [2005], Kap. 2.5.

$\square$

Johann Radon (16. Dezember 1887 bis 25. Mai 1956). 1910 Promotion in Wien, 1913 Beweis der Existenz von Dichten im $\mathbb{R}^n$, Professor in Hamburg, Greifswald, Erlangen und Breslau. Nach dem Krieg Professor und Rektor der Universität Wien.
http://de.wikipedia.org/wiki/Johann_Radon

Otton Marcin Nikodym (13. August 1887 bis 4. Mai 1974). Studium der Mathematik und Physik in Lemberg, Promotion in Warschau, bis 1945 Professor in Warschau. 1930 Beweis des allgemeinen Falls. Nach dem Krieg am Kenyon College in Ohio tätig.

Satz 6.11. *Sei $f : \mathbb{R} \to [0, \infty[$ und $\mathbb{P} = f \odot \lambda$. Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß genau dann, wenn f auf $\mathbb{R}$ integrierbar ist und $\displaystyle\int f \, d\lambda = 1$.*

Beweis: $\mathbb{P}([a, b]) = \displaystyle\int_{[a,b]} f \, d\lambda \quad \in [0, 1] \quad \forall a, b$

Insbesondere $\displaystyle\int_{\mathbb{R}} f \, d\lambda = 1$. □

Kapitel 7

Diskrete und stetige Verteilungen

7.1 Diskrete Verteilungen

Definition 7.1 (diskreter W'keitsraum)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum. Gibt es eine abzählbare (endlich oder unendlich) Menge* $T \in \mathcal{F}$ *mit* $\mathbb{P}(T) = 1$*, so heißt* $(\Omega, \mathcal{F}, \mathbb{P})$ *diskreter Wahrscheinlichkeitsraum,* $\mathbb{P}$ *diskretes Wahrscheinlichkeitsmaß und* T *ein abzählbarer Träger von* $\mathbb{P}$*.*

Bemerkung: Ω selbst muß nicht abzählbar sein, ist es dies jedoch, so ist $T = \Omega$.

Satz 7.1. *Ist* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein diskreter Wahrscheinlichkeitsraum mit Träger* T*, so ist die Funktion*

$$\begin{aligned} f : \Omega &\to [0,1] \\ \omega &\mapsto f(\omega) = \begin{cases} \mathbb{P}(\{\omega\}) & \text{für } \omega \in T \\ 0 & \text{sonst} \end{cases} \end{aligned}$$

eine Dichte von $\mathbb{P}$ *bezüglich des Zählmaßes* μ_Z*, also* $\mathbb{P} = f \odot \mu_Z$*. Insbesondere gilt*

$$\mathbb{P}(A) = \int_A f \, d\mu_Z = \sum_{\omega \in A \cap T} f(\omega) \quad \forall A \in \mathcal{F}. \tag{7.1}$$

Umgekehrt gibt es zu jeder Funktion f *der obigen Gestalt* <u>*genau*</u> *ein diskretes Wahrscheinlichkeitsmaß* $\mathbb{P}$ *auf* $(\Omega, \mathcal{F})$*, so daß (7.1) gilt.*

Definition 7.2 (Zähldichte)**.** f *aus Satz 7.1 heißt Zähldichte.*

Beweis: Zunächst zu zeigen:
$\int_A g\, d\mu_Z = \sum_{\omega \in A} g(\omega);\; g \in M^+$
Der Träger $\{\omega \in \Omega | g(\omega) > 0\}$ ist abzählbar!
Zunächst einfach: $g = I_B$

$$\int_A g\, d\mu_Z = \int_A I_B\, d\mu_Z = \int I_A \cdot I_B\, d\mu_Z = \int I_{A\cap B}\, d\mu_Z = \mu_Z(A\cap B) = |A\cap B| = \sum_{\omega\in A} \underbrace{g(\omega)}_{=I_B!}$$

etwas schwieriger: $g = \sum_{i=1}^n c_i I_{B_i} \in T^+$

$$\begin{aligned}\int_A g\, d\mu_Z &= \int_A \sum_{i=1}^n c_i I_{B_i}\, d\mu_Z \overset{\text{Linearität}}{=} \sum_{i=1}^n c_i \int_A I_{B_i}\, d\mu_Z = \sum_{i=1}^n c_i \sum_{\omega\in A} I_{B_i}(\omega) \\ &= \sum_{\omega\in A}\sum_{i=1}^n c_i I_{B_i}(\omega) = \sum_{\omega\in A} g(\omega)\end{aligned}$$

ganz schwierig: $g \in M^+$ (Träger immer noch abzählbar)
mit Lemma 6.2 existiert eine Folge von Treppenfunktionen $(g_n)_{n\in\mathbb{N}} \in T^+$ mit $g_n \uparrow g$

$$\begin{aligned}\int_A g\, d\mu_Z &= \int_A \lim_{n\to\infty} g_n\, d\mu_Z \overset{\text{monotone Konvergenz}}{=} \lim_{n\to\infty}\int_A g_n\, d\mu_Z \\ &= \lim_{n\to\infty}\sum_{\omega\in A} g_n(\omega) \overset{g_n\uparrow g}{=} \sum_{\omega\in A} g(\omega)\end{aligned}$$

f hat per Definition abzählbaren Träger und damit

$$\int_A f\, d\mu_Z = \sum_{\omega\in A} f(\omega) \quad \forall A \in \mathcal{F}.$$

Desweiteren

$$\mathbb{P}(A) = \mathbb{P}(A\cap T) \overset{\mathbb{P}\text{ ist diskret}}{=} \sum_{\omega\in A\cap T} \mathbb{P}(\{\omega\}) \overset{\text{Def. } f}{=} \sum_{\omega\in A} f(\omega) = \int_A f\, d\mu_Z \overset{\text{Lemma 6.4}}{=} (f\odot\mu_Z)(A)$$

$\Leftrightarrow$ f ist Dichte bzgl. μ_Z und $(f \odot \mu_Z) = \mathbb{P}$ ist Maß mit der Dichte f bzgl. μ_Z.

Umgekehrt: Ist eine Zähldichte f gegeben, so definieren wir $\mathbb{P} = f \odot \mu_Z$ und per Definition

$$\mathbb{P}(A) = (f \odot \mu_Z)(A) = \int\limits_A f\, d\mu_Z = \sum_{\omega \in A} f(\omega)$$

$\square$

Bemerkung: Jedes diskrete Wahrscheinlichkeitsmaß $\mathbb{P}$ hat genau eine Zähldichte f bezüglich μ_Z. Dies folgt aus Satz 6.10.

Beispiel 7.1 (diskrete Gleichverteilung). $\Omega = \{1, \dots, n\}$ *endlich.*
Zähldichte $f(\omega) := \dfrac{1}{|\Omega|} \quad \forall \omega \in \Omega$
$\mathbb{P}(A) = \displaystyle\int_A f\, d\mu_Z \overset{\text{Satz 7.1}}{=} \sum_{\omega \in A} f(\omega) = \frac{|A|}{|\Omega|} \quad \forall A \in \mathcal{P}(\Omega)$
$U := \mathbb{P}$

Beispiel 7.2 (Bernoulli-Verteilung). $\Omega = \{\underbrace{\text{"Kopf"}}_{\omega_1}, \underbrace{\text{"Zahl"}}_{\omega_2}\} \quad \Omega = \{\text{"defekt"}, \text{"OK"}\}$
endlich
Zähldichte $f(\omega) = \begin{cases} p & \omega = \omega_1 \\ 1-p & \omega = \omega_2 \end{cases}$
$p \in [0,1]$ *"Erfolgswahrscheinlichkeit" Parameter*
$\mathbb{P}(A) = \int\limits_A f\, d\mu_Z = \sum_{\omega \in A} f(\omega) \quad \forall A \in \mathcal{P}(\Omega) = \{\{\omega_1\}, \{\omega_2\}, \Omega, \emptyset\}$
$B(1,p) := \mathbb{P}$

Beispiel 7.3 (Binomial-Verteilung). $\Omega = \{0, 1, 2, \dots, n\}$ *endlich*
Zähldichte $f(\omega) = \dbinom{n}{\omega} p^\omega (1-p)^{n-\omega} \quad \forall \omega \in \Omega$
$p \in [0,1], n \in \mathbb{N}$
$\mathbb{P}(A) = \displaystyle\int\limits_A f\, d\mu_Z = \sum_{\omega \in A} \binom{n}{\omega} p^\omega (1-p)^{n-\omega} \quad \forall A \in \mathcal{F} = \mathcal{P}(\Omega)$
$\mathbb{P}(\Omega) = \displaystyle\sum_{\omega \in \Omega} \binom{n}{\omega} p^\omega (1-p)^{n-\omega} = \sum_{\omega=0}^{n} \binom{n}{\omega} p^\omega (1-p)^{n-\omega} \overset{\text{Bin. Formel}}{=} (p + (1-p))^n = 1^n = 1 \quad \checkmark$
Achtung: n und p sind Parameter dieser Verteilung.
$B(n,p) := \mathbb{P}$

Beispiel 7.4 (Poisson-Verteilung). $\Omega = \mathbb{N}_0$
Zähldichte $f(\omega) = \dfrac{\lambda^\omega e^{-\lambda}}{\omega!} \quad \forall \omega \in \Omega$

$\lambda \in \mathbb{R}^+$ *Parameter*

$\mathbb{P}(A) = \int_A f \, d\mu_Z = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A} \frac{\lambda^\omega e^{-\lambda}}{\omega!}$

$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \frac{\lambda^\omega e^{-\lambda}}{\omega!} = \sum_{\omega=0}^{\infty} \frac{\lambda^\omega e^{-\lambda}}{\omega!} = \frac{1}{e(\lambda)} \sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!} = \frac{1}{\sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!}} \sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!} = 1 \quad \checkmark$

$Po(\lambda) := \mathbb{P}$

7.2 Stetige Verteilungen

Sei $\mu = (f \odot \lambda)$ ein Maß mit der Dichte f bezüglich des Lebesgue-Maßes λ, also

$$\mu(A) = \int_A f \, d\lambda \quad \forall A \in \mathcal{B}.$$

Satz 7.2. *Sei $f : \mathbb{R} \to [0, \infty[$ und $\mathbb{P} = f \odot \lambda$. Dann ist $\mathbb{P}$ ein Wahrscheinlichkeitsmaß genau dann, wenn f auf $\mathbb{R}$ integrierbar ist und $\int f \, d\lambda = 1$.*

Beweis: $\mathbb{P}([a,b]) = \int_{[a,b]} f \, d\lambda \quad \in [0,1] \quad \forall a, b$

Insbesondere $\int_{\mathbb{R}} f \, d\lambda = 1.$ □

Bemerkung: Ist $g : \mathbb{R} \to \mathbb{R}$ stetig und $f = g \cdot I_{]a,b[}$, so heißt f stetige Dichte, $F_{\mathbb{P}}$ ist auf ganz $\mathbb{R}$ stetig. $\mathbb{P}$ heißt stetige Verteilung.

Beispiel 7.5 (Stetige Gleichverteilung). $\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B},\ a < b \in \mathbb{R}$

$f(\omega) = \frac{1}{b-a} I_{]a,b[}(\omega) \quad \forall \omega \in \Omega$

$\mathbb{P}(A) = \int_A f \, d\lambda = \frac{1}{b-a} \int I_{]a,b[\cap A} \, d\lambda = \frac{\lambda\left(]a,b[\cap A\right)}{b-a} = 1 \quad A =]a,b[$

$F(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \quad \textit{Verteilungsfunktion} \\ 1 & x > b \end{cases}$

$U(a,b) := \mathbb{P}$

Beispiel 7.6 (Normalverteilung). $\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B},\ \mu, \sigma \in \mathbb{R},\ \sigma > 0$

$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$ *Dichte der Normalverteilung*

$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ *Dichte der Standardnormalverteilung* ($\mu = 0, \sigma = 1$)

$\mathbb{P}(A) = \int_A f \, d\lambda \quad \forall A \in \mathcal{B}$

$\mathbb{P}(]-\infty, x]) = \int\limits_{-\infty}^{x} \varphi(y)\, dy =: \Phi(x)$ *Verteilungsfunktion der Standardnormalverteilung*

$F(x) = \int\limits_{-\infty}^{x} f(y)\, dy$ *Verteilungsfunktion der Normalverteilung*

z.Z. $\Phi(\infty) = 1$

Zunächst: $\int\limits_0^\infty \exp\left(-\frac{x^2}{2}\right) dx = \sqrt{\pi}$

$$\left(\int\limits_0^\infty \exp(-y^2)\, dy\right)^2 = \int\limits_0^\infty \exp(-y^2)\, dy \int\limits_0^\infty \exp(-t^2)\, dt$$

$$\overset{Fubini}{=} \int\limits_0^\infty \left(\int\limits_0^\infty \exp(-y^2)\, dy\right) \exp(-t^2)\, dt$$

$$\overset{t=x\cdot y}{=} \int\limits_0^\infty \int\limits_0^\infty \exp(-y^2)\, dy \; \exp(-x^2y^2)\, y\, dx$$

$$\overset{Fubini}{=} \int\limits_0^\infty \underbrace{\int\limits_0^\infty y \cdot \exp(-(1+x^2)y^2)\, dy}_{\frac{1}{2}\frac{1}{1+x^2}} dx$$

$$= \frac{1}{2}\int\limits_0^\infty \frac{1}{1+x^2}\, dx = \frac{1}{2}[\arctan]_0^\infty = \frac{1}{2}\left(\frac{\pi}{2} - 0\right) = \frac{\pi}{4}$$

$$\Rightarrow \int\limits_0^\infty \exp(-y^2)\, dy = \frac{\sqrt{\pi}}{2}$$

$$\overset{\exp(-y^2) \text{ symmetrisch um } y=0}{\Rightarrow} \int\limits_{-\infty}^\infty \exp(-y^2)\, dy = \sqrt{\pi}$$

Weiterhin:

$$\int\limits_{-\infty}^\infty \exp\left(-\frac{y^2}{2}\right) dy \overset{x=\frac{y}{\sqrt{2}}}{=} \int\limits_{-\infty}^\infty \exp(-x^2)\sqrt{2}\, dx = \sqrt{2\pi}$$

$$\Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = \frac{1}{\sqrt{2\pi}}\sqrt{2\pi} = 1$$

$$\int\limits_{-\infty}^{\infty} f(x)\,dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sigma}\,\varphi\left(\frac{x-\mu}{\sigma}\right) dx \overset{u=\frac{x-\mu}{\sigma}}{=} \int\limits_{-\infty}^{\infty} \varphi(u)\,du = 1$$

$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$ *(nur numerisch zu approximieren.)*

$N(\mu, \sigma^2) := \mathbb{P}$

Sei $(\mathbb{R}, \mathcal{B}, N(10,4))$ *unser Wahrscheinlichkeitsraum und* $A =]12, 20]$ *das Ereignis* $12 < \omega \leq 20$.
Dann ist

$$\mathbb{P}(A) = \int_A f\,d\lambda = \int\limits_{12}^{20} f(x)\,dx = \int\limits_{-\infty}^{20} f(x)\,dx - \int\limits_{-\infty}^{12} f(x)\,dx = \Phi\left(\frac{20-10}{2}\right) - \Phi\left(\frac{12-10}{2}\right)$$

$= 0.99... - 0.8413447 \approx 0.1586550$

Beispiel 7.7 (Exponentialverteilung)**.** $\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B}$

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \lambda\,\exp(-\lambda x) & x > 0 \end{cases}$$

$\lambda \in \mathbb{R}^+$ *Parameter*

$\mathbb{P}(A) = \int_A f\,d\lambda$

$$F(x) = \int\limits_{-\infty}^{x} f(y)\,dy = \int\limits_{-\infty}^{x} \lambda\,\exp(-\lambda y)\,dy = \lambda \int\limits_{0}^{x} \exp(-\lambda y)\,dy = \lambda\frac{1}{-\lambda}[\exp(-\lambda y)]_0^x$$

$= 1 - \exp(-\lambda x)$ *für* $x > 0, 0$ *sonst*

$\to 1$ *für* $x \to \infty$

7.3 Gemischte Verteilungen

Beispiel 7.8 (Gemischte Verteilung)**.** *Ein Smartphone-Hersteller verkauft eine erweiterte Reparaturversicherung. Aus langjähriger Erfahrung ist ihm bekannt, dass* $\alpha\%$ *seiner Smartphones während dieser Phase keine Reparatur benötigen. Die Kosten der Reparatur seien Exponentialverteilt mit Erwartungswert* β *Euro. Sei* X *die zukünftigen Reparaturkosten eines neu verkauften Smartphones. Wie ist* X *verteilt.*

Wir konstruieren uns ein Wahrscheinlichkeitsmaß $\mathbb{P}$ *auf* $(\mathbb{R}, \mathcal{B})$. *Es gilt:*

- $\mathbb{P}$ *ist auf* $\mathbb{R}^+$ *absolut stetig,*

- $\mathbb{P}(0) > 0$.

$\mathbb{P}$ besitzt also einen absolut stetigen und einen diskreten Teil. $\mathbb{P}$ wird nicht vom Lebesguemaß dominiert und besitzt somit keine Lebesgue-Dichte (vergeiche hierzu auch den Satz von Radon-Nikodym 6.10).

Wähle das Maß μ, so daß $\mathbb{P} = f^\star \odot \mu$ ein Maß mit Dichte $f^\star$ bezüglich μ ist. Wähle

$$\mu := \lambda + \delta_0.$$

μ ist ein σ-endliches Maß!

Wir zeigen

$$\mathbb{P} \ll \mu. \tag{7.2}$$

Sei dazu $B \in \mathcal{B}$, so dass $\mu(B) = 0$. Also

$$\underbrace{\lambda(B)}_{\geq 0} + \underbrace{\delta_0(B)}_{\geq 0} = 0$$

und somit

$$\lambda(B) = 0, \qquad \delta_0(B) = 0.$$

Schließlich folgt daher

$$\mathbb{P}(B) \overset{0 \notin B}{=} \int_B f(x)\, d\lambda(x) = \int f(x) I_B(x)\, d\lambda(x) \overset{\text{Bem. zu Satz 6.4}}{=} 0.$$

Also ist

$$\mathbb{P} \ll \lambda + \delta_0 = \mu$$

und damit existiert mit dem Satz von Radon-Nikodym auch eine Dichte $f^\star \in M^+$ von $\mathbb{P}$ bzgl. μ. Es gilt

$$\mathbb{P}(B) = \int_B f(x)\, d\lambda(x) \qquad \forall\, B \in \mathcal{B} \quad \text{mit } \{0\} \notin B$$

und

$$\mathbb{P}(0) = \alpha > 0.$$

Setze nun

$$f^\star(x) := f(x) \qquad \forall\, x \in \mathbb{R} \setminus \{0\}$$

und

$$f^\star(x) := \alpha \qquad \text{falls } x = 0.$$

Dann ist

$$f^\star : \mathbb{R} \to \mathbb{R}, \quad x \mapsto f^\star(x)$$

die Dichte von $\mathbb{P}$ bzgl. $\mu = \lambda + \delta_0$, das heißt

$$\mathbb{P}(A) = \int_A f^\star(x)\, d\mu(x) \qquad \forall\, A \in \mathcal{B}.$$

Bemerkung: Dank der in den letzten Kapitel erarbeiteten, einheitlichen Theorie können wir alle erdenklichen Verteilungen behandeln. Lästige (und verwirrende) Fallunterscheidungen (stetig, diskret,...) sind damit nicht mehr nötig!

Kapitel 8

Bedingte Wahrscheinlichkeit und Unabhängigkeit

8.1 Bedingte Wahrscheinlichkeit

Definition 8.1 (bedingte W'keit)**.** *Ist* $(\Omega, \mathcal{F}, \mathbb{P})$ *ein W'keitsraum und* $\mathbb{P}(B) > 0$ *für ein* $B \in \mathcal{F}$*, so heißt*

$$\begin{aligned} \mathbb{P}(\cdot \,|\, B) : \mathcal{F} &\to [0,1] \\ A &\mapsto \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \end{aligned}$$

die bedingte W'keitverteilung bzw. bedingte Wahrscheinlichkeit von A *gegeben* B*.*

Satz 8.1. $\mathbb{P}(\cdot \,|\, B)$ *ist W'keitsmaß auf* $(\Omega, \mathcal{F})$*.*

Beweis:

M1) $\surd$

M2) $\mathbb{P}(\cdot \,|\, B) \geq 0 \quad \surd$

M3) Sei $A_i, i \in \mathbb{N} \in \mathcal{F}$ disjunkt
$\mathbb{P}\left(\dot{\bigcup}_{i=1}^{\infty} A_i \middle| B\right) \overset{!}{=} \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$

$$
\begin{aligned}
\sum_{i=1}^{\infty} \mathbb{P}(A_i \,|\, B) &= \frac{1}{\mathbb{P}(B)} \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B) \\
&\overset{\mathbb{P} \text{ ist W'keitsmaß}}{=} \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\dot{\bigcup_{i=1}^{\infty}} (A_i \cap B) \right) \\
&= \frac{1}{\mathbb{P}(B)} \mathbb{P}\left(\left(\dot{\bigcup_{i=1}^{\infty}} A_i \right) \cap B \right) \\
&= \mathbb{P}\left(\dot{\bigcup_{i=1}^{\infty}} A_i \,\middle|\, B \right)
\end{aligned}
$$

Normiertheit $\mathbb{P}(\Omega|B) = \mathbb{P}(B)/\mathbb{P}(B) = 1$.

□

Bemerkung:
Die Definitionen übertragen sich auf Zufallsvariablen, also z.B.

$$
\begin{aligned}
\mathbb{P}(X > t \,|\, X > 0) &= \frac{\mathbb{P}(X > t\,,\, X > 0)}{\mathbb{P}(X > 0)} \\
&= \frac{\mathbb{P}(\{\omega \in \Omega \,|\, X(\omega) > t \,\wedge\, X(\omega) > 0\})}{\mathbb{P}(\{\omega \in \Omega \,|\, X(\omega) > 0\})}
\end{aligned}
$$

Beispiel 8.1 (Gedächtnislosigkeit der Exponenentialverteilung)**.** $X \sim \mathit{Exp}(\lambda)$

$\mathbb{P}(X > t + s \,|\, X > t) = \mathbb{P}(X > s)$,

denn: $\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - F_X(t) = 1 - (1 - \exp(-\lambda t)) = \exp(-\lambda t) \quad \forall t \geq 0$.

$\Rightarrow \forall s, t \geq 0$

$$
\begin{aligned}
\mathbb{P}(X > t + s \,|\, X > t) &= \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} \\
&= \frac{\exp(-\lambda(t + s))}{\exp(-\lambda t)} \\
&= \frac{\exp(-\lambda t) \cdot \exp(-\lambda s)}{\exp(-\lambda t)} \\
&= \exp(-\lambda s) = \mathbb{P}(X > s)
\end{aligned}
$$

Satz 8.2. *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein W'keitsraum und $A_n \in \mathcal{F}, n \in \mathbb{N}$. Dann gilt*

$$\begin{aligned} \textit{Ist} \quad \mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) &> 0 \Rightarrow \\ \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \ldots \cdot \mathbb{P}\left(A_n \,\middle|\, \bigcap_{i=1}^{n-1} A_i\right). \end{aligned}$$

Beweis: $\mathbb{P}(\bigcap_{i=1}^{n-1} A_i) > 0 \;\Rightarrow\; \mathbb{P}(\bigcap_{i=1}^{j-1} A_i) > 0 \;\forall j \leq n-1$, da $\bigcap_{i=1}^{n-1} A_i \subset \bigcap_{i=1}^{j-1} A_i$, d.h. die bedingten Wahrscheinlichkeiten sind alle definiert.
Rest über vollständige Induktion:

(IA) $n = 1$: trivial
$n = 2$: Sei $\mathbb{P}(A_1) > 0$. Nach Def. 8.1:

$$\mathbb{P}\left(A_2|A_1\right) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \quad \Leftrightarrow \quad \mathbb{P}\left(A_1 \cap A_2\right) = \mathbb{P}(A_1) \cdot \mathbb{P}\left(A_2|A_1\right) \quad \checkmark$$

(IV) Für ein $n \in \mathbb{N}$ gelte

$$\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) > 0 \;\Rightarrow\; \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1) \cdot \mathbb{P}\left(A_2|A_1\right) \cdot \ldots \cdot \mathbb{P}\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right).$$

(IS) $n \to n+1$:
Sei $\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) > 0$. Dann

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\bigcap_{i=1}^{n} A_i \cap A_{n+1}\right) \overset{n=2}{=} \mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \mathbb{P}\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right) \\ &\overset{\text{(IV)}}{=} \mathbb{P}(A_1) \cdot \mathbb{P}\left(A_2|A_1\right) \cdot \ldots \cdot \mathbb{P}\left(A_n \middle| \bigcap_{i=1}^{n-1} A_i\right) \mathbb{P}\left(A_{n+1} \middle| \bigcap_{i=1}^{n} A_i\right). \end{aligned}$$

□

Satz 8.3 (Satz von der totalen Wahrscheinlichkeit)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ W'keitsraum. Sei $\dot{\bigcup}_{i=1}^{\infty} A_i = \Omega$ eine disjunkte Zerlegung von Ω mit $\mathbb{P}(A_i) > 0 \quad \forall i \in \mathbb{N}$. Dann gilt*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \cdot \mathbb{P}(A \mid A_i) \quad \forall A \in \mathcal{F}.$$

Beweis:
$A = \dot{\bigcup}_{i=1}^{\infty} A_i \cap A$

$$\Rightarrow \mathbb{P}(A) = \mathbb{P}\left(\dot{\bigcup}_{i=1}^{\infty} (A_i \cap A)\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i \cap A) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \cdot \mathbb{P}(A \mid A_i)$$

□

Satz 8.4 (Satz von Bayes)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *W'keitsraum mit* $\dot{\bigcup}_{i=1}^{\infty} A_i = \Omega$ *einer disjunkten Zerlegung von* Ω *mit* $\mathbb{P}(A_i) > 0 \quad \forall i \in \mathbb{N}$ *und* $B \in \mathcal{F}$ *mit* $\mathbb{P}(B) > 0$. *Dann gilt*

$$\mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{\infty} \mathbb{P}(A_i)\mathbb{P}(B \mid A_i)}$$

Beweis: $\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \cdot \mathbb{P}(B \mid A_i)$ (totale W'keit)

$$\Rightarrow \mathbb{P}(A_i \mid B) = \frac{\mathbb{P}(B \mid A_i) \cdot \mathbb{P}(A_i)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i \cap B)}{\underbrace{\mathbb{P}(B)}_{=\sum \mathbb{P}(A_i)\ldots}} \text{ (Def. 8.1)}$$

□

Beispiel 8.2. *Population von n Personen.*

$(\Omega = \{1, \ldots, n\}, \mathcal{P}, U)$ *Gleichverteilung* $U(\{\omega\}) = \dfrac{1}{n}$

$K = \{1, \ldots, 0.02 \times n\}$ *Kranke*
$G = \{0.02 \times n + 1, \ldots, n\}$ *Gesunde*

$\mathbb{P}(K) = 0.02$
$\mathbb{P}(G) = \mathbb{P}(\bar{K}) = 1 - 0.02 = 0.98$

$P \ldots$ *Test positiv,* $N \ldots$ *Test negativ mit*

$\mathbb{P}(P \mid K) = 0.95$

$\mathbb{P}(P \mid G) = 0.1$

$$\mathbb{P}(K \mid P) = \frac{\mathbb{P}(P \mid K) \cdot \mathbb{P}(K)}{\underbrace{\mathbb{P}(P \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(P \mid G) \cdot \mathbb{P}(G)}_{=\mathbb{P}(P)}} = 16.2\%$$

$\mathbb{P}(G \mid P) = 83.8\%$

8.2 Unabhängigkeit

Definition 8.2 (Unabhängigkeit von Ereignissen)**.** *Ereignisse $A_i \in \mathcal{F}, i \in I \neq \emptyset$, eines W'keitsraumes $(\Omega, \mathcal{F}, \mathbb{P})$ heißen stochastisch unabhängig (stu), wenn für jede endliche Teilmenge $\emptyset \neq J \subset I$ gilt*

$$\mathbb{P}\left(\bigcap\nolimits_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}(A_j).$$

Bemerkung:
Insbesondere $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ für $A, B \in \mathcal{F}$.

Satz 8.5.

a) A, B stu, $\mathbb{P}(B) > 0 \Rightarrow \mathbb{P}(A \,|\, B) = \mathbb{P}(A)$

b) A, B stu $\Rightarrow A, \bar{B}$ stu.

Beweis:

a) $\mathbb{P}(A \,|\, B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \overset{\text{stu}}{=} \dfrac{\mathbb{P}(A) \cdot \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$

b)

$$\begin{aligned}
\mathbb{P}(A) \cdot \mathbb{P}(B) &= \mathbb{P}(A \cap B) \\
\Leftrightarrow \mathbb{P}(A) \cdot (1 - \mathbb{P}(\bar{B})) &= \mathbb{P}(A \cap B) \\
\Leftrightarrow \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(\bar{B}) &= \mathbb{P}(A \cap B) \\
\Leftrightarrow \mathbb{P}(A \cap B) + \mathbb{P}(A \cap \bar{B}) - \mathbb{P}(A) \cdot \mathbb{P}(\bar{B}) &= \mathbb{P}(A \cap B) \qquad | - \mathbb{P}(A \cap B) \\
\Leftrightarrow \mathbb{P}(A \cap \bar{B}) &= \mathbb{P}(A) \cdot \mathbb{P}(\bar{B})
\end{aligned}$$

□

Satz 8.6 (Borel-Cantelli-Lemma)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ W'keitsraum und $A_i, i \in \mathbb{N}$, eine Folge von Ereignissen mit $A_i \in \mathcal{F}$.*

a) Ist $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$, so folgt

$$\mathbb{P}(\limsup A_i) = 0$$

.

b) Sind die A_i stu und $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$, so folgt

$$\mathbb{P}(\limsup A_i) = 1$$

.

Beweis:

a) Sei $B_n := \bigcup_{i \geq n} A_i, n \in \mathbb{N}$. $\quad B_n \downarrow \limsup A_i$ und somit

$$\mathbb{P}(\limsup A_i) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i\geq n} A_i\right) \leq \lim_{n\to\infty} \underbrace{\sum_{i\geq n} \mathbb{P}(A_i)}_{\to 0 \text{ für } n\to\infty} = 0.$$

b) $1 - x \leq \exp(-x) \quad \forall x \in \mathbb{R}$

Damit

$$\begin{aligned} 0 \leq \mathbb{P}\left(\bigcap_{n\leq i\leq m} \bar{A}_i\right) &\overset{\text{stu}}{=} \prod_{n\leq i\leq m} (1 - \mathbb{P}(A_i)) \\ &\leq \prod_{n\leq i\leq m} \exp(-\mathbb{P}(A_i)) \\ &= \underbrace{\exp\underbrace{\left(-\underbrace{\sum \mathbb{P}(A_i)}_{=\infty}\right)}_{=-\infty \text{ für } m\to\infty}}_{=0} \to 0 \end{aligned}$$

$\Rightarrow \mathbb{P}(\overline{\limsup A_i}) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \bigcap_{i\geq n} \bar{A}_i\right) \leq \sum_{n=1}^{\infty} \underbrace{\mathbb{P}\left(\bigcap_{i\geq n} \bar{A}_i\right)}_{=0} = 0$

$\Rightarrow \mathbb{P}(\limsup A_i) = 1.$

□

Kapitel 9

Momente und erzeugende Funktionen

9.1 Momente

Definition 9.1 (Erwartungswert). *Ist $X : \Omega \to \mathbb{R}$ eine quasi-integrierbare reelle ZV, so heißt*

$$\mathbb{E}(X) := \int X \, d\mathbb{P} = \int x \, d\mathbb{P}_X(x)$$

der Erwartungswert von X.

Bemerkung: Sei $\mathbb{P}_X = f_X \odot \mu$ (d.h., $f_X : \mathbb{R} \to \mathbb{R}$ ist die μ-Dichte des Bildmaßes von $\mathbb{P}$ unter X) und $g(x) = x$ (Identität). Damit gilt:

$$\begin{aligned}
\mathbb{E}(X) &= \int X \, d\mathbb{P} = \int g \circ X \, d\mathbb{P} \\
&\overset{\text{Satz 6.6}}{=} \int g \, d\mathbb{P}_X = \int g \, d(f_X \odot \mu) \overset{\text{Lemma 6.4}}{=} \int g \cdot f_X \, d\mu \\
&= \begin{cases} \sum\limits_{x \in T} x f_X(x) & \mu = \mu_Z \text{ (diskret)} \\[2ex] \int\limits_{-\infty}^{\infty} x f_X(x) \, dx & \mu = \lambda \text{ (stetig)} \end{cases}
\end{aligned}$$

Letzteres vorausgesetzt, daß entweder T abzählbarer Träger von f_X (diskreter Fall) oder f_X Riemann-integrierbar (stetiger Fall) ist.

Bemerkung:

$$\begin{aligned}\mathbb{E}(g(X)) &= \int (g \circ X)\, d\mathbb{P} \qquad \text{wenn } g \circ X \text{ quasi-integrierbar} \\ &= \int g\, d\mathbb{P}_X \\ &= \begin{cases} \sum g(x) f_X(x) & \mathbb{P} \text{ diskret} \\ \int\limits_{-\infty}^{\infty} g(x) f_X(x)\, dx & \mathbb{P} \text{ stetig} \end{cases}\end{aligned}$$

Bemerkung: Alle Eigenschaften des Lebesgue-Integrals übertragen sich auf den Erwartungswert, insbesondere

$$\mathbb{E}(aX + b) = a\,\mathbb{E}(X) + b$$

$$X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$$

Definition 9.2 (Varianz einer ZV)**.** *Ist X eine ZV mit $\mathbb{E}(|X|) < \infty$ so heißt*

$$\mathbb{V}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Varianz von X.

Bemerkung: $\mathbb{P}_X = f_X \odot \lambda \Rightarrow \mathbb{V}(X) = \int (x - \mathbb{E}(X))^2 f(x)\, dx$
$\mathbb{P}_X = f_X \odot \mu_Z \Rightarrow \mathbb{V}(X) = \sum\limits_{x \in T} (x - \mathbb{E}(X))^2 f(x)$

Beispiel 9.1.

$$\begin{aligned}\mathbb{V}(X) = 0 &\Rightarrow \{\omega \in \Omega | (X(\omega) - \mathbb{E}(X))^2 > 0\} \textit{ ist } \mathbb{P}\textit{-Nullmenge} \\ &\Rightarrow X = \mathbb{E}(X) \quad \mathbb{P}\textit{-fast-sicher}\end{aligned}$$

Beispiel 9.2 (Bernoulli-Verteilung)**.** $(\Omega_1 = \{\text{"Kopf"}, \text{"Zahl"}\}, \mathcal{P}(\Omega_1), B(1,p))$ *W'keitsraum*
$(\Omega_2 = \{0,1\}, \mathcal{P}(\Omega_2))$ *Meßraum*

$X : \Omega_1 \to \Omega_2$
$\omega \mapsto X(\omega) = \begin{cases} 0 & \omega = \text{"Kopf"} \\ 1 & \omega = \text{"Zahl"} \end{cases}$ *(meßbar)*

$$\begin{aligned}
\mathbb{P}_X(A) &= B(1,p)(X^{-1}(A)), \ \textit{d.h. z.B.} \\
\mathbb{P}_X(\{0\}) &= B(1,p)(X^{-1}(\{0\})) \\
&= B(1,p)(\{\textit{"Kopf"}\}) \\
&= \int_{\textit{"Kopf"}} f \, d\mu_Z = f(\textit{"Kopf"}) = p.
\end{aligned}$$

Weiterhin:
$\mathbb{P}_X = f_X \odot \mu_Z$ *mit* $f_X : \underbrace{\{0,1\}}_{!!! \ \textit{nicht} \ \Omega} \to [0,1]$
$f_X(x) = p^x(1-p)^{1-x} \quad x \in \Omega_2, 0$ *sonst*

$(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_X)$ *ist W'keitsraum.*

Dann:

$$\begin{aligned}
\mathbb{E}(X) &= \int X \, d\mathbb{P} = \int x \, d\mathbb{P}_X(x) = \int x \, d(f_X \odot \mu_Z)(x) = \int x f_X(x) \, d\mu_Z(x) \\
&= \sum_{x \in \{0,1\}} x \cdot f_X(x) = 1 \cdot p + 0 \cdot (1-p) = p, \\
\mathbb{V}(X) &= \int (X - \mathbb{E}(X))^2 \, d\mathbb{P} \\
&= \sum_{x \in T} x^2 \, f_X(x) - p^2 = 1^2 \cdot p + 0^2 \cdot (1-p) - p^2 = p\,(1-p).
\end{aligned}$$

Beispiel 9.3 (Binomialverteilung)**.** $(\Omega_1 = \{0, \dots, n\}, \mathcal{P}(\Omega_1), B(n,p))$
$(\Omega_2 = \Omega_1, \mathcal{P}(\Omega_2))$

$X : \Omega_1 \to \Omega_2$
$\omega \mapsto X(\omega) = \omega \quad$ *Identität*

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&\overset{\text{Index } x=1}{=} n\,p \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\
&= n\,p \underbrace{\sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-x-1}}_{\text{Binomische Formel}} \\
&= n\,p \underbrace{(p + (1-p))^{n-1}}_{1} = n\,p,
\end{aligned}$$

$$\begin{aligned}
\mathbb{V}(X) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} - (np)^2 \\
&\overset{q:=1-p}{=} n\,p \sum_{x=1}^{n} x \binom{n-1}{x-1} p^{x-1} q^{n-x} - (n\,p)^2 \\
&= n\,p \sum_{x=0}^{n-1} (x+1) \binom{n-1}{x} p^x q^{n-1-x} - (n\,p)^2 \\
&= n\,p \left(\sum_{x=0}^{n-1} x \binom{n-1}{x} p^x q^{n-1-x} + 1 \right) - (n\,p)^2 \\
&= n\,p\,((n-1)\,p + 1) - (n\,p)^2 = n\,p\,(1-p).
\end{aligned}$$

Beispiel 9.4 (Gleichverteilung)**.** $\mathbb{E}(X) = \dfrac{1}{b-a} \displaystyle\int_a^b x\,dx = \dfrac{a+b}{2}$,

$\mathbb{V}(X) = \dfrac{1}{b-a} \displaystyle\int_a^b x^2\,dx - \left(\dfrac{a+b}{2}\right)^2 = \dfrac{(b-a)^2}{12}$.

Definition 9.3 (Momente)**.** *Sei X eine integrierbare Zufallsvariable und $n \in \mathbb{N}$, so daß X^n quasi-integrierbar ist. Dann heißt*

$m_{(n)}(X) = \mathbb{E}(\lvert X\rvert^n)$	*n-tes absolutes Moment*
$m_n(X) = \mathbb{E}(X^n)$	*n-tes Moment und*
$m_n^0(X) = \mathbb{E}((X - \mathbb{E}(X))^n)$	*n-tes zentriertes Moment von X*

Bemerkung: $\mathbb{E}(X)$ heißt erstes Moment.
$\mathbb{V}(X)$ ist das zweite zentrierte Moment.
$\mathbb{E}(X - \mathbb{E}(X)) = 0 = m_1^0(X)$
Bemerkung:

- $m_3^0(X) = m_3(X) - 3m_1(X)m_2(X) + 2m_1^3(X)$ ist die Schiefe von X.
- $K := \dfrac{m_4(X)}{(m_2^0(X))^2}$ Kurtosis oder Wölbung von X.
- $m_{(k)} < \infty \quad \Rightarrow \quad m_j < \infty \quad \forall j \leq k$ (vgl Satz 9.11)

Bemerkung:

$$\begin{aligned} &X \text{ diskret} \\ \Rightarrow m_k^0(X) &= \sum_x (x - m_1(x))^k \, f(x) \\ &= \sum_x x^k \, f(x - m_1(x)) \end{aligned}$$

$$\begin{aligned} &X \text{ stetig} \\ \Rightarrow m_k^0(X) &= \int\limits_{-\infty}^{\infty} (x - m_1(x))^k \, f(x) \, dx \\ &= \int\limits_{-\infty}^{\infty} x^k \, f(x - m_1(x)) \, dx \end{aligned}$$

Definition 9.4 (symmetrische Verteilung)**.** *Sei $\mathbb{P}$ Verteilung auf* $(\mathbb{R}, \mathcal{B})$*.*
$\mathbb{P}$ heißt symmetrisch um $a \in \mathbb{R}$ $\quad \Leftrightarrow \quad \mathbb{P}(]-\infty, a - x]) = \mathbb{P}([a + x, \infty[)$

Satz 9.1. *Sei $\mathbb{P}$ eine um a symmetrische Verteilung auf* $(\mathbb{R}, \mathcal{B})$; $X \sim \mathbb{P}_X = \mathbb{P}$ *und g eine integrierbare Funktion. Dann gilt*

i) $g \in L^1 \Rightarrow \displaystyle\int g \, d\mathbb{P}_X = \int g(x) \, d\mathbb{P}_X(x) = \int g\,(2a - x) \, d\mathbb{P}_X(x)$

ii) $m_1(X) = a$

iii) $m_{2n+1}^0(X) = 0 \quad \forall n \in \mathbb{N}.$

Beweis:

i) $\mathbb{P}_{X-a} = \mathbb{P}_{-(X-a)}$ (da $\mathbb{P}$ symmetrisch)

$$\begin{aligned}
\Rightarrow \mathbb{P}((X-a) \leq t) \quad &= \quad \mathbb{P}(X \leq a+t) \\
&\overset{\text{Symmetrie}}{=} \quad \mathbb{P}(X \geq a-t) \\
&= \quad \mathbb{P}(-X+a \leq t)
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \int g \circ X \, d\mathbb{P} = \int g \, d\mathbb{P}_X \quad &= \quad \mathbb{E}(g(X)) \\
&\overset{\text{Kreative 0}}{=} \quad \mathbb{E}(g(a+(X-a))) \\
&\overset{\text{Symmetrie}}{=} \quad \mathbb{E}(g(a-(X-a))) \\
&= \quad \mathbb{E}(g(2a-X)) \\
&= \quad \int g(2a-x) \, d\mathbb{P}_X(x)
\end{aligned}$$

ii)

$$\begin{aligned}
m_1(X) = \int X \, d\mathbb{P} \quad &\overset{\text{Kreative 0}}{=} \quad \frac{1}{2} \underbrace{\left[\int X \, d\mathbb{P} + \int (2a-X) \, d\mathbb{P}\right]}_{2a} \\
&= \quad \frac{1}{2} \, 2\, a = a
\end{aligned}$$

iii)

$$\begin{aligned}
\int (X-a)^k \, d\mathbb{P} \quad &= \quad \frac{1}{2}\left[\int (X-a)^k \, d\mathbb{P} + \int (2a-X-a)^k \, d\mathbb{P}\right] \\
&= \quad \frac{1}{2}\left[\int (X-a)^k \, d\mathbb{P} + \int (-1)^k (X-a)^k \, d\mathbb{P}\right] \\
&\overset{k \text{ ungerade}}{=} \quad \frac{1}{2}\int \ldots - \frac{1}{2}\int \ldots = 0
\end{aligned}$$

$\square$

Beispiel 9.5. $X \sim N(\mu, \sigma^2)$ *symmetrisch um* μ.
$\Rightarrow m_1(X) = \mathbb{E}(X) = \mu$

9.2 Momenterzeugende Funktion

Definition 9.5 (momenterzeugende Funktion)**.** *Ist X eine reelle ZV und $\mathcal{D} := \{s \in \mathbb{R} | \mathbb{E}(\exp(sX)) < \infty\}$, so heißt die Funktion*

$$\begin{aligned} M : \mathcal{D} &\to \mathbb{R} \\ s &\mapsto \mathbb{E}(\exp(sX)) = \int \exp(sx)\, d\mathbb{P}_X(x) \end{aligned}$$

momenterzeugende Funktion.

Satz 9.2. *Sei X eine ZV mit momenterzeugender Funktion $M : \mathcal{D} \to \mathbb{R}$. Ist $]-a, a[\subset \mathcal{D}$ für ein beliebiges $a > 0$, so gilt*

i) $\mathbb{E}(X^n) < \infty$,

ii) $M(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!} \mathbb{E}(X^n)$,

iii) $\left.\frac{\partial^n M(s)}{\partial^n s}\right|_{s=0} = \mathbb{E}(X^n)$.

Bemerkung:

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Beweis: M momenterzeugende Funktion

$$\begin{aligned} &\overset{\text{Def.}}{\Rightarrow} \quad \mathbb{E}(\exp(sX)) < \infty \quad \forall s \in]-a, a[\\ &\qquad \text{d.h. } \exp(sx) \text{ ist bzgl. } \mathbb{P}_X \text{ integrierbar} \\ &\Rightarrow \quad \exp(|sx|) = \sum_{k=0}^{\infty} \frac{|sx|^k}{k!} \text{ ist bzgl. } \mathbb{P}_X \text{ integrierbar} \end{aligned}$$

Betrachte Folge von Partialsummen

$f_n(X) = \sum_{i=0}^{n} \frac{(sX)^i}{i!} \overset{n\to\infty}{\to} \exp(sX)$

$g_n(X) = \exp(|sX|)$

$|f_n(X)| \leq g_n(X)$

$\overset{\text{Satz 6.9}}{\Rightarrow} \lim_{n\to\infty} \int f_n \, d\mathbb{P}_X = \int \lim_{n\to\infty} f_n \, d\mathbb{P}_X = \int \exp(sX) \, d\mathbb{P}_X$ existiert und

$$
\begin{aligned}
M(s) &= \mathbb{E}(\exp(sX)) \\
&= \mathbb{E}(\lim_{n\to\infty} f_n) \\
&= \lim_{n\to\infty} \mathbb{E}(f_n) = \sum_{n=0}^{\infty} \frac{s^n}{n!}\,\mathbb{E}(X^n) \quad s \in]-a, a[.
\end{aligned}
$$

Wegen der Darstellung

$$
M(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!}\,\mathbb{E}(X^n) = 0 + \frac{s}{1!}\,\mathbb{E}(X) + \frac{s^2}{2!}\,\mathbb{E}(X^2) + \frac{s^3}{3!}\,\mathbb{E}(X^3) + \dots
$$

ist dies eine Taylorreihe der Form

$$
M(s) = \sum_{n=1}^{\infty} \frac{(s-a)^n}{n!}\,\frac{\partial^n M(s)}{\partial^n s}\bigg|_{s=a}
$$

$$
\Rightarrow \frac{\partial^n M(s)}{\partial^n s}\bigg|_{s=0} = \mathbb{E}(X^n)
$$

$\square$

Beispiel 9.6 (Normalverteilung)**.**

$$X \sim N(0,1)$$

$$\begin{aligned}
M(s) &= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp(sx)\, \exp\left(-\frac{x^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left(sx - \frac{x^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \exp\left(\frac{s^2}{2} - \frac{(x-s)^2}{2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{s^2}{2}\right) \int\limits_{-\infty}^{\infty} \exp\left(-\frac{(x-s)^2}{2}\right) dx \\
&\overset{u=x-s}{=} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{s^2}{2}\right) \underbrace{\int\limits_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) du}_{\sqrt{2\pi}} \\
&= \exp\left(\frac{s^2}{2}\right)
\end{aligned}$$

$$\left.\frac{\partial M(s)}{\partial s}\right|_{s=0} = \left. s \cdot \exp\left(\frac{s^2}{2}\right)\right|_{s=0} = 0$$

$$\left.\frac{\partial^2 M(s)}{\partial^2 s}\right|_{s=0} = \left. s^2 \cdot \exp\left(\frac{s^2}{2}\right) + \exp\left(\frac{s^2}{2}\right)\right|_{s=0} = 1$$

$\Rightarrow \mathbb{E}(X) = 0$
$\mathbb{E}(X^2) = 1$
$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1$

$Y \sim N(\mu, \sigma^2)$

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

$F_X(x) = \Phi(x)$

$\Leftrightarrow Y$ *und* $\sigma X + \mu$ *haben dieselbe Verteilungsfunktion*

$\Leftrightarrow Y$ *und* $\sigma X + \mu$ *haben gleiche Verteilung:* $Y = \sigma X + \mu$

$$\begin{aligned}
\Rightarrow\ \mathbb{E}(Y) &= \mathbb{E}(\sigma X+\mu) = \sigma\,\mathbb{E}(X)+\mu = \mu \\
\Rightarrow \mathbb{E}(Y^2) &= \mathbb{E}((\sigma X+\mu)^2) \\
&= \mathbb{E}(\sigma^2X^2+2\sigma X\mu+\mu^2) \\
&= \sigma^2+\mu^2 \\
\Rightarrow\ \mathbb{V}(Y) &= \mathbb{E}(Y^2)-E(Y)^2 \\
&= \sigma^2+\mu^2-\mu^2=\sigma^2
\end{aligned}$$

9.3 Charakteristische Funktion

Definition 9.6 (Charakteristische Funktion)**.** *Sei X eine Zufallsvariable. Die Funktion $\varphi_X : \mathbb{R}\to\mathbb{C}$ mit*

$$\varphi_X(t) := \mathbb{E}(\exp(itX)) = \int \exp(itX)d\mathbb{P}$$

heißt die charakteristische Funktion von X.

Bemerkung: Mathematisch gesehen ist die charakteristische Funktion die Fourier-Transformierte der Verteilung.

Beispiel 9.7 (Charakteristische Funktionen)**.** *i) $X\sim B(n,p)$:*

$$\varphi_X(t) = (1+p+p\exp(it))^n$$

ii) $X\sim Po(\lambda)$

$$\varphi_X(t) = \exp\left(-\lambda(1-\exp(it)\right)$$

iii) $X\sim N(0,1)$

$$\varphi_X(t) = \exp(-t^2/2)$$

Satz 9.3. *Seien φ_X und φ_Y charakteristische Funktionen und $a,b\in\mathbb{R}$. Dann gilt:*

$$\varphi_X = \varphi_Y \Leftrightarrow \mathbb{P}_X = \mathbb{P}_Y.$$

Beweis: Beweis zu 16.4.7 in Schmidt [2011]. □

Satz 9.4. *Seien φ_X und φ_Y charakteristische Funktionen und $a,b\in\mathbb{R}$. Dann gilt:*

i) $\varphi_{aX+b}(t) = \exp(itb)\varphi_X(at)$.

ii) Sind X und Y unabhängig, dann ist

$$\varphi_{X+Y} = \varphi_X \varphi_Y.$$

Beweis:

i) Linearität des Lebesgue-Integrals.

ii) Wegen Unabhängigkeit gilt:

$$\begin{aligned} \varphi_{X+Y}(t) &= \mathbb{E}(\exp(it(X+Y)) = \mathbb{E}(\exp(itX)\exp(itY)) && (9.1) \\ &= \mathbb{E}(\exp(itX))\mathbb{E}(\exp(itY)) = \varphi_X(t)\varphi_Y(t). && (9.2) \end{aligned}$$

□

Satz 9.5. *Sei φ_X eine charakteristische Funktion. Dann gilt:*

i) φ_X ist eine gleichmäßig stetige Funktion.

ii) Ist $\mathbb{E}(|X|^k) < \infty$, dann ist für alle $k \in \mathbb{N}$

$$\mathbb{E}(X^k) = \frac{\varphi_X^{(k)}(0)}{i^k}.$$

Beweis: Beweis zu 16.4.5 in Schmidt [2011]. □

Satz 9.6. *Sei φ_X die integrierbare charakteristische Funktion von X. Dann gilt: Die Verteilung von X hat die Dichte*

$$f_X(x) = \frac{1}{2\pi} \int \exp(-itx)\varphi_X(t)dt.$$

Beweis: Die Funktion f_X ist wohldefiniert und stetig, da die charakteristische Funktion stetig ist.
Für alle $a, b \in \mathbb{R}$ mit $a < b$, die Stetigkeitspunkte von F_X sind, gilt:

$$\begin{aligned} \int_a^b f_X(x)dx &= \int_a^b \frac{1}{2\pi} \left(\int_{-\infty}^{\infty} \exp(-itx)\varphi_X(t)dt \right) dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_a^b \exp(-itx)\varphi_X(t)dx \right) dt \text{ (Fubini)} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\varphi_X(t) \int_a^b \exp(-itx)dx \right) dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t) \frac{\exp(-ita) - \exp(-itb)}{it} dt \\ &= \mathbb{P}(a < X < b). \end{aligned}$$

Differentieren ergibt die Behauptung. □

Bemerkung: Satz 9.6 gilt analog auch im multivariaten Fall.

9.4 Ungleichungen

Satz 9.7 (Markow- und Tschebyschow-Ungleichungen)**.** *Sei $X : \Omega \to \mathbb{R}$ eine reelle ZV. Dann gilt für jedes $\epsilon > 0$:*

$$\begin{aligned}\mathbb{P}(|X| \geq \epsilon) &\leq \frac{1}{\epsilon^n} \int |X^n| \, I_{\{|X| \geq \epsilon\}} \, d\mathbb{P} \\ &\leq \frac{1}{\epsilon^n} \, \mathbb{E}(|X|^n).\end{aligned}$$

Insbesondere ($n = 1$)

$$\mathbb{P}(|X| \geq \epsilon) \leq \frac{1}{\epsilon} \, \mathbb{E}(|X|) \quad \textit{(Markow-Ungleichung)}$$

und für $\mathbb{E}(|X|) < \infty$

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq \epsilon) \leq \frac{\mathbb{E}((X - \mathbb{E}(X))^2)}{\epsilon^2} = \frac{\mathbb{V}(X)}{\epsilon^2} \quad \textit{(Tschebyschow-Ungleichung).}$$

Beweis: Sei $Y \geq 0$ ZV. Dann gilt für jedes $\alpha > 0$:

$$\alpha \cdot I_{\{Y \geq \alpha\}} \leq Y \cdot I_{\{Y \geq \alpha\}} \leq Y.$$

$$\begin{aligned}\overset{\int \text{ monoton}}{\Rightarrow} \quad & \int \alpha \cdot I_{\{Y \geq \alpha\}} \, d\mathbb{P} = \alpha \cdot \int I_{\{Y \geq \alpha\}} \, d\mathbb{P} = \alpha \cdot \mathbb{P}(\{Y \geq \alpha\}) \leq \int Y \, I_{\{Y \geq \alpha\}} \, d\mathbb{P} \\ & \leq \int Y \, d\mathbb{P} = \mathbb{E}(Y) \\ \overset{|:\alpha}{\Rightarrow} \quad & \mathbb{P}(Y \geq \alpha) \leq \frac{1}{\alpha} \int\limits_{\{Y \geq \alpha\}} Y \, d\mathbb{P} \leq \frac{1}{\alpha} \, \mathbb{E}(Y)\end{aligned}$$

Mit $Y := |X|^n$ und $\alpha = \epsilon^n$ folgt

$$\mathbb{P}(|X|^n \geq \epsilon^n) \leq \frac{1}{\epsilon^n} \, \mathbb{E}(|X|^n).$$

Markov: Spezialfall für $n = 1$.
Tschebyschew: $n = 2$ und $X = X - \mathbb{E}(X)$. □

Andrei Andrejewitsch Markow (2./14. Juni 1856 bis 20. Juli 1922) in Rußland geborener Mathematiker, 1885 verteidigte er seine Habilitationsschrift an der Universität Sankt Petersburg und wurde dort 1886 Professor an der Fakultät für Mathematik und Physik. Er berechnete 1913 die Buchstabensequenzen in russischer Literatur, um die Notwendigkeit der Unabhängigkeit für das Gesetz der großen Zahlen nachzuweisen. Aus diesem Ansatz entwickelte sich der Markow-Prozess.
http://de.wikipedia.org/wiki/Andrei_Andrejewitsch_Markow

Pafnuti Lwowitsch Tschebyschow (14. Mai/26. Mai 1821 bis 26. November/8. Dezember 1894) entstammte einer russischen Großgrundbesitzerfamilie und wurde 1850 Professor in Sankt Petersburg. Zahlreiche Beiträge zu Interpolation, Approximationstheorie, Funktionentheorie, Wahrscheinlichkeitsrechnung, Zahlentheorie, Mechanik und Ballistik.
http://de.wikipedia.org/wiki/Pafnuti_Lwowitsch_Tschebyschow

Satz 9.8 (Jensen'sche Ungleichung)**.** *Sei $f : I \to \mathbb{R}$ eine konvexe Funktion auf einem Intervall $I \subset \mathbb{R}$ und $X : \Omega \to I$ eine integrierbare ZV. Dann gilt*

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X)).$$

Beweis:

$$f : I \to \mathbb{R} \text{ konvex} \quad :\Leftrightarrow \quad f(\alpha\, x + (1-\alpha)\, y) \leq \alpha\, f(x) + (1-\alpha)\, f(y) \quad \forall\, x, y \in I,\ \alpha \in [0,1]$$

$\Rightarrow \quad f(x) = \sup_{v \in V} v(x)$

mit $V = \{v \,|\, v(x) = a + bx \leq f(x) \quad \forall\, x \in I\}$ die Menge aller linearen Funktionen unterhalb f

$$\Rightarrow \quad \mathbb{E}(f(X)) = \mathbb{E}(\sup_{v \in V} v(X)) \overset{\forall\, v_0 \in V}{\geq} \mathbb{E}(v_0(X)) = v_0(\mathbb{E}(X))$$

$$\Rightarrow \quad \underbrace{\sup \mathbb{E}(f(X))}_{\mathbb{E}(f(X))} \geq \underbrace{\sup_{v_0 \in V} v_0\,(\mathbb{E}(X))}_{f(\mathbb{E}(X))}$$

$\square$

Johan Ludwig William Valdemar Jensen (8. Mai 1859 bis 5. März 1925) war ein dänischer Mathematiker und leistete wichtige Beiträge bei der Erforschung der riemannschen Vermutung. Er arbeitete bei der Bell Telephone Company und später der Kopenhagener Telephongesellschaft.
http://de.wikipedia.org/wiki/Johann_Ludwig_Jensen

Beispiel 9.8.

- $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$ *mit* $f(x) = x^2$ *konvex*
- $\mathbb{P}(X > 0) = 1 \Rightarrow \mathbb{E}\left(\frac{1}{X}\right) \geq \frac{1}{\mathbb{E}(X)}$ *mit* $f(x) = \frac{1}{x}$

Definition 9.7 (Norm)**.** *Sei* $f : \Omega \to \mathbb{R}$ $\mathcal{F}$-$\mathcal{B}$*-meßbar auf* $(\Omega, \mathcal{F}, \mu)$*. Dann heißt*

$$\|f\|_p := \left(\int |f|^p d\mu\right)^{1/p} \in [0, \infty[$$

die "p-Norm" von f*.*

Satz 9.9 (Ungleichung von Hölder)**.** *Es sei* $1 < p, q < \infty$ *mit* $\frac{1}{p} + \frac{1}{q} = 1$*. Dann gilt für zwei meßbare Funktionen* $f, g : \Omega \to \mathbb{R}$*:*

$$\|f \cdot g\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Beweis:

- $\|f\|_p = 0$ oder $\|g\|_q = 0 \Rightarrow f \cdot g = 0$ μ-f.ü. $\checkmark$
- $(\|f\|_p = \infty$ oder $\|g\|_q = \infty)$ <u>und</u> $\|f\|_p \|g\|_q > 0$ $\checkmark$
- $0 < \|f\|_p, \|g\|_q < \infty$:
  Sei $x, y \geq 0,\ \alpha, \beta \geq 0$ mit $\alpha + \beta = 1$. Dann gilt
  $x^\alpha y^\beta \leq \alpha x + \beta y$, denn:
  $\ln(x^\alpha y^\beta) = \alpha \ln x + \beta \ln y \overset{\text{Konkavität}}{\leq} \ln(\alpha x + \beta y)$
  $\Rightarrow x := \frac{|f|^p}{\|f\|_p^p};\quad y := \frac{|g|^q}{\|g\|_q^q};\quad \alpha = \frac{1}{p},\ \beta = \frac{1}{q}$
  $\frac{|f|\,|g|}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \frac{|f|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g|^q}{\|g\|_q^q}$
  $\Rightarrow \int \frac{|f|\,|g|\,d\mu}{\|f\|_p \|g\|_q} \leq \frac{1}{p} \cdot \frac{1}{\|f\|_p^p} \underbrace{\int |f|^p\,d\mu}_{=\|f\|_p^p} + \frac{1}{q} \cdot \frac{1}{\|g\|_q^q} \underbrace{\int |g|^q\,d\mu}_{=\|g\|_q^q}$
  $\frac{\int |f|\,|g|\,d\mu}{\|f\|_p \|g\|_q} \leq \frac{1}{p} + \frac{1}{q} \overset{\text{n.V.}}{=} 1$
  $\Leftrightarrow \int |fg|\,d\mu \leq \|f\|_p \|g\|_q$
  $\Leftrightarrow \|fg\|_1 \leq \|f\|_p \|g\|_q$

$\square$

Otto Ludwig Hölder (22. Dezember 1859 bis 29. August 1937). 1882 Promotion in Tübingen, seit 1890 Professor in Tübingen.
http://de.wikipedia.org/wiki/Otto_H%C3%B6lder

Satz 9.10 (Ungleichung von Minkowski). *Sind $f, g : \Omega \to \mathbb{R}$ meßbar und $p \geq 1$, so gilt*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p. \quad \textit{(Dreiecksungleichung)}$$

Beweis: $p = 1$ oder $\|f\|_p = \infty$ oder $\|g\|_p = \infty$ oder $\|f+g\|_p = 0$ $\quad\surd$
$p > 1, \|f\|_p < \infty, \|g\|_p < \infty$ und $\|f+g\|_p > 0$
$q := (1-\frac{1}{p})^{-1} \Rightarrow \frac{1}{p} + \frac{1}{q} = 1$
Damit

$$\begin{aligned}
\|f+g\|_p^p &= \int |f+g|^p \, d\mu \\
&= \int |f+g| \, |f+g|^{p-1} \, d\mu \\
&\overset{\text{Dreiecksungleichung}}{\leq} \int (|f|+|g|) \, |f+g|^{p-1} \, d\mu \\
&= \int |f||f+g|^{p-1} \, d\mu + \int |g||f+g|^{p-1} \, d\mu \\
&\overset{\text{zweimal Hölder}}{\leq} (\|f\|_p + \|g\|_p) \| |f+g|^{p-1} \|_q
\end{aligned}$$

Mit $(p-1)q = p$ gilt
$\| \, |f+g|^{p-1} \, \|_q = \left(\int |f+g|^p \, d\mu\right)^{\frac{1}{q}} = \|f+g\|_p^{\frac{p}{q}}$
und somit zusammen

$$\begin{aligned}
\|f+g\|_p^p &\leq (\|f\|_p + \|g\|_p) \, \|f+g\|_p^{\frac{p}{q}} \\
\|f+g\|_p^p \|f+g\|_p^{-\frac{p}{q}} &\leq (\|f\|_p + \|g\|_p) \, \|f+g\|_p^{\frac{p}{q}} \|f+g\|_p^{-\frac{p}{q}} \\
\|f+g\|_p &\leq \|f\|_p + \|g\|_p
\end{aligned}$$

(denn $p - p \cdot q^{-1} = p - p(1-\frac{1}{p}) = 1$) $\square$

Hermann Minkowski (22. Juni 1864 bis 12. Januar 1909). Studium in Königsberg zusammen mit Hurwitz und Hilbert. Lehrtätigkeit in Bonn, Königsberg und Zürich (dort u.a. Lehrer von Einstein), 1902 Ordinarius in Göttingen.
http://de.wikipedia.org/wiki/Hermann_Minkowski

Definition 9.8 (L^p-Raum). *Für $p \geq 1$ sei*

$$L^p := L^p(\Omega, \mathcal{F}, \mu) := \{f | f : \Omega \to \mathbb{R}, f \textit{ meßbar}, \|f\|_p < \infty\}.$$

Satz 9.11. *Ist $\mu(\Omega) < \infty$ und $q > p \geq 1$, so ist $L^q \subset L^p$ und es gibt $c \geq 0$, so daß*

$$\|f\|_p \leq c\|f\|_q \quad \forall f \in L^q.$$

Beweis: $r := \frac{q}{p} \quad s := \left(1 - \frac{1}{r}\right)^{-1}$, so daß $\left(\frac{1}{r} + \frac{1}{s}\right) = 1$.

$$\begin{aligned}
\|f\|_p^p &= \int |f|^p \, d\mu = \| \, |f|^p \cdot I_\Omega\|_1 \\
&\overset{\text{Hölder}}{\leq} \int |f|^{pr} \, d\mu^{\frac{1}{r}} \int |I_\Omega|^s \, d\mu^{\frac{1}{s}} \\
&= \|f\|_q^p \cdot \underbrace{\mu(\Omega)^{\frac{1}{s}}}_{<\infty} < \infty
\end{aligned}$$

und die p-te Wurzel liefert die Behauptung. $\square$

Kapitel 10

Mehrdimensionale Verteilungen

10.1 Definition

Definition 10.1 (Produktmaßraum)**.** *Seien* $(\Omega_1, \mathcal{F}_1, \mu)$ *und* $(\Omega_2, \mathcal{F}_2, \nu)$ *zwei Maßräume. Dann heißt der Maßraum*

$$(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu \otimes \nu)$$

mit Basismenge

$$\Omega_1 \times \Omega_2,$$

Produkt-σ*-Algebra*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A \times B | A \in \mathcal{F}_1, B \in \mathcal{F}_2\})$$

(wobei $\{A \times B | A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$ *ein durchschnittsstabiler Erzeuger von* $\mathcal{F}_1 \otimes \mathcal{F}_2$ *ist)*
und Produktmaß

$$\mu \otimes \nu(A \times B) = \mu(A) \cdot \nu(B) \quad \forall A \in \mathcal{F}_1, B \in \mathcal{F}_2$$

Produktmaßraum.

Satz 10.1 (Satz von Fubini für das Lebesgue-Integral)**.** *Seien* $(\Omega_1, \mathcal{F}_1, \mu)$ *und* $(\Omega_2, \mathcal{F}_2, \nu)$ *zwei Maßräume mit* σ*-endlichen Maßen* μ *und* ν*. Ist* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *eine nicht-negative,* $(\mathcal{F}_1 \otimes \mathcal{F}_2)$*-*$\mathcal{B}$*-meßbare Funktion oder eine* $(\mu \otimes \nu)$*-integrierbare Funktion, so gilt*

$$\begin{aligned} \int f \, d(\mu \otimes \nu) &= \int \left(\int f(\omega_1, \omega_2) \, d\mu(\omega_1) \right) d\nu(\omega_2) \\ &= \int \left(\int f(\omega_1, \omega_2) \, d\nu(\omega_2) \right) d\mu(\omega_1). \end{aligned}$$

Beweis: Satz 2.24 Meintrup and Schäffler [2005] □

Guido Fubini (19. Januar 1879 bis 6. Juni 1943). Studium und Promotion in Pisa, 1908 Professor in Turin. Emigrierte 1939 in die U.S.A., lehrte bis zu seinem Tod in New York.
http://www-history.mcs.st-andrews.ac.uk/Biographies/Fubini.html

Bemerkung: Die Definition 10.1 und der Satz von Fubini erstreckt sich in gleicher Weise auf Produktmaßräume

$$\left(\prod_{i=1}^{n}\Omega_i, \bigotimes_{i=1}^{n}\mathcal{F}_i, \bigotimes_{i=1}^{n}\mu_i\right)$$

basierend auf den Maßräumen $(\Omega_i, \mathcal{F}_i, \mu_i), i = 1, \dots, n$.

Definition 10.2 (n-dimensionale ZV, Verteilungsfunktion)**.** *Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}^n$ eine n-dimensionale reelle ZV, $X = (X_1, \dots, X_n)$. Dann heißt die Funktion*

$$\begin{aligned} F_X : \mathbb{R}^n &\to [0,1] \\ x = (x_1, \dots, x_n) &\mapsto \mathbb{P}(X \le x) \\ &= \mathbb{P}(X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n) \\ &= \mathbb{P}(\{\omega \in \Omega \,|\, X_i(\omega) \le x_i \,\forall i = 1, \dots, n\}) \end{aligned}$$

die n-dimensionale Verteilungsfunktion von X.

Bemerkung: Für $x \in \mathbb{R}^n$ und $h \in \mathbb{R}$ sei $(x + h) := (x_1 + h, \dots, x_n + h)$. Dann gilt

i) $F_X(x)$ ist monoton wachsend in jeder Komponente von x.

ii) $F_X(x)$ ist rechtsstetig

$$\lim_{h \downarrow 0} F_X(x + h) = F_X(x) \quad \forall x \in \mathbb{R}^n.$$

iii) $\lim_{x_k \to -\infty} F_X(x) = 0$ für alle $k \in \{1, \dots, n\}$, $\lim_{h \to \infty} F_X(x + h) = 1$.

Satz 10.2. *X, Y n-dimensionale ZV, $X \sim F_X, Y \sim F_Y$.*

$$F_X = F_Y \Rightarrow \mathbb{P}_X = \mathbb{P}_Y.$$

Beweis: Satz 4.30 in Meintrup and Schäffler [2005] □

Bemerkung: Wahrscheinlichkeitsmaße auf $\mathbb{R}^n$ sind durch die zugehörigen n-dimensionalen Verteilungsfunktionen vollständig definiert.

Bemerkung: $(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P} = f \odot \lambda^n)$ mit ZV $X(\omega) = \omega \in \mathbb{R}^n$

$$\Rightarrow \quad \mathbb{P}_X(]a,b]) = \int f\, I_{]a,b]}\, d\lambda^n \overset{f \text{ Riemann-integrierbar, Fubini}}{=} \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f(x_1, \dots, x_n)\, dx_n\, dx_{n-1} \dots dx_1$$

Satz 10.3. *Sei $X : \Omega \to \mathbb{R}^n$ eine n-dimensionale ZV und $g : \mathbb{R}^n \to \mathbb{R}^k$ $\mathcal{B}^n$-$\mathcal{B}^k$-meßbar (Baire-Funktion). Dann ist*

$$g \circ X : \Omega \to \mathbb{R}^k, \quad \omega \mapsto g(X(\omega))$$

eine k-dimensionale reelle ZV mit Verteilung

$$\mathbb{P}_{g(X)}(A) = \mathbb{P}(g(X) \in A) = \mathbb{P}(\{\omega \in \Omega \,|\, g(X(\omega)) \in A\}).$$

Beweis: Satz 5.4 (Meßbarkeit der Komposition) □

Definition 10.3 (Randverteilung). *Ist $g : \mathbb{R}^n \to \mathbb{R}$ mit $g(x) = x_j$ für ein festes $j \in \{1, \dots, n\}$ (Projektion auf j-te Koordinate), so gilt:*

$$\begin{aligned} g(X) &= X_j, \\ \mathbb{P}_{X_j}(A) &= \mathbb{P}_X(\{x \in \mathbb{R}^n : x_j \in A\}) = \mathbb{P}(X_j \in A) \text{ für } A \in \mathcal{B}. \end{aligned}$$

$\mathbb{P}_{X_j}$ heißt Randverteilung oder marginale Verteilung von X_j.

Bemerkung:
Besitzt X die Dichte f, so hat X_j die Dichte

$$\begin{aligned} f_j : \mathbb{R} &\to \mathbb{R} \\ y &\mapsto \int \dots \int f(x_1, x_2, \dots, x_{j-1}, y, x_{j+1}, \dots, x_n)\, d\nu(x_n) \dots d\nu(x_{j+1}) d\nu(x_{j-1}) \dots d\nu(x_1) \end{aligned}$$

mit z.B. $\nu = \lambda$ oder $\nu = \mu_Z$

f_j heißt Randdichte von X_j.

10.2 Transformationssatz für Dichten

Satz 10.4 (Dichtetransformationssatz). *Sei $X : \Omega \to \mathbb{R}$ ZV mit stetiger Verteilungsfunktion F_X und Dichte $f_X(x) = \dfrac{\partial F_X(x)}{\partial x}$ bezüglich des Lebesgue-Maßes λ. Weiterhin sei $g : \mathbb{R} \to \mathbb{R}$ bijektiv und stetig differenzierbar, $\dfrac{\partial g(x)}{\partial x} \neq 0$ mit $h = g^{-1}$.*

$\Rightarrow$ *$g \circ X$ hat Dichte* $\quad f_{g\circ X}(y) = f_X(h(y)) \cdot \left|\dfrac{\partial h(y)}{\partial y}\right| \quad$ *bezüglich λ.*

Beweis:
Fallunterscheidung:

(A) h monoton wachsend: $\dfrac{\partial h(y)}{\partial y} > 0$

(B) h monoton fallend: $\dfrac{\partial h(y)}{\partial y} < 0$

(A)

$$\begin{aligned} F_{g\circ X}(y) &= \mathbb{P}(g \circ X \leq y) \\ &= \mathbb{P}(X \leq g^{-1}(y)) \quad (h = g^{-1} \text{ monoton wachsend}) \\ &= \mathbb{P}(X \leq h(y)) = F_X(h(y)) \end{aligned}$$

$$f_{g\circ X}(y) = \frac{\partial F_{g\circ X}(y)}{\partial y} = \frac{\partial F_X(h(y))}{\partial y} = f_X(h(y)) \cdot \underbrace{\frac{\partial h(y)}{\partial y}}_{\geq 0}$$

(B)

$$\begin{aligned} F_{g\circ X}(y) &= \mathbb{P}(g \circ X \leq y) \\ &= \mathbb{P}(X \geq g^{-1}(y)) \quad (h = g^{-1} \text{ monoton fallend}) \\ &= \mathbb{P}(X \geq h(y)) \\ &\overset{X \text{ stetig}}{=} \mathbb{P}(X > h(y)) \\ &= 1 - F_X(h(y)) \end{aligned}$$

$$\Rightarrow f_{g\circ X}(y) = \frac{\partial F_{g\circ X}(y)}{\partial y} = 0 - f_X(h(y)) \cdot \underbrace{\frac{\partial h(y)}{\partial y}}_{\leq 0}$$

zusammen (A) und (B):
$f_{g\circ X}(y) = f_X(h(y)) \cdot \left|\frac{\partial h(y)}{\partial y}\right|$

$\square$

Beispiel 10.1. *$X \sim Exp(\lambda)$ Gesucht: Verteilung von X^2*

$g(x) = x^2$

$h(y) = g^{-1}(y) = \sqrt{y}$ *(da $f_X(x) = 0$ für $x \leq 0$ keine Fallunterscheidung nötig)*

$\frac{\partial h(y)}{\partial y} = \frac{1}{2\sqrt{y}}$

$$\begin{aligned} f_{X^2}(y) &= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \\ &= \lambda \exp(-\lambda\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} \cdot I_{[0,\infty[}(\sqrt{y}) \\ F_{X^2}(x) &= \int\limits_0^x f_{X^2}(y)\, dy \end{aligned}$$

Satz 10.5 (Trafo für injektive g)**.** *Sei $X : \Omega \to \mathbb{R}$ mit Verteilungsfunktion F_X und stetiger Dichte f_X sowie $g : \mathbb{R} \to \mathbb{R}$ eine Baire-Funktion. Sei $\mathbb{R} \supset I = \dot{\bigcup\limits_{m\in M}} I_m$ eine disjunkte Vereinigung offener Intervalle ($I_m =]a_m, b_m[$) derart, daß gilt:*

a) F_X ist stetig differenzierbar auf I:

$$f_X(x) = \frac{\partial F_X(x)}{\partial x},$$

b) $\mathbb{P}(X \in I) = 1$,

c) Sei $g_m = g|_{I_m}$; $g_m : I_m \to \mathbb{R}$ die Einschränkung von g auf I_m mit

1) g_m ist stetig differenzierbar und bijektiv,

2) $\frac{\partial g_m(x)}{\partial x}. \neq 0$

Dann gilt: $g \circ X$ *hat die Dichte*

$$f_{g \circ X}(y) = \sum_{m \in M} v_m(y),$$

$$\text{wobei } v_m(y) = \begin{cases} f_X(h_m(y)) \cdot \left| \frac{\partial h_m(y)}{\partial y} \right| & y \in g(I_m) \\ 0 & \text{sonst} \end{cases}$$

und $h_m = g_m^{-1}$.

Beweis:
Sei $V_m(y) := \mathbb{P}(X \in I_m, g \circ X \leq y)$ die gemeinsame Verteilungsfunktion der Indikatorfunktion $I_{I_m} \circ X$ und der Komposition $g \circ X$.

Zu zeigen:

$$V_m \text{ ist Stammfunktion von } v_m \quad \Leftrightarrow \quad v_m \text{ ist Dichte von } V_m$$

Fallunterscheidung:

(A) $\dfrac{\partial h_m(x)}{\partial x} > 0 \quad \forall x \in I_m =]a_m, b_m[$

(B) $\dfrac{\partial h_m(x)}{\partial x} < 0 \quad \forall x \in I_m =]a_m, b_m[$

(A)

$$\begin{aligned} V_m(y) &= \mathbb{P}(a_m \leq X \leq b_m,\, g_m \circ X \leq y) \\ &= \mathbb{P}(a_m \leq X \leq b_m,\, X \leq h_m(y)) \\ &= \mathbb{P}(a_m \leq X \leq h_m(y)) \quad \text{weil } h_m(t) \in I_m \Rightarrow h_m(t) \leq b_m \\ &= F_X(h_m(y)) - F_X(a_m) \end{aligned}$$

$$\Rightarrow \frac{\partial V_m(y)}{\partial y} = f_X(h_m(y)) \cdot \underbrace{\frac{\partial h_m(y)}{\partial y}}_{>0} = v_m(y)$$

(B) analog, zusammen folgt Form von $v_m(y)$

$$\Rightarrow V_m(y) = \int_{-\infty}^{y} v_m(t)\, dt$$

Bleibt zu zeigen: $f_{g\circ X}$ ist Dichte von $F_{g\circ X}$

$$
\begin{aligned}
\int_{-\infty}^{y} f_{g\circ X}(t)\,dt &= \int_{-\infty}^{y} \sum_{m\in M} v_m(t)\,dt \\
&\overset{\text{Linearität}}{=} \sum_{m\in M} \underbrace{\int_{-\infty}^{y} v_m(t)\,dt}_{V_m(y)} \\
&= \sum_{m\in M} \mathbb{P}(X\in I_m,\, g\circ X\le y) \\
&\overset{I=\dot{\bigcup}_m I_m,\, \mathbb{R}=I\dot{\cup}\bar{I}}{=} \mathbb{P}(X\in I,\, g\circ X\le y) + \underbrace{\mathbb{P}(X\notin I,\, g\circ X\le y)}_{=0 \text{ nach Voraussetzung}} \\
&= \mathbb{P}(X\in I,\, g\circ X\le y) \\
&= \mathbb{P}(g\circ X\le y) = F_{g\circ X}(y)
\end{aligned}
$$

□

Beispiel 10.2. $X\sim N(0,1)$. *Gesucht:* f_{X^2}

$g(x)=x^2$
$I_1 =]-\infty, 0[$
$I_2 =]0,\infty[$

$g_1 = g|_{I_1} \Rightarrow h_1(y) = -\sqrt{y}$
$g_2 = g|_{I_2} \Rightarrow h_2(y) = \sqrt{y}$

$$\left|\frac{\partial h_i(y)}{\partial y}\right| = \frac{1}{2\sqrt{y}} \quad i=1,2$$

Fallunterscheidung:

$$
\begin{aligned}
y < 0: \qquad f_{X^2}(y) &= 0 \\
y > 0: \qquad f_{X^2}(y) &= \underbrace{f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}}_{v_2(y)} + \underbrace{f_X(-\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}}_{v_1(y)} \\
&= \frac{1}{2\sqrt{y}} \left(\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y}{2}\right) \right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{2\sqrt{y}} \left(2 \cdot \exp\left(-\frac{y}{2}\right)\right) \\
&= \frac{1}{\sqrt{2\pi y}} \exp\left(-\frac{y}{2}\right) \\
&\widehat{=} \textit{ Dichte der } \chi^2\textit{-Verteilung mit einem Freiheitsgrad.}
\end{aligned}
$$

Satz 10.6 (Verallgemeinerungen auf den n-dimensionalen Fall)**.** *Sei $X : \Omega \to \mathbb{R}^n$ ZV mit Dichte f_X bzgl. λ^n.*
Sei $g : \mathbb{R}^n \to \mathbb{R}^n$ Baire-Funktion.
Ferner sei $G_m \in \mathcal{B}^n, m \in M$ derart, daß

i) $\mathbb{P}\left(X \in \dot{\bigcup\limits_{m \in M}} G_m\right) = 1,$

ii) $g_m := g|_{G_m}$ *bijektiv und stetig differenzierbar.*

Sei $H_m(y) = \underbrace{\left(\frac{\partial h_m(y)}{\partial y_1 \partial y_2 \cdots \partial y_n}\right)}_{=:J \text{ Jacobi-Matrix von } h_m}$ *und* $h_m = g_m^{-1}$.

Dann gilt

$$f_{g \circ X}(y) = \sum_{m \in M} v_m(y)$$

mit $v_m(y) = \begin{cases} f_X(h_m(y)) \cdot |\det(H_m(y))| & \textit{falls } h_m(y) \in G_m, \\ 0 & \textit{sonst.} \end{cases}$

Beweis: geschenkt. □

Beispiel 10.3. *Sei $X = (X_1, X_2)$ eine zwei-dimensionale ZV mit Dichte f_X bzgl. λ^2. Gesucht ist $f_{X_1+X_2}$.*

$g(x_1, x_2) = (X_1 + X_2, X_2) \in \mathbb{R}^2$

$\Rightarrow$ *g bijektiv und damit*

$$\begin{aligned}
h(y) &= g^{-1}(y_1, y_2) = (y_1 - y_2, y_2) \\
g \circ X &= (\underbrace{X_1 + X_2}_{=Y}, X_2) \\
\det H &= \det\left(\frac{\partial g^{-1}}{\partial y}\right) = \det\begin{pmatrix} \frac{\partial y_1 - y_2}{\partial y_1} & \frac{\partial y_1 - y_2}{\partial y_2} \\ \frac{\partial y_2}{\partial y_1} & \frac{\partial y_2}{\partial y_2} \end{pmatrix} \\
&= \det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = 1
\end{aligned}$$

$\overset{\textit{Trafo}}{\Rightarrow}$

$$f_{Y,X_2}(y, x_2) = f_X(y_1 - x_2, x_2) \cdot 1$$

$\Rightarrow$

$$\begin{aligned}
f_Y(y) &\overset{\textit{Randdichte}}{=} \int f_{Y,X_2}(y, x_2)\, dx_2 \quad \textit{(Bemerkung zu Definition 10.3)} \\
&\overset{x_1 = y - x_2}{=} \int f_X(y - x_2, x_2)\, dx_2 \quad \textit{"Faltungsformel"}
\end{aligned}$$

Falls X_1 und X_2 stu:

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \quad \textit{(kommt noch!)}$$

und damit

$$f_Y(y) = \int f_{X_1}(y - x_2) \cdot f_{X_2}(x_2)\, dx_2$$

Beispiel 10.4. $X_1 \sim N(0,1),\ X_2 \sim N(0,1)$ *stu.*

$Y_1 := \dfrac{X_2}{X_1} \quad Y_2 := X_1$

$g(x_1, x_2) = \left(\dfrac{x_2}{x_1}, x_1\right)$

$g^{-1}(y_1, y_2) = (y_2, y_1\, y_2)$

$$\det J = \det \begin{pmatrix} 0 & 1 \\ y_2 & y_1 \end{pmatrix} = -y_2$$

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1}(y_2) \overset{stu}{\cdot} f_{X_2}(y_1\, y_2) \cdot |\det J| \\ &= |-y_2| \cdot \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(y_2^2 + (y_1\, y_2)^2\right)\right) \\ &= |y_2| \cdot \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left(y_2^2 \cdot (1+y_1^2)\right)\right) \end{aligned}$$

Randverteilung von Y_1:

$$\begin{aligned} f_{Y_1}(y_1) &= \int\limits_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1,y_2)\, dy_2 \\ &\overset{Symmetrie}{=} 2 \cdot \int\limits_{0}^{\infty} \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}\, y_2^2(1+y_1^2)\right) \cdot y_2\, dy_2 \\ &= \frac{1}{\pi} \cdot \int\limits_{0}^{\infty} \exp\left(-\frac{1}{2}\, y_2^2(1+y_1^2)\right) \cdot y_2\, dy_2 \\ &\overset{Stammfkt.\ bilden}{=} \frac{1}{\pi} \cdot \left[\exp\left(-\frac{1}{2}\, y_2^2\,(1+y_1^2)\right) \cdot \frac{-1}{1+y_1^2}\right]_{y_2=0}^{\infty} \\ &= \frac{1}{\pi} \cdot \frac{-1}{1+y_1^2} \cdot \underbrace{\left[\exp\left(-\frac{1}{2}\, y_2^2\,(1+y_1^2)\right)\right]_{y_2=0}^{\infty}}_{=-1} \\ &= \frac{1}{\pi} \cdot \frac{1}{1+y_1^2} \quad \cong \quad \textit{Cauchy-Verteilung} \end{aligned}$$

Augustin Louis Cauchy (21. August 1789 bis 23. Mai 1857) war ein französischer Mathematiker. Jede Kurzfassung seines Lebenswerkes muß scheitern, deshalb sei an dieser Stelle auf
http://de.wikipedia.org/wiki/Augustin_Louis_Cauchy
verwiesen.

10.3 Unabhängigkeit von Zufallsvariablen

Definition 10.4.

i) Eine Familie von Mengensystemen $\mathcal{F}_i \subset \mathcal{F}$, $i \in I \neq \emptyset$, heißt stochastisch unabhängig, wenn dies für jede Familie von Ereignissen $A_i \in \mathcal{F}_i$, $i \in I$, für alle $i \in I$ gilt. Vgl. Definition 8.2 *ii) Eine Familie von Zufallsvariablen $X_i : \Omega \to \Omega_i$, auf einem W'keitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißt stochastisch unabhängig, wenn dies für die Familie der von den Zufallsvariablen X_i erzeugten σ-Algebra*

$$\sigma(X_i) = \sigma(X_i^{-1}(\mathcal{F}_i)) \quad i \in I$$

gilt.

Bemerkung:

$$\sigma(X_i^{-1}(\mathcal{F}_i)) \stackrel{\text{Satz 5.2}}{=} \underbrace{X_i^{-1}(\mathcal{F}_i) \subset \mathcal{F}}_{X_i \text{ ist ZV, also } \mathcal{F}\text{-}\mathcal{F}_i\text{-meßbar}}$$

Bemerkung:
$X_1, \dots, X_n$ sind reelle ZV auf $(\mathbb{R}, \mathcal{B}, \mathbb{P})$. Dann gilt:

$$X_1, \dots, X_n \text{ stu} \quad \Leftrightarrow \quad \mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) \quad \forall B_i \in \mathcal{B},$$

denn $\sigma(X_i) = X_i^{-1}(\mathcal{B}) \subset \mathcal{B}$ und nach Definition 10.4 muß obige Gleichung für alle möglichen Ereignisse B_i gelten.

Satz 10.7. *Ist $X = (X_1, \dots, X_n)$ reelle n-dimensionale ZV, so sind $X_1, \dots, X_n$ stu $\quad \Leftrightarrow \quad \forall c \in \mathbb{R}^n$ gilt*

$$\begin{aligned} \mathbb{P}(X \le c) &= \mathbb{P}(X_1 \le c_1, \dots, X_n \le c_n) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \le c_i), \\ \textit{bzw.} \quad F_X(c) &= \prod_{i=1}^{n} F_{X_i}(c_i). \end{aligned}$$

Ist X diskret, so gilt: $X_1, \dots, X_n$ stu $\quad \Leftrightarrow \quad \forall x \in T_1 \times T_2 \times \dots, \times T_n \quad$ (T_i Träger von X_i) gilt

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

Beweis: X diskret

$$\begin{aligned}
\mathbb{P}(X_1 \le c_1, \dots, X_n \le c_n) &= \sum_{X_1 \in T_1, x_1 \le c_1} \cdots \sum_{X_n \in T_n, x_n \le c_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\
&= \sum_{x_1 \le c_1} \cdots \sum_{x_n \le c_n} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\
&= \left(\sum_{x_1 \le c_1} \mathbb{P}(X_1 = x_1) \right) \cdots \left(\sum_{x_n \le c_n} \mathbb{P}(X_n = x_n) \right) \\
&= \mathbb{P}(X_1 \le c_1) \cdots \mathbb{P}(X_n \le c_n)
\end{aligned}$$

X stetig: Weise die Unabhängigkeit der Mengensysteme $\{\omega \in \Omega \,|\, X_i(\omega) \le c,\ c \in \mathbb{R}\}$ nach und nutze $\{]-\infty, c] | c \in \mathbb{R}\}$ als durchschnittsstabilen Erzeuger von $\mathcal{B}$. Details Satz 5.9/5.10 Meintrup and Schäffler [2005]. $\square$

Satz 10.8. *Sind $X_1, \dots, X_n$ reelle unabhängige ZV mit Dichten $f_{X_1}, \dots, f_{X_n}$ bzgl. λ, dann und genau dann ist die gemeinsame Dichte von $X = (X_1, \dots, X_n)$ das Produkt der einzelnen Dichten, d.h.*

$$f_X(x) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

ist Dichte bzgl. λ^n.

Beweis: Sei F_X die Verteilungsfunktion von X und $c \in \mathbb{R}^n$. Dann gilt für $I_c := (-\infty, c_1] \times \ldots \times (-\infty, c_n]$ nach Definition:

$$F_X(c) = \mathbb{P}_X(I_c) = \int_{I_c} f d\lambda^n$$

Unabhängigkeit der $X_1, \dots, X_n$ ist nach Satz 10.7 äquivalent zu:

$$F_X(c) = \mathbb{P}(X_1 \le c_1, \dots, X_n \le c_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \le c_i) = \prod_{i=1}^{n} \int_{(-\infty, c_i]} f_{X_i} d\lambda$$

$$\overset{\text{Fubini}}{=} \int_{(-\infty, c_1]} \cdots \int_{(-\infty, c_n]} \prod_{i=1}^{n} f_{X_i} d\lambda \cdots d\lambda \overset{\text{Fubini}}{=} \int_{I_c} \prod_{i=1}^{n} f_{X_i} d\lambda^n$$

Mit Satz 10.2 folgt daraus $f \odot \lambda^n = \prod_{i=1}^{n} f_{X_i} \odot \lambda^n$ und mit dem Satz von Radon-Nikodym (Satz 6.10) erhält man:

$$f = \prod_{i=1}^{n} f_{X_i} \quad \lambda^n - \text{f.ü.},$$

d.h. die Dichte f_X von X bzgl. λ^n ist das Produkt der Einzeldichten. $\square$

Satz 10.9. *Seien $X_1 : \Omega \to \mathbb{R}^{n_1}$ und $X_2 : \Omega \to \mathbb{R}^{n_2}$ zwei n_i-dimensionale stochastisch unabhängige Zufallsvariablen $(i = 1, 2)$ und $h_i : \mathbb{R}^{n_i} \to \mathbb{R}^{m_i}$ Baire-Funktionen.*

$$\Rightarrow \quad Y_1 = h_1 \circ X_1 \text{ und } Y_2 = h_2 \circ X_2 \text{ stu.}$$

Beweis: Nach Voraussetzung gilt

$$\sigma(X_1) = X_1^{-1}(\mathcal{B}^{n_1}) \quad \text{und} \quad \sigma(X_2) = X_2^{-1}(\mathcal{B}^{n_2}) \quad \text{stu.}$$

z.Z.

$$Y_1^{-1}(\mathcal{B}^{m_1}) \quad \text{und} \quad Y_2^{-1}(\mathcal{B}^{m_2}) \quad \text{stu.}$$

$$\begin{aligned} Y_i^{-1}(\mathcal{B}^{m_i}) &= (h_i \circ X_i)^{-1}(\mathcal{B}^{m_i}) \\ &= X_i^{-1}(\underbrace{h_i^{-1}(\mathcal{B}^{m_i})}_{\subset \mathcal{B}^{n_i} \text{ da } h \text{ Baire-Funktion}}) \\ &\subset X_i^{-1}(\mathcal{B}^{n_i}) \end{aligned}$$

$\Rightarrow$ wg. $X_1^{-1}(\mathcal{B}^{n_1}),\ X_2^{-1}(\mathcal{B}^{n_2})$ stu,
auch $Y_1^{-1}(\mathcal{B}^{m_1}),\ Y_2^{-1}(\mathcal{B}^{m_2})$ stu. □

Satz 10.10 (Erwartungswert des Produkts unabhängiger ZV)**.** *Sind $X_1, \ldots, X_n$ unabhängige integrierbare ZV, so folgt*

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

(Die Umkehrung folgt nicht!)

Beweis: Betrachte $n = 2$

Sei $f_X = f_{X_1} \cdot f_{X_2}$ (stu) und

$\mathbb{P}_X = f_X \odot (\mu_1 \otimes \mu_2) = f_{X_1} \cdot f_{X_2} \odot (\mu_1 \otimes \mu_2)$

Betrachte $g : \mathbb{R}^2 \to \mathbb{R} \quad x \mapsto g(x) = x_1 \cdot x_2$

$$\begin{aligned}
\mathbb{E}(g \circ X) &= \int g \, d\mathbb{P}_X \\
&= \int g \, d(f_X \odot (\mu_1 \otimes \mu_2)) \\
&\overset{\text{stu und 6.4}}{=} \int g \cdot (f_{X_1} \cdot f_{X_2}) \, d(\mu_1 \otimes \mu_2) \\
&\overset{\text{Fubini}}{=} \int \int x_1 \, x_2 \, f_{X_1}(x_1) \cdot f_{X_2}(x_2) \, d\mu_1(x_1) \, d\mu_2(x_2) \\
&= \int x_1 \, f_{X_1}(x_1) d\mu_1(x_1) \int x_2 \, f_{X_2}(x_2) d\mu_2(x_2) \\
&= \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)
\end{aligned}$$

□

Satz 10.11. *Seien $X_1, \ldots, X_n$ unabhängige reelle ZV, deren momenterzeugende Funktionen $M_1, \ldots, M_n$ alle auf dem Intervall $]-a, a[$ definiert sind. Sei $S_n := \sum_{i=1}^n X_i$, so ist auch*

$$M(s) = \mathbb{E}(\exp(s \, S_n))$$

auf $]-a, a[$ definiert und es gilt

$$M(s) = \prod_{i=1}^{n} M_i(s), \quad s \in]-a, a[.$$

Beweis: Nach Satz 10.9 sind $\exp(s\,X_1), \ldots, \exp(s\,X_n)$ als Komposition stochastisch unabhängiger ZV wieder stu und, nach Voraussetzung, auf $]-a, a[$ integrierbar.
$\Rightarrow \exp\left(s \cdot \sum_{i=1}^n X_i\right) = \exp(s\,X_1) \cdot \ldots \cdot \exp(s\,X_n)$ ist integrierbar
$\Rightarrow M(s) = \mathbb{E}(\exp(s\,S_n)) \overset{\text{Satz 10.10, stu}}{=} \prod_{i=1}^n \mathbb{E}(\exp(s \cdot X_i)) = \prod_{i=1}^n M_i(s)$ □

10.4 Mehrdimensionaler Erwartungswert und Kovarianzmatrix

Definition 10.5. *Sei $X : \Omega \to \mathbb{R}^n$, $X = (X_1, \ldots, X_n)$ eine n-dimensionale ZV, dann heißt*

$$\mathbb{E}(X) := (\mathbb{E}(X_1), \ldots, \mathbb{E}(X_n))$$

der (n-dimensionale) Erwartungswert von X und

$$\mathbb{V}(X) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^\top)$$

die Kovarianzmatrix von X.

$$\begin{aligned} Cov(X, Y) &= \mathbb{V}((X, Y))_{1,2} \\ &= \mathbb{V}((X, Y))_{2,1} \\ &= \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) \end{aligned}$$

heißt Kovarianz von zwei Zufallsvariablen X und Y.

Bemerkung:

$$X \in \mathbb{R}^{n\times 1} \Rightarrow XX^\top \in \mathbb{R}^{n\times n}$$

Satz 10.12. *Sei $X : \Omega \to \mathbb{R}^n$ n-dimensionale ZV, $A \in \mathbb{R}^{m\times n}$ und $b \in \mathbb{R}^m$. Dann gilt*

i) $\mathbb{E}(AX + b) = A\,\mathbb{E}(X) + b$

ii) $\mathbb{V}(AX + b) = A\,\mathbb{V}(X)\,A^\top$

iii) $\mathbb{V}(X)$ *positiv semi definit (psd).*

Beweis:

i) Linearität des $\mathbb{E}$-Wertes

ii)

$$\begin{aligned} \mathbb{V}(AX + b) &= \mathbb{E}((AX + b - \mathbb{E}(AX + b))\,(AX + b - \mathbb{E}(AX + b))^\top) \\ &= \mathbb{E}((AX + b - A\,\mathbb{E}(X) - b)\,(AX + b - A\,\mathbb{E}(X) - b)^\top) \\ &= \mathbb{E}((AX - A\,\mathbb{E}(X))(AX - A\,\mathbb{E}(X))^\top) \\ &= \mathbb{E}(A\,(X - \mathbb{E}(X))\,(X - \mathbb{E}(X))^\top\,A^\top) \\ &= A\,\mathbb{E}((X - \mathbb{E}(X))\,(X - \mathbb{E}(X))^\top)\,A^\top \\ &= A\,\mathbb{V}(X)\,A^\top \end{aligned}$$

iii) z.Z. $x^\top\,\mathbb{V}(X)\,x \geq 0 \quad \forall x \in \mathbb{R}^n$ HA

□

Satz 10.13. *X, Y ZV, $X_1, \ldots, X_n$ ZV* *i)* $Cov(X,Y) = \mathbb{E}(X\,Y) - \mathbb{E}(X)\,\mathbb{E}(Y)$

ii) X, Y *stu* $\quad\Rightarrow\quad Cov(X,Y) = 0$

iii)

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) &= \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{1\le i<j\le n} Cov(X_i, X_j) \\ &\overset{X_1,\ldots,X_n\ stu}{=} \sum_{i=1}^{n} \mathbb{V}(X_i) \end{aligned}$$

Beweis:

i)

$$\begin{aligned} \mathrm{Cov}(X,Y) &= \mathbb{E}((X-\mathbb{E}(X))\,(Y-\mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(Y)\,\mathbb{E}(X) - \mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(X)\,\mathbb{E}(Y) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y) \end{aligned}$$

ii) X, Y stu $\quad\Rightarrow\quad \mathbb{E}(X\,Y) \overset{\text{Satz 10.10}}{=} \mathbb{E}(X)\,\mathbb{E}(Y)$

iii)

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^{n} X_i\right) &= \mathbb{E}\left(\left(\sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))\right)^2\right) \\ &= \mathbb{E}\left(\sum_{i,j} (X_i - \mathbb{E}(X_i))\,(X_j - \mathbb{E}(X_j))\right) \\ &= \sum_{i,j} \underbrace{\mathbb{E}((X_i - \mathbb{E}(X_i))\,(X_j - \mathbb{E}(X_j)))}_{\mathrm{Cov}(X_i,X_j)} \end{aligned}$$

□

Definition 10.6 (Korrelationskoeffizient)**.** *Seien* X, Y *ZV mit* $\mathbb{V}(X) < \infty,\ \mathbb{V}(Y) < \infty$. *Dann heißt*

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{\mathbb{V}(X)\,\mathbb{V}(Y)}}$$

der Korrelationskoeffizient von X *und* Y.

Satz 10.14.

$$|\rho(X,Y)| \leq 1$$

Beweis:

$$\begin{aligned} |\mathrm{Cov}(X,Y)| &\leq \mathbb{E}(|(X-\mathbb{E}(X))\cdot(Y-\mathbb{E}(Y))|) \\ &\overset{\text{Hölder-Ungleichung}}{\leq} \mathbb{E}((X-\mathbb{E}(X))^2)^{\frac{1}{2}}\cdot\mathbb{E}((Y-\mathbb{E}(Y))^2)^{\frac{1}{2}} \\ &= \mathbb{V}(X)^{\frac{1}{2}}\cdot\mathbb{V}(Y)^{\frac{1}{2}} = \sqrt{\mathbb{V}(X)\cdot\mathbb{V}(Y)} \end{aligned}$$

□

Satz 10.15 (Standardisierung)**.** *Sei* $X = (X_1,\ldots,X_n)$ *n-dimensionale ZV mit* $\mathbb{E}(X) = \mu$ *und positiv definiter Kovarianzmatrix* $\mathbb{V}(X) = \Sigma$. *Dann existiert eine invertierbare Matrix* $B \in \mathbb{R}^{n\times n}$ *mit* $Y = B^{-1}(X-\mu)$ *und* $\mathbb{E}(Y) = 0_n$ *sowie* $\mathbb{V}(Y) = 1_n$.

Beweis: Σ ist nach Voraussetzung positiv definit. Damit existiert die Cholesky-Zerlegung

$$\Sigma = B\,B^\top \quad \text{mit } B \in \mathbb{R}^{n\times n},\ B^{-1} \text{ existiert.}$$

Also
$\mathbb{E}(Y) = \mathbb{E}(B^{-1}(X-\mu)) = B^{-1}(\mathbb{E}(X)-\mu) = 0_n$

$\mathbb{V}(Y) = B^{-1}\,\mathbb{V}(X)\,{B^{-1}}^\top = B^{-1}\,B\,B^\top\,{B^{-1}}^\top = 1_n\,(B^{-1}B)^\top = 1_n\,1_n = 1_n$ □

Kapitel 11

Konvergenz

11.1 Konvergenzarten

Beispiel 11.1. $X_i \sim B(1, \frac{1}{2}) \quad i = 1, 2, \ldots$ *stu.*

$Y_n = \sum_{i=1}^{n} X_i \sim B\left(n, \frac{1}{2}\right)$

Wie groß ist der Abstand $\quad \frac{1}{n} Y_n - \frac{1}{2} \quad$ *?*

$Z_i = 1 - X_i$

$Z_i \neq X_i$ *aber* Z_i *und* X_i *haben die gleiche Verteilung.*

Definition 11.1 (Konvergenzarten)**.** *Sei* $(\Omega, \mathcal{F}, \mathbb{P})$ *W'keitsraum und* $X_i : \Omega \to \mathbb{R}$ *ZV,* $i = 1, 2, \ldots$. *Dann konvergiert die Folge* $(X_n)_{n \in \mathbb{N}}$ *gegen eine ZV* $X : \Omega \to \mathbb{R}$

a) *fast sicher, wenn*

$$\mathbb{P}(\{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega) \text{ für } n \to \infty\}) = 1;$$

Schreibweisen:

$$X_n \overset{f.s.}{\to} X,\ X_n \overset{a.s.}{\to} X,\ X_n \overset{a.e.}{\to} X,\ X_n \to X \ wp\ 1.$$

b) *im r-ten Moment,* $r \geq 1$, *wenn*

$$\mathbb{E}(|X_n|^r) < \infty \,\forall n \quad und \quad \mathbb{E}(|X_n - X|^r) \to 0 \quad \text{für } n \to \infty;$$

Schreibweise:

$$X_n \overset{r}{\to} X$$

c) in Wahrscheinlichkeit, falls

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0 \quad \textit{für } n \to \infty \; \forall \epsilon > 0;$$

Schreibweise:

$$X_n \xrightarrow{\mathbb{P}} X$$

d) in Verteilung, falls

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

und alle Stetigkeitsstellen von $F_X(x) = \mathbb{P}(X \leq x)$*; Schreibweise:*

$$X_n \xrightarrow{\mathcal{D}} X$$

Bemerkung:

a) $\Omega = \{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\} =: A$ ist uninteressant, da keine W'keit vorkommt. Konvergenz fast sicher meint, daß $\bar{A}$ eine Nullmenge bzgl. $\mathbb{P}$ ist.

b) Schreibweisen: $X_n \xrightarrow{2} X$ heißt

$$\mathbb{E}(|X_n - X|^2) \to 0.$$

c) $\mathbb{P}(|X_n - X| > \epsilon) = \displaystyle\int I_{A_n} \, d\mathbb{P}$

mit $A_n = \{\omega \in \Omega \,|\, |X_n(\omega) - X(\omega)| > \epsilon\}$

d) $X_n(\omega) \neq X(\omega) \quad \forall \omega \in \Omega, \forall n$

aber $X_n \xrightarrow{D} X$ ist erlaubt.

Beispiel 11.2 (Fortsetzung). • $Z_i \xrightarrow{\mathcal{D}} X_1$ *denn* $Z_i \sim B\left(1, \frac{1}{2}\right), X_1 \sim B\left(1, \frac{1}{2}\right)$

aber $Z_i \neq X_1$

$$\mathbb{E}\left(\frac{1}{n} Y_n\right) = \frac{1}{n} n\, p = p$$

$$\Rightarrow \mathbb{E}\left(\left(\frac{1}{n} Y_n - p\right)^2\right) = \mathbb{V}\left(\frac{1}{n} Y_n\right) = \frac{1}{n^2} n\, p\,(1-p) = \frac{1}{n} p\,(1-p) \to 0$$

und damit $\displaystyle\frac{1}{n} Y_n \xrightarrow{r=2} p$

$\mathbb{E}\left(\frac{1}{n}\,Y_n\right) = \frac{1}{n}\,\mathbb{E}(Y_n) = \frac{1}{n}\,n\,p = p$

$\mathbb{P}\left(\left|\frac{1}{n}\,Y_n - \mathbb{E}\left(\frac{1}{n}\,Y_n\right)\right| > \epsilon\right) \overset{\textit{Tschebyschev}}{\leq} \frac{\mathbb{V}\left(\frac{1}{n}Y_n\right)}{\epsilon^2} = \frac{\frac{1}{n^2}\,n\,p\,(1-p)}{\epsilon^2} = \frac{\frac{1}{n}\,p\,(1-p)}{\epsilon^2} \to 0$

für $n \to \infty$

und damit $\quad \frac{1}{n}Y_n \xrightarrow{\mathbb{P}} p = \frac{1}{2}$

Satz 11.1. *Sei* $(X_n)_{n\in\mathbb{N}}$*, Folge von ZV* $X_i : \Omega \to \mathbb{R}$ *und* $X : \Omega \to \mathbb{R}$ *ZV. Dann gilt*

a) $X_n \xrightarrow{\mathbb{P}} X \quad \Rightarrow \quad X_n \xrightarrow{\mathcal{D}} X$

b) $X_n \xrightarrow{f.s.} X \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} X$

c) $X_n \xrightarrow{r} X \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} X \quad$ *für* $r \geq 1$

d) $X_n \xrightarrow{r} X \quad \Rightarrow \quad X_n \xrightarrow{s} X \quad r > s \geq 1$

Beweis:

a)

$$\begin{array}{rcl} X_n \xrightarrow{\mathbb{P}} X; \quad F_n(x) & = & \mathbb{P}(X_n \leq x) \\ F(x) & = & \mathbb{P}(X \leq x) \end{array}$$

Es gilt für $\epsilon > 0$

$F_n(x) = \mathbb{P}(X_n \leq x) = \mathbb{P}(X_n \leq x, X \leq x+\epsilon) + \mathbb{P}(X_n \leq x, X > x+\epsilon)$
$\leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$

und

$F(x-\epsilon) = \mathbb{P}(X \leq x-\epsilon) = \mathbb{P}(X \leq x-\epsilon, X_n \leq x) + \mathbb{P}(X \leq x-\epsilon, X_n > x)$
$\leq F_n(x) + \mathbb{P}(|X_n - X| > \epsilon)$

Zusammen:

$$F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Für $n \to \infty$ gilt

$$F(x-\epsilon) - \underbrace{\mathbb{P}(|X_n - X| > \epsilon)}_{\to 0} \leq F_n(x) \leq F(x+\epsilon) + \underbrace{\mathbb{P}(|X_n - X| > \epsilon)}_{\to 0}$$

$\Rightarrow F(x-\epsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+\epsilon)$

Falls F stetig ist, gilt

$$F(x) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x)$$

$\Rightarrow F(x) = \lim_{n\to\infty} F_n(x)$

$\Rightarrow X_n \xrightarrow{\mathcal{D}} X$

c) $r = 1$ reicht. Dann

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\epsilon} \quad \text{(Markov für } X_n - X\text{)}$$

b) Sei

$$A_n(\epsilon) = \{\omega \in \Omega \,|\, |X_n(\omega) - X(\omega)| > \epsilon\}$$

und

$$B_m(\epsilon) = \cup_{n\geq m} A_n(\epsilon)$$

sowie

$$C = \{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega) \text{ für } n \to \infty\}$$

sowie

$$A(\epsilon) = \cap_m \cup_{n=m}^\infty A_n(\epsilon) = \{\omega \in \Omega \,|\, \omega \in A_n(\epsilon) \text{ für unendlich viele } n\}$$

$X_n(\omega) \to X(\omega) \quad \Leftrightarrow \quad \omega \notin A(\epsilon) \quad \forall \epsilon > 0$

somit: $\mathbb{P}(C) = 1 \Rightarrow \mathbb{P}(A(\epsilon)) = 0 \quad \forall \epsilon > 0$

Falls $\mathbb{P}(A(\epsilon)) = 0 \quad \forall \epsilon > 0 \quad \Rightarrow$

$$\mathbb{P}(\bar{C}) = \mathbb{P}\left(\cup_{\epsilon>0} A(\epsilon)\right) = \mathbb{P}\left(\cup_{m=1}^\infty A\left(\frac{1}{m}\right)\right) \leq \underbrace{\sum_{m=1}^\infty \mathbb{P}\left(A\left(\frac{1}{m}\right)\right)}_{=0 \text{ nach Vor.}} = 0$$

$\Rightarrow \mathbb{P}(C) = 1 \quad \Leftrightarrow \quad \mathbb{P}(A(\epsilon)) = 0 \quad \forall \epsilon > 0$

Weiterhin $B_m(\epsilon) \downarrow A(\epsilon) \quad \Rightarrow$

$$\mathbb{P}(A(\epsilon)) = 0 \quad \Leftrightarrow \quad \lim_{m\to\infty} \mathbb{P}(B_m(\epsilon)) = 0$$

Zusammen:

$$X_n \overset{\text{f.s.}}{\to} X \quad \Leftrightarrow \quad \mathbb{P}(B_m(\epsilon)) \to 0 \quad \text{für } m \to \infty$$

Weiterhin

$X_n \overset{\text{f.s.}}{\to} X \quad \Leftarrow \quad \sum_n \mathbb{P}(A_n(\epsilon)) < \infty \quad \forall \epsilon > 0$, denn

$$\mathbb{P}(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} \mathbb{P}(A_n(\epsilon))$$

(Borel-Cantelli) Schlußendlich: $\quad A_n(\epsilon) \subseteq B_n(\epsilon)$

$\Rightarrow \mathbb{P}(A_n(\epsilon)) \leq \mathbb{P}(B_n(\epsilon))$

$\Leftrightarrow \mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(B_n(\epsilon))$

$\Rightarrow \mathbb{P}(|X_n - X| > \epsilon) \to 0 \quad$ falls $\quad \mathbb{P}(B_n(\epsilon)) \to 0$

$\Leftrightarrow X_n \overset{\mathbb{P}}{\to} X \quad \Leftarrow \quad X_n \overset{\text{f.s.}}{\to} X \quad \Leftarrow \quad \mathbb{P}(B_n(\epsilon)) \to 0 \quad$ für $n \to \infty$.

$\square$

Bemerkung: Andere Beziehungen gelten so allgemein nicht.

- $X_n \overset{s}{\to} X \quad \not\Rightarrow \quad X_n \overset{r}{\to} X \quad r > s \geq 1$

  Betrachte

$$X_n = \begin{cases} n & \text{mit Wahrscheinlichkeit } n^{-\frac{1}{2}(r+s)} \\ 0 & \text{mit Wahrscheinlichkeit } 1 - n^{-\frac{1}{2}(r+s)} \end{cases}$$

$$\begin{aligned} \mathbb{E}(|X_n|^s) &= n^s \cdot n^{-\frac{1}{2}(r+s)} \\ &= n^{\frac{1}{2}(s-r)} \to 0 \quad \text{weil } s < r \\ \mathbb{E}(|X_n|^r) &= n^r \cdot n^{-\frac{1}{2}(r+s)} \\ &= n^{\frac{1}{2}(r-s)} \to \infty \quad \text{weil } r > s \end{aligned}$$

  und somit $\quad X_n \overset{s}{\to} 0 \quad$ aber nicht $\quad X_n \overset{r}{\to} 0$

- $X_n \xrightarrow{\mathbb{P}} X \quad \not\Rightarrow \quad X_n \xrightarrow{1} X$

  Betrachte

  $$X_n = \begin{cases} n^3 & \text{mit Wahrscheinlichkeit } n^{-2} \\ 0 & \text{mit Wahrscheinlichkeit } 1 - n^{-2} \end{cases}$$

  $\mathbb{P}(|X_n| > \epsilon) = n^{-2} \to 0 \quad \Leftrightarrow \quad X_n \xrightarrow{\mathbb{P}} 0$

  Aber $\quad \mathbb{E}(|X_n|) = n^3 \cdot n^{-2} = n \to \infty$

Satz 11.2.

a) $X_n \xrightarrow{\mathcal{D}} c$ *konstant* $\quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} c$

b) $X_n \xrightarrow{\mathbb{P}} X$ *und* $\mathbb{P}(|X_n| \leq k) = 1 \quad \forall n$ *und ein* $k \in \mathbb{R}^+$

$\Rightarrow X_n \xrightarrow{r} X \quad \forall r \geq 1$

c) $P_n(\epsilon) = \mathbb{P}(|X_n - X| > \epsilon)$ *mit* $\sum_n P_n(\epsilon) < \infty \quad \forall \epsilon > 0$

$\Rightarrow X_n \xrightarrow{f.s.} X$

Beweis:

a)

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n < c - \epsilon) + \mathbb{P}(X_n > c + \epsilon) \\ &\leq \mathbb{P}(X_n \leq c - \epsilon) + (1 - \mathbb{P}(X_n \leq c + \epsilon)) \\ &\to \mathbb{P}(X \leq c - \epsilon) + (1 - \mathbb{P}(X \leq c + \epsilon)) \\ &= 0 + (1 - 1) = 0 \end{aligned}$$

b) $X_n \xrightarrow{\mathbb{P}} X, \quad \mathbb{P}(|X_n| \leq k) = 1 \quad \Rightarrow \quad \mathbb{P}(|X| \leq k) = 1$

weil $\quad \mathbb{P}(|X| \leq k + \epsilon) = \lim_{n\to\infty} \mathbb{P}(|X_n| \leq k + \epsilon) = 1 \quad \forall \epsilon > 0$

Sei $\quad A_n(\epsilon) = \{\omega \in \Omega \,|\, |X_n(\omega) - X(\omega)| > \epsilon\}$

Dann gilt:

$$|X_n(\omega) - X(\omega)|^r \leq \epsilon^r \quad \omega \in \overline{A_n(\epsilon)}$$

bzw.

$$|X_n(\omega) - X(\omega)|^r \leq (2k)^r \quad \omega \in A_n(\epsilon)$$

jeweils mit W'keit 1. Also

$$|X_n - X|^r \leq \epsilon^r \, I_{\overline{A_n(\epsilon)}} + (2k)^r \, I_{A_n(\epsilon)}$$

$$\begin{aligned}
\mathbb{E}(|X_n - X|^r) &\leq \epsilon^r \, \mathbb{P}(\overline{A_n(\epsilon)}) + (2k)^r \, \mathbb{P}(A_n(\epsilon)) \\
&= ((2k)^r - \epsilon^r) \underbrace{\mathbb{P}(A_n(\epsilon))}_{\to 0} + \epsilon^r \\
&\to \epsilon^r \quad \text{für } n \to \infty \\
&\to 0 \quad \text{für } \epsilon \downarrow 0 \\
&\Rightarrow X_n \xrightarrow{r} X
\end{aligned}$$

c) Siehe Beweis zu Satz 11.1 b)

□

11.2 Konvergenz in Wahrscheinlichkeit

Satz 11.3 (Schwaches Gesetz der großen Zahlen)**.** *Seien $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 < \infty \quad \forall i = 1, \ldots, n$. Dann gilt*

$$\bar{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mu.$$

Beweis: $\bar{X} \xrightarrow{2} \mu$ denn

$$\mathbb{E}(\bar{X}) = \frac{1}{n} \sum (\mathbb{E}(X_i)) = \frac{1}{n} \, n \, \mu = \mu$$

und

$$\mathbb{E}(|\bar{X} - \mu|^2) = \mathbb{V}(\bar{X}) = \frac{1}{n^2} \, n \, \sigma^2 = \frac{1}{n} \, \sigma^2 \to 0$$

$\overset{\text{Satz 11.1c)}}{\Rightarrow} \bar{X} \xrightarrow{\mathbb{P}} \mu$ □

Beispiel 11.3. *$X_i \sim Po(\lambda)$ stu. $\Rightarrow \mathbb{E}(X_i) = \lambda$, $\mathbb{V}(X_i) = \lambda, i = 1, \ldots, n$ und $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \lambda$*

Bemerkung: $\mathbb{V}(X_i) = \sigma^2 < \infty$ ist nicht notwendig, $\sigma^2 = \infty$ funktioniert auch (sog. Satz von Khinchine).

Satz 11.4 (Eindeutigkeit der Konvergenz)**.**

$$X_n \xrightarrow{\mathbb{P}} X \quad \textit{und} \quad X_n \xrightarrow{\mathbb{P}} \tilde{X} \quad \Rightarrow \quad \mathbb{P}(X = \tilde{X}) = 1$$

Beweis: Sei $\epsilon > 0$, dann

$$\begin{aligned} 0 &\leq \mathbb{P}(|X - \tilde{X}| \geq \epsilon) = \mathbb{P}(|X - X_n + X_n - \tilde{X}| \geq \epsilon) \\ &\leq \mathbb{P}(|X - X_n| + |\tilde{X} - X_n| \geq \epsilon) \\ &\leq \underbrace{\mathbb{P}\left(|X - X_n| \geq \frac{\epsilon}{2}\right)}_{\to 0 \text{ n.V.}} + \underbrace{\mathbb{P}\left(|\tilde{X} - X_n| \geq \frac{\epsilon}{2}\right)}_{\to 0 \text{ n.V.}} \end{aligned}$$

$\to 0 \quad \forall \epsilon$

$$\begin{aligned} \mathbb{P}(X \neq \tilde{X}) &= \mathbb{P}\left(\cup_{k=1}^{\infty} |X - \tilde{X}| \geq \frac{1}{k}\right) \\ &\leq \sum_{k=1}^{\infty} \mathbb{P}\left(|X - \tilde{X}| \geq \frac{1}{k}\right) \\ &= \sum_{k=1}^{\infty} 0 = 0 \end{aligned}$$

□

Satz 11.5. *Seien A_n und B_n Folgen von ZV mit $A_n \xrightarrow{\mathbb{P}} a$ und $B_n \xrightarrow{\mathbb{P}} b$. Dann gilt*

a) $A_n + B_n \xrightarrow{\mathbb{P}} a + b$
$A_n - B_n \xrightarrow{\mathbb{P}} a - b$

b) $A_n \cdot B_n \xrightarrow{\mathbb{P}} a \cdot b$

c) $\dfrac{A_n}{B_n} \xrightarrow{\mathbb{P}} \dfrac{a}{b} \quad \forall b \neq 0$

Beweis: Vorarbeit:

Seien E_n und F_n Folgen von Ereignissen ($E_n \in \mathcal{F}, F_n \in \mathcal{F}$), dann gilt:

$$\mathbb{P}(E_n) \to 1 \quad \text{und} \quad \mathbb{P}(F_n) \to 1 \quad \Rightarrow \quad \mathbb{P}(E_n \cap F_n) \to 1$$

weil

$$1 - \mathbb{P}(E_n \cap F_n) = \mathbb{P}(\overline{E_n \cap F_n}) = \mathbb{P}(\bar{E_n} \cup \bar{F_n}) \leq \mathbb{P}(\bar{E_n}) + \mathbb{P}(\bar{F_n}) \to 0.$$

Hauptteil:

$\mathbb{P}(|A_n - a| < \epsilon) \to 1$
$\mathbb{P}(|B_n - b| < \epsilon) \to 1$

$$\begin{aligned} \Rightarrow \quad & \mathbb{P}(|A_n - a| < \epsilon \wedge |B_n - b| < \epsilon) \\ \leq \quad & \mathbb{P}(|A_n - a + B_n - b| < \epsilon) \\ = \quad & \mathbb{P}(|(A_n + B_n) - (a+b)| < \epsilon) \to 1 \end{aligned}$$

□

Satz 11.6. *Sei X_n eine Folge von ZV mit $X_n \xrightarrow{\mathbb{P}} c$ und $f : \mathbb{R} \to \mathbb{R}$ eine in $c \in \mathbb{R}$ stetige Funktion*

$$\Rightarrow f(X_n) \xrightarrow{\mathbb{P}} f(c).$$

Beweis: f stetig in $c \quad \Leftrightarrow \quad |f(x) - f(c)| < a,\ a > 0 \quad$ für ein b mit $|x - c| < b$.

$\mathbb{P}(|f(X_n) - f(c)| < a) \geq \mathbb{P}(|X_n - c| < b) \to 1 \quad \forall a > 0$

$\Rightarrow f(X_n) \xrightarrow{\mathbb{P}} f(c)$ □

Bemerkung: Die Ergebnisse lassen sich im Wesentlichen auf den k-dimensionalen Fall übertragen, siehe Lehmann [2001].

Bemerkung: $X_n \xrightarrow{\mathbb{P}} c$ konstant $\quad \not\Rightarrow \quad \mathbb{E}(X_n) \to c$

z.B. $\quad X_n = \begin{cases} 1 & \text{mit W'keit} \quad 1 - p_n \\ n & \text{mit W'keit} \quad p_n \end{cases}$

Wenn $p_n \to 0$, dann auch $\mathbb{P}(|X_n - 1| > \epsilon) \to 0 \quad \Rightarrow \quad X_n \xrightarrow{\mathbb{P}} 1$

Aber

$\mathbb{E}(X_n) = (1 - p_n) + n\, p_n = 1 + (n-1)\, p_n \to \infty \quad$ für $\quad p_n = \frac{1}{\sqrt{n}} \quad$ z.B.

Beispiel 11.4. $X_n \sim N(\mu, \sigma_n^2), n = 1, \ldots$

$X_n \xrightarrow{\mathbb{P}} \mu \quad \Leftrightarrow \quad \sigma_n^2 \to 0.$ *Denn*

$\mathbb{P}(|X_n - \mu| \leq \epsilon) = \mathbb{P}\left(\left|\underbrace{\frac{X_n - \mu}{\sigma_n}}_{\sim N(0,1)}\right| \leq \frac{\epsilon}{\sigma_n}\right) \to 1$ *wenn* $\frac{\epsilon}{\sigma_n} \to \infty$ *und somit* $\sigma_n \to 0$.

11.3 Konvergenz in Verteilung

Beispiel 11.5. *$X_n \sim N(0, \sigma_n^2)$ mit Verteilungsfunktion H_n mit $\sigma_n^2 \to \infty$*

$$H_n(x) = \mathbb{P}\left(\frac{X_n}{\sigma_n} \le \frac{x}{\sigma_n}\right) = \Phi\left(\frac{x}{\sigma_n}\right)$$

Mit $\sigma_n^2 \to \infty$ folgt $\frac{x}{\sigma_n} \to 0$ und somit $H_n(x) \to \Phi(0) = \frac{1}{2} \quad \forall x \in \mathbb{R}$. Deshalb

$$\lim_{n\to\infty} H_n(x) = H(x) \equiv \frac{1}{2}$$

und somit ist H keine Verteilungsfunktion und X_n konvergiert nicht in Verteilung.

Satz 11.7. *Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von ZV mit Verteilungsfunktion $X_n \sim F_n$. Dann existiert $X \sim F$ mit $X_n \stackrel{\mathcal{D}}{\to} X$ genau dann, wenn $\forall \epsilon > 0$ eine Konstante $k \in \mathbb{R}^+$ und $n_0 \in \mathbb{N}$ existiert, so daß*

$$\mathbb{P}(|X_n| \le k) \ge 1 - \epsilon \quad \forall n \ge n_0.$$

Beweis: Sei F derart, daß

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \forall x \in \mathbb{R}.$$

z.Z. F ist Verteilungsfunktion, also

- $F(x) \in [0, 1] \quad \surd$
- F monoton wachsend $\quad \surd$
- F rechtsstetig $\quad \surd$
- $\lim_{x\to-\infty} F(x) = 0, \ \lim_{x\to\infty} F(x) = 1$

$F_n(x) = \mathbb{P}(X_n \le x) \ge 1 - \epsilon \quad$ für $x \ge k, n > n_0$ und $\forall \epsilon > 0$

also $\lim_{n\to\infty} F_n(x) = F(x) \ge 1 - \epsilon \quad$ für $x \ge k$ und $\forall \epsilon > 0$

und $\lim_{x\to\infty} F(x) = 1 \quad$ für $\epsilon \downarrow 0$ □

Beispiel 11.6. *$X \sim U(0, 1)$*

$$X_n = \begin{cases} X & n \text{ gerade} \\ 1 - X & n \text{ ungerade} \end{cases}$$

Dann

$X_n \xrightarrow{\mathcal{D}} X$ *und*

$X_n \xrightarrow{\mathcal{D}} 1 - X$ *nämlich*

$X_n \xrightarrow{\mathcal{D}} U(0,1)$ *(d.h. jede ZV, welche gleichverteilt auf* $[0,1]$ *ist).*

$$\begin{aligned} Y_n \sim N(0,1) &\xrightarrow{\mathcal{D}} Y \sim N(0,1) \\ &\xrightarrow{\mathcal{D}} -Y \sim N(0,1) \end{aligned} \qquad \textit{Grenzwert nicht eindeutig!}$$

Beispiel 11.7. $X_n \sim N(\mu_n, \sigma_n^2) \quad \mu_n \to 0, \sigma_n^2 \to 1$

$\Rightarrow X_n \xrightarrow{\mathcal{D}} N(0,1)$ *weil*

$X_n \sim H_n$ *mit*

$$H_n(x) = \mathbb{P}\left(\frac{X_n - \mu_n}{\sigma_n} \le \frac{x - \mu_n}{\sigma_n}\right) = \Phi\left(\frac{x - \mu_n}{\sigma_n}\right)$$

Wegen $\dfrac{x - \mu_n}{\sigma_n} \to x$ *folgt aus der Stetigkeit von* Φ *auch* $H_n(x) \to \Phi(x)$ *und somit*
$X_n \xrightarrow{\mathcal{D}} N(0,1)$

Satz 11.8.

$$X_n \xrightarrow{\mathcal{D}} X,\ a, b \neq 0 \in \mathbb{R} \text{ konstant} \quad \Rightarrow \quad b\,X_n + a \xrightarrow{\mathcal{D}} b\,X + a.$$

Beweis: Sei o.B.d.A $b > 0$.

$$\mathbb{P}(b\,X_n + a \le x) = \mathbb{P}\left(X_n \le \frac{x-a}{b}\right) = F_n\left(\frac{x-a}{b}\right)$$
$$\to F\left(\frac{x-a}{b}\right) = \mathbb{P}\left(X \le \frac{x-a}{b}\right)$$ □

Satz 11.9. $X_n \xrightarrow{\mathcal{D}} X \quad \Leftrightarrow \quad \mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$

$\forall$ *beschränkten und stetigen Funktionen* $g : \mathbb{R} \to \mathbb{R}$.

Beweis: ohne Beweis □

Lemma 11.1.

$$X_n \xrightarrow{\mathcal{D}} X \quad \Leftrightarrow \quad \varphi_{X_n}(t) \to \varphi_X(t) \quad \forall t$$

Beweis: i.W. Konvergenz der $\mathbb{E}$-Werte $\quad \mathbb{E}(X_n^r) \to \mathbb{E}(X^r)$ □

Bemerkung: Die momenterzeugende Funktion ist hier eigentlich ungeeignet, weil alle Momente $\mathbb{E}(X_n^r)$, $r \geq 1$, existieren müssen. Besser wäre die charakteristische Funktion $\varphi_X(t) = \mathbb{E}(\exp(itX))$. Siehe Meintrup and Schäffler [2005].

Satz 11.10 (Slutsky's Theorem)**.**

$$X_n \xrightarrow{\mathcal{D}} X,\; A_n \xrightarrow{\mathbb{P}} a \textit{ und } B_n \xrightarrow{\mathbb{P}} b \quad \Rightarrow \quad A_n + B_n\, X_n \xrightarrow{\mathcal{D}} a + b\, X$$

Beweis: (Skizze)

Sei $Y_n \xrightarrow{\mathcal{D}} Y$ unabhängig von X_n, X. Dann

$\varphi_{(X_n,Y_n)}(s) \overset{\text{stu}}{=} \varphi_{X_n}(s) \cdot \varphi_{Y_n}(s)$

$\xrightarrow{\text{Lemma 11.1}} \varphi_X(t) \cdot \varphi_Y(t) = \varphi_{(X,Y)}(t)$

Mit $Y_n \xrightarrow{\mathbb{P}} a$ und Faltung folgt $X_n + Y_n \to X + a$, analog für $Y_n \xrightarrow{\mathbb{P}} b$. □

Evgeny Evgenievich Slutsky (7. April 1880 bis 10. März 1948)
Ukrainischer Ökonom und Statistiker.
http://en.wikipedia.org/wiki/Eugen_Slutsky

Beispiel 11.8. $X_n \sim Po(\lambda) \quad \mathbb{E}(X_n) = \lambda = \mathbb{V}(X_n)$

$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \lambda$

Was passiert mit $\quad \bar{X}_n - \lambda$?

$\bar{X}_n - \lambda \xrightarrow{\mathbb{P}} 0$

mit $\mathbb{E}(\bar{X}_n - \lambda) = 0$ *und*

$\mathbb{V}(\bar{X}_n - \lambda) = \mathbb{V}(\bar{X}_n) = \frac{1}{n^2}\mathbb{V}\left(\sum X_i\right) = \frac{1}{n^2}\, n\, \lambda = \frac{\lambda}{n} \to 0$

Aber $\quad \frac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}} \quad$ *hat Varianz* 1.

Was ist die Verteilung dieser Größe?

Satz 11.11 (Zentraler Grenzwertsatz)**.** *Seien $X_i, i = 1, \ldots$ unabhängig und identisch verteilt (u.i.v.) mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 < \infty \ \forall i = 1, \ldots$. Dann gilt*

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} N(0,1)$$

bzw.

$$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

Beweis:(Skizze) bekannt:

$$\begin{aligned} M_{\sum_{i=1}^n Y_i}(s) &\overset{Y_i \text{ stu}}{=} \prod_{i=1}^n M_{Y_i}(s) \\ &\overset{Y_i \text{ u.i.v.}}{=} (M_{Y_1}(s))^n \end{aligned}$$

Jetzt: $Y_i = \dfrac{X_i - \mu}{\sigma}$

$\mathbb{E}(Y_i) = 0, \mathbb{V}(Y_i) = 1$

$e^{t\,y} \overset{\text{Taylorreihe}}{=} 1 + t\,y + \dfrac{1}{2} t^2\, y^2 + \text{Rest}$

$$\begin{aligned} \Rightarrow M_{Y_1}(s) &= \mathbb{E}(e^{s\,Y_1}) \\ &= 1 + s\,\mathbb{E}(Y_1) + \frac{1}{2}\,s^2\,\mathbb{E}(Y_1^2) + \mathbb{E}(\text{Rest}) \\ &= 1 + \frac{1}{2}\,s^2 + \mathbb{E}(\text{Rest}) \\ \Rightarrow M_{\sum Y_i}(s) &= \left[1 + \frac{1}{2}\,s^2 + \mathbb{E}(\text{Rest})\right]^n \\ \Rightarrow M_{\frac{1}{\sqrt{n}}\sum Y_i}(s) &= \left[1 + \frac{1}{2\,n}\,s^2 + \mathbb{E}\left(\frac{\text{Rest}}{n}\right)\right]^n \end{aligned}$$

mit $\mathbb{E}\left(\dfrac{\text{Rest}}{n}\right) \to 0$ (o.B.)

und damit

$$M_{\frac{1}{\sqrt{n}}\sum Y_i}(s) = \left(1 + \frac{\frac{1}{2}\,s^2}{n}\right)^n \to \exp\left(\frac{1}{2}\,s^2\right) \quad \text{für } n \to \infty.$$

Die Behauptung folgt aus $M_{\mathrm{N}(0,1)}(s) = \exp\left(\dfrac{s^2}{2}\right)$. Um den Beweis vollständig zu haben, muß man von der momenterzeugenden Funktion zur charakteristischen Funktion übergehen. □

Bemerkung: Der ZGWS gilt für alle X_i mit beliebiger Verteilung!

Beispiel 11.9 (Bernoulliverteilung)**.** $X_i \sim B(1,p), i = 1,\dots,n$ *mit* $p = 0.5$*. Die Verteilung des standardisierten Mittelwertes konvergiert nach dem Zentralen Grenzwertsatz gegen eine Normalverteilung*

$$\sqrt{n}\frac{n^{-1}\sum_{i=1}^{n} X_i - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} N(0,1)$$

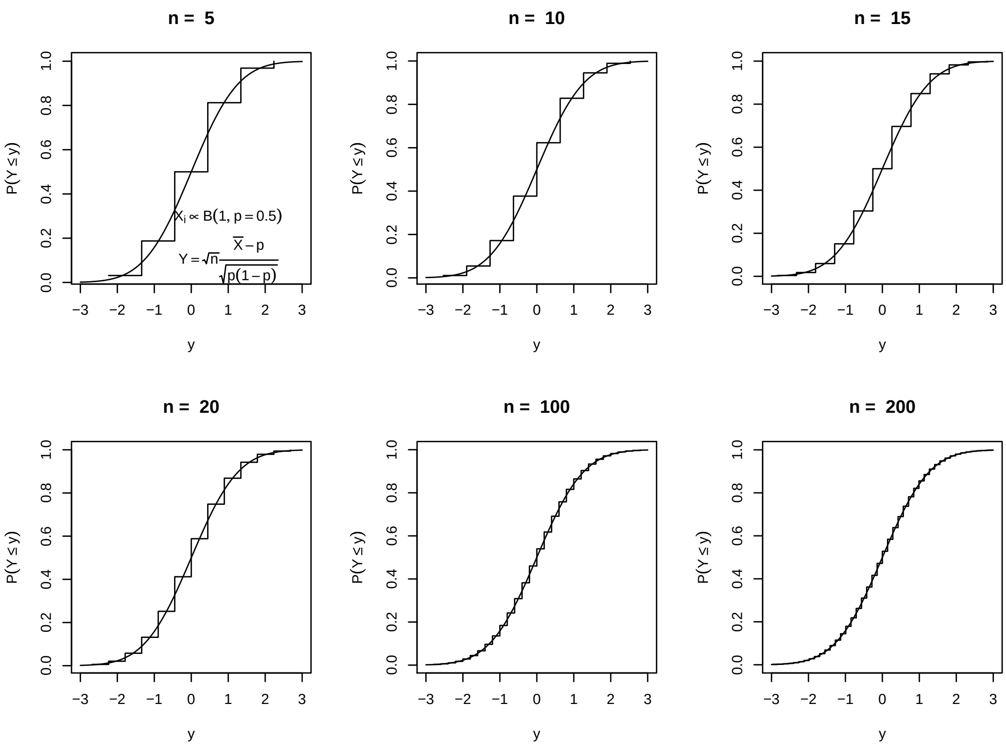


Abbildung 11.1: Approximation der Binomialverteilung durch die Normalverteilung.

Beispiel 11.10 (Poissonverteilung)**.** $X_i \sim Po(\lambda), i = 1,\dots,n$ *mit* $\lambda = 5$*. Die Verteilung des standardisierten Mittelwertes konvergiert nach dem Zentralen Grenzwertsatz gegen eine Normalverteilung*

$$\sqrt{n}\frac{n^{-1}\sum_{i=1}^{n} X_i - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0,1)$$

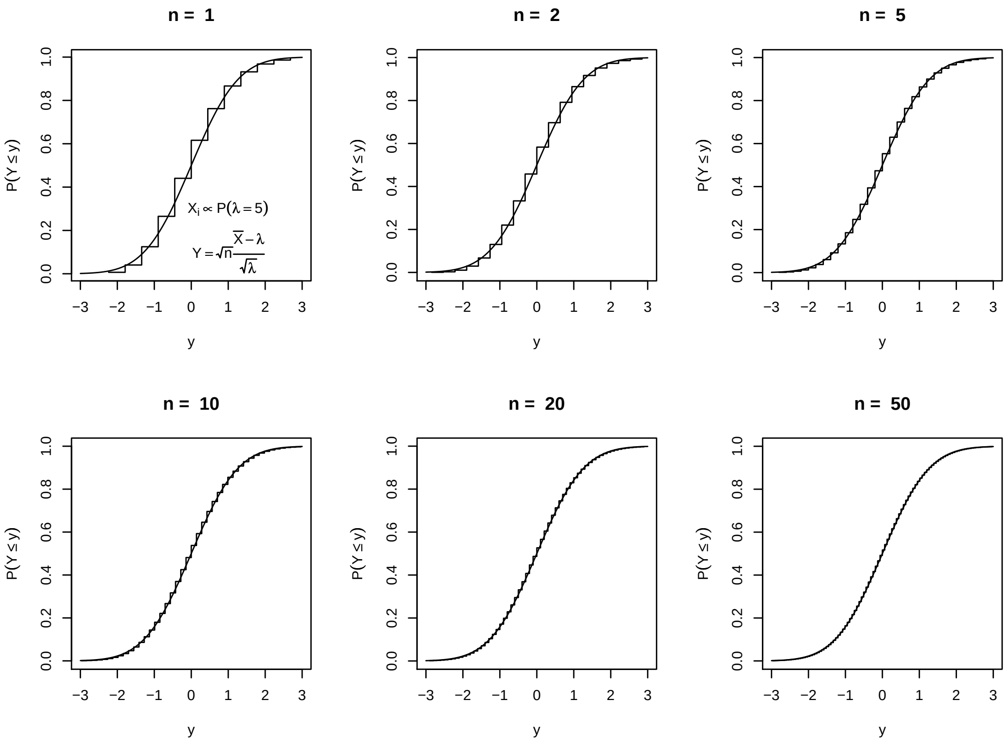


Abbildung 11.2: Approximation der Poissonverteilung durch die Normalverteilung.

Beispiel 11.11 (χ^2-Verteilung)**.** *$X \sim \chi^2_n$ mit $\mathbb{E}(X) = n$ und $\mathbb{V}(X) = 2n$. Betrachte die Dichte der standardisierten ZV*

$$Y = \frac{X - n}{\sqrt{2n}}$$

also die Transformation $y = g(x) = (x - n)(2n)^{-1/2}$. Mit dem Transformationssatz folgt

$$f_Y(y) = f_{\chi^2_n}(y\sqrt{2n} + n)\sqrt{2n}$$

Da X die Summe von unabhängigen standardnormalverteilten ZV ist, folgt mit dem ZGWS, daß die zugehörigen Verteilungsfunktionen konvergieren, und zwar $F_Y(y) \to \Phi(y) \iff Y \xrightarrow{\mathcal{D}} N(0,1)$.

Beispiel 11.12. *$X_1, \dots, X_n$ iid mit $\mathbb{E}(X_i) = 0$, $\mathbb{V}(X_i) = \sigma^2$ und $\mathbb{V}(X_i^2) = \tau^2 < \infty$.*

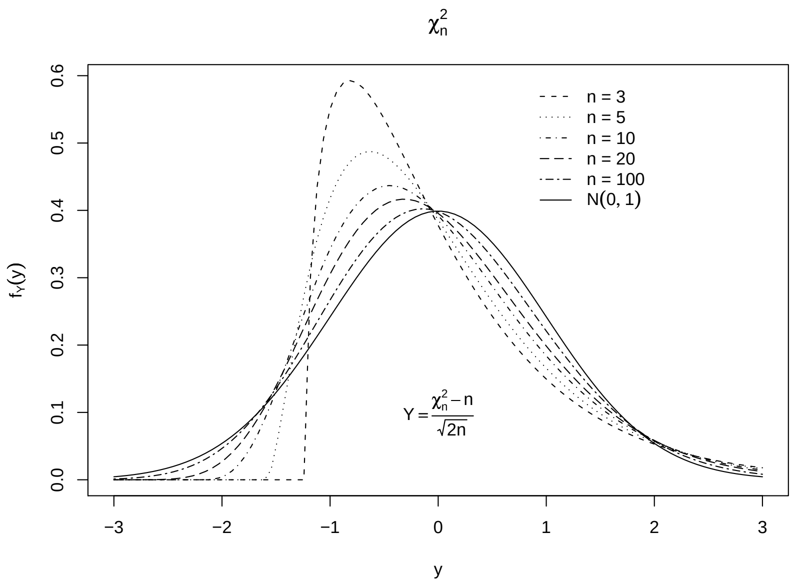


Abbildung 11.3: Approximation der χ^2-Verteilung durch die Normalverteilung.

$\sqrt{n}\,(S^2 - \sigma^2) \xrightarrow{\mathcal{D}} \quad ?$

$S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{\mathbb{P}} \sigma^2$ *(siehe HA)*

S^2 hängt nicht von μ ab, also o.B.d.A. $\mu = 0$

ZGWS: $\quad \sqrt{n}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} N(0, \tau^2)$

Gesetz der großen Zahlen: $\quad \bar{X}^2 \xrightarrow{\mathbb{P}} 0$

Damit folgt: $\quad \sqrt{n}\,(S^2 - \sigma^2) = \sqrt{n}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2 - \underbrace{\bar{X}^2}_{\to 0} - \sigma^2\right) \xrightarrow{\mathcal{D}} N(0, \tau^2)$

Satz 11.12 (Berry-Esseen). *Es seien $X_1, \ldots, X_n$ u.i.v., $X_i \sim F$ mit $\mathbb{E}(X_i) = \mu$, $\mathbb{V}(X_i) = \sigma^2$, $\mathbb{E}(X_i^3) < \infty \quad \Rightarrow$*

$\exists\, c$ (unabhängig von F) mit

$$|G_n(x) - \Phi(x)| \leq \frac{c}{\sqrt{n}} \frac{\mathbb{E}\left(|X_1 - \mu|^3\right)}{\sigma^3}$$

wobei $\quad G_n(x) = \mathbb{P}\left(\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \leq x\right).$

Beweis: geschenkt □

Bemerkung: Satz 11.12 gilt für alle n, nicht nur für $n \to \infty$! c ist unabhängig von F und es ist bekannt, daß $c_{\min} \leq 0.7975$.
F darf ohne zusätzliche Annahmen nicht von n abhängen!

Beispiel 11.13. $X_1, \ldots, X_n, \quad X_i \sim Po\left(\frac{1}{n}\right)$

$\Rightarrow \quad \sum_{i=1}^n X_i \sim Po(1)$

Aber $\quad \sqrt{n}\,\frac{\left(\bar{X} - \frac{1}{n}\right)}{\sqrt{\frac{1}{n}}} = \frac{\sum X_i - n\,\frac{1}{n}}{\sqrt{n}\,\frac{1}{\sqrt{n}}} = \sum X_i - 1 \quad \sim \quad Po(1) - 1 \neq N(0,1) \quad \forall n$

Satz 11.13. *Es seien $X_1, \ldots, X_n$ u.i.v. mit $X_i \sim F$, $\mathbb{E}(X_i) = \mu$, $\mathbb{V}(X_i) = \sigma^2$, $\mathbb{E}(X_i^3) < \infty$ und k-tem zentralen Moment $\mu_k^0 = \mathbb{E}((X_i - \mu)^k)$. Dann gilt:*

$$\begin{aligned} G_n(x) &= \mathbb{P}\left(\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \leq x\right) \\ &= \Phi(x) + \underbrace{\frac{\mu_3^0}{6\,\sigma^3\,\sqrt{n}}\,(1 - x^2)\,\varphi(x)}_{\textit{erste Edgeworth-Korrektur}} + o\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

Bemerkung:

- $o\left(\frac{1}{\sqrt{n}}\right)$ meint eine Folge, die wie $\frac{1}{\sqrt{n}} \to 0$ für $n \to \infty$, aber schneller als $\frac{1}{\sqrt{n}}$.
- für symmetrische Verteilungen ist $\mu_3 = 0$ und damit ist z.B. für die t-Verteilung die Approximation durch N$(0,1)$ schneller als $\frac{1}{\sqrt{n}}$. Die erste Edgeworth-Korrektur ist also für schiefe Verteilungen, etwa $\chi^2\,(n)$ von Interesse.

Satz 11.14 (Δ-Methode)**.** *Falls* $\quad \sqrt{n}\,(X_n - \nu) \xrightarrow{\mathcal{D}} N(0, \tau^2)$

dann auch $\quad \sqrt{n}\,(f(X_n) - f(\nu)) \xrightarrow{\mathcal{D}} N\left(0, \tau^2 \cdot \frac{\partial f(\nu)^2}{\partial \nu}\right)$

falls $f : \mathbb{R} \to \mathbb{R}$ *in* ν *differenzierbar ist und* $\frac{\partial f(\nu)}{\partial \nu} \neq 0.$

Beweis: Entwickle f in Taylorreihe

$$f(x+\Delta) = f(x) + \Delta\, \frac{\partial f(x)}{\partial x}\bigg|_{x=x} + \ldots + \frac{\Delta^r}{r!}\, \frac{\partial^r f(x)}{\partial^r x}\bigg|_{x=x} + \text{Rest}$$

mit $\Delta = X_n - \nu$ und $x = \nu$ folgt

$$f(X_n) = f(\nu) + (X_n - \nu)\,\frac{\partial f(\nu)}{\partial \nu} + \text{Rest}$$

$$\Rightarrow \sqrt{n}\,(f(X_n) - f(\nu)) = \underbrace{\sqrt{n}\,(X_n - \nu) \cdot \frac{\partial f(\nu)}{\partial \nu}}_{\text{linear}} + \underbrace{\text{Rest}}_{\to 0 \text{ ohne Beweis}}$$

$$\overset{11.8}{\Rightarrow} \sqrt{n}\,(f(X_n) - f(\nu)) \xrightarrow{\mathcal{D}} \mathrm{N}\left(0, \tau^2 \cdot \frac{\partial f(\nu)^2}{\partial \nu}\right) \qquad \square$$

Beispiel 11.14. $X_1, \ldots, X_n$ *u.i.v.* $\quad X_i \sim Po(\lambda)$

$\overset{ZGWS}{\Rightarrow} \quad \sqrt{n}\,(\bar{X} - \lambda) \xrightarrow{\mathcal{D}} N(0, \lambda)$

Problem: Varianz der Grenzverteilung hängt von unbekanntem λ *ab.*

Suche f *mit*

$$\sqrt{n}\,(f(\bar{X}) - f(\lambda)) \xrightarrow{\mathcal{D}} N\left(0, \lambda \cdot \frac{\partial f(\lambda)^2}{\partial \lambda}\right)$$

wobei $\quad \lambda \cdot \frac{\partial f(\lambda)^2}{\partial \lambda} \equiv$ *konstant.*

zum Beispiel $\quad \frac{\partial f(\lambda)}{\partial \lambda} = \frac{c}{\sqrt{\lambda}}$

$\Rightarrow f(\lambda) = 2\,c\sqrt{\lambda} \quad$ *denn* $c \int \frac{1}{\sqrt{\lambda}}\,d\lambda = c \cdot 2\sqrt{\lambda}$

und damit $(c \equiv 1)$

$$2\sqrt{n}\left(\sqrt{\bar{X}} - \sqrt{\lambda}\right) \xrightarrow{\mathcal{D}} N(0,1)$$

Beispiel 11.15. $X_1, \ldots, X_n$ *u.i.v.* $\quad X_i \sim N(0, \sigma^2)$

$\Rightarrow \mathbb{E}(X_i^2) = \sigma^2,\ \mathbb{V}(X_i^2) = 2\,\sigma^4$

$$\overset{\text{ZGWS}}{\Rightarrow} \quad \sqrt{n}\left(\frac{1}{n}\sum X_i^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} N(0, 2(\sigma^2)^2)$$

Gesucht: f *mit* $\dfrac{\partial f(\sigma^2)}{\partial \sigma^2} = \dfrac{c}{\sqrt{2}\,\sigma^2}$

$$f = \int \frac{\partial f(\sigma^2)}{\partial \sigma^2}\, d\sigma^2 = \frac{c}{\sqrt{2}} \int \frac{1}{\sigma^2}\, d\sigma^2 = \frac{c}{\sqrt{2}} \log(\sigma^2)$$

Setzt $c = 1 \quad \Rightarrow$

$$\sqrt{\frac{n}{2}} \log\left(\frac{\frac{1}{n}\sum X_i^2}{\sigma^2}\right) \xrightarrow{\mathcal{D}} N(0,1)$$

Definition 11.2. *Seien* $X_1, \dots, X_n$ *u.i.v. mit Verteilungsfunktion* F*. Dann heißt*

$$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

die *empirische Verteilungsfunktion* *von* $X_1, \dots, X_n$*.*

Satz 11.15 (Satz von Glivenko-Cantelli)**.**

a) $\widehat{F}(x) \xrightarrow{f.s.} F(x) \quad \forall x \in \mathbb{R}$

b) $\sqrt{n}\,(\widehat{F}_n(x) - F(x)) \xrightarrow{\mathcal{D}} N(0, F(x)(1-F(x))) \quad \forall x \in \mathbb{R}$

c) $\sup_{x\in\mathbb{R}} |\widehat{F}_n(x) - F(x)| \xrightarrow{f.s.} 0$

Beweis: $I_{X_i \le x} \sim \mathrm{B}(1, F(x))$ u.i.v. $\quad \forall i = 1, \dots, n$ und $x \in \mathbb{R}$.

$$\overset{\text{ZGWS}}{\Rightarrow} \quad \underbrace{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} I_{X_i\le x} - F(x)\right)}_{=\sqrt{n}\,(\widehat{F}_n(x) - F(x))} \xrightarrow{\mathrm{d}} \mathrm{N}(0, F(x)(1-F(x)))$$

$\Rightarrow$ b)

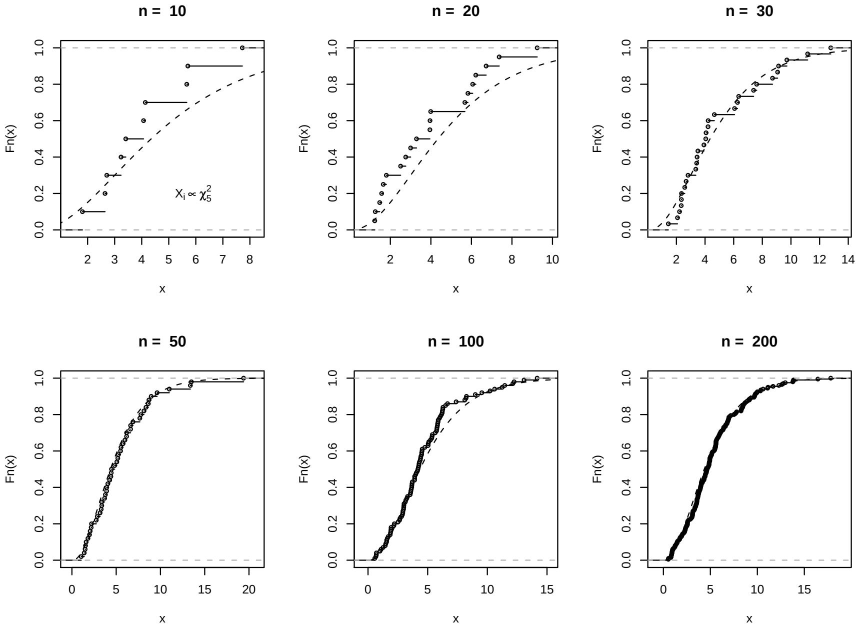
$\widehat{F}_n(x) = \frac{1}{n}\sum_{X_i \le x} \xrightarrow{\mathbb{P}} \mathbb{E}(I_{X_i \le x}) = \mathbb{P}(X_i \le x) = F(x)$

$\Rightarrow \widehat{F}_n(x) \xrightarrow{\mathbb{P}} F(x)$ (f.s. braucht ein starkes Gesetz der großen Zahlen mit $\xrightarrow{\text{f.s.}}$ statt $\xrightarrow{\mathbb{P}}$).

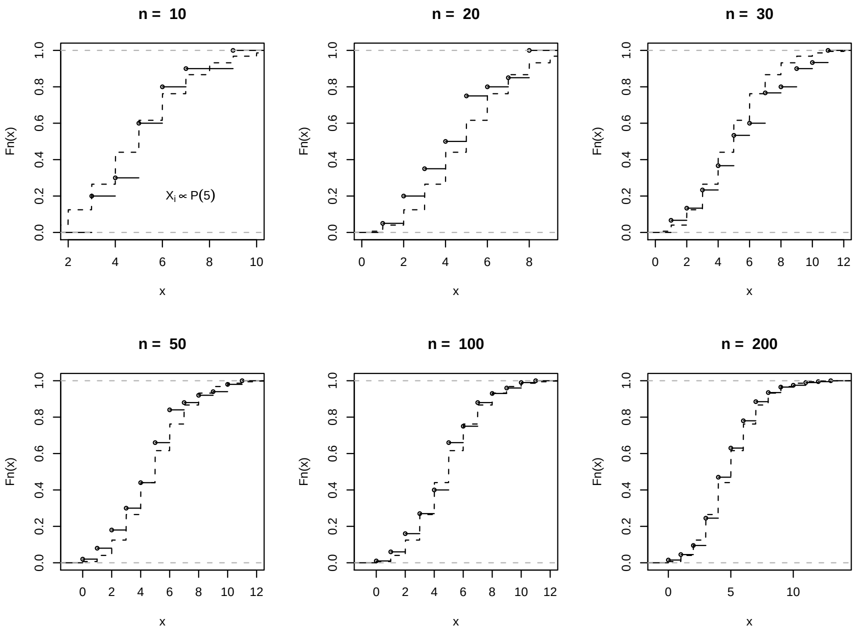
c) ohne Beweis $\square$

Waleri Iwanowitsch Gliwenko (2. Januar 1897 bis 15. Februar 1940).
Russischer Mathematiker und Logiker.
http://de.wikipedia.org/wiki/Waleri_Iwanowitsch_Gliwenko

Beispiel 11.16. *(Glivenko-Cantelli)*



Approximation der Verteilungsfunktion der $\chi^2(5)$ Verteilung durch die empirische Verteilungsfunktion.



Approximation der Verteilungsfunktion der $Po(5)$ Verteilung durch die empirische Verteilungsfunktion.

Kapitel 12

Spezielle Verteilungen und deren Eigenschaften

12.1 Diskrete Verteilungen

Satz 12.1 (Erwartungswert und Varianz der B(n,p))**.** *Seien $X_1,\dots,X_n$ stu mit $X_i \sim B(1,p),\ i=1,\dots,n$. Dann gilt für $X=\sum_{i=1}^n X_i$*

$$\begin{aligned}\mathbb{E}(X) &= n\cdot p\\ \mathbb{V}(X) &= n\cdot p\cdot(1-p)\end{aligned}$$

Beweis: $\mathbb{E}(X)=\mathbb{E}(\sum_{i=1}^n X_i)=\sum_{i=1}^n \underbrace{\mathbb{E}(X_i)}_{\equiv p}=n\cdot p$

$\mathbb{V}(X)=\mathbb{V}(\sum_{i=1}^n X_i)\overset{\text{stu}}{=}\sum_{i=1}^n \underbrace{\mathbb{V}(X_i)}_{p\,(1-p)}=n\cdot p\cdot(1-p)$ □

Satz 12.2 (Zusammenhang zwischen B$(1,p)$ und B(n,p))**.** *Seien $X_1,\dots,X_n$ stu mit $X_i\sim B(1,p)$. Dann gilt:*

$$X=\sum_{i=1}^n X_i \sim B(n,p).$$

Beweis: $\mathbb{P}(X_1=1,X_2=1,\dots,X_m=1,X_{m+1}=0,\dots,X_n=0)\overset{\text{stu}}{=}\prod_{i=1}^m \mathbb{P}(X_i=1)\prod_{i=m+1}^n \mathbb{P}(X_i=0)$

$=\prod_{i=1}^m p\prod_{i=m+1}^n (1-p)=p^m\,(1-p)^{n-m}$

Da es $\binom{n}{m}$ Möglichkeiten gibt, die 1/0 zu permutieren, folgt

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i = m\right) = \binom{n}{m} p^m (1-p)^{n-m}$$

$\Rightarrow \sum_{i=1}^{n} X_i \sim \mathrm{B}(n,p)$ □

Korollar 12.1. *$X_1 \sim B(n_1,p)$, $X_2 \sim B(n_2,p)$ stu. $\Rightarrow X_1+X_2 \sim B(n_1+n_2,p)$ Summenstabilität*

Definition 12.1 (Hypergeometrische Verteilung)**.** *Modell: Ziehen ohne Zurücklegen $X \sim Hyp(n,K,N)$,*

$$\mathbb{P}(X=k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Satz 12.3 (Erwartungswert und Varianz von $X \sim \mathrm{Hyp}(n,K,N)$)**.** *$X \sim Hyp(n,K,N)$*

$$\begin{aligned}\mathbb{E}(X) &= n \cdot \frac{K}{N}\\ \mathbb{V}(X) &= n \cdot \frac{K}{N}\left(1-\frac{K}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)\end{aligned}$$

Bemerkung: $\mathbb{E}(X)$ ist analog zu $\mathrm{B}(n,p)$, da $\frac{K}{N}$ der Erfolgswahrscheinlichkeit p entspricht.
$\mathbb{V}(X)$ ist für $n > 1$ kleiner als die Varianz von $\mathrm{B}(n,p)$
$\Rightarrow$ *Endlichkeitskorrektur*

Beweis: Wir betrachten die einzelnen Ziehungen mit $Y_1,\dots,Y_n$, $X = \sum_{i=1}^{n} Y_i$, Y_i binär.
Betrachte gemeinsame Verteilung der Y_i z.B. $k=2$, $n=4$,

$\mathbb{P}(Y_1=1,Y_2=1,Y_3=0,Y_4=0) = \frac{K}{N} \cdot \frac{K-1}{N-1} \cdot \frac{N-K}{N-2} \cdot \frac{N-K-1}{N-3}$

$\mathbb{P}(Y_1=1,Y_2=0,Y_3=0,Y_4=1) = \frac{K}{N} \cdot \frac{N-K}{N-1} \cdot \frac{N-K-1}{N-2} \cdot \frac{K-1}{N-3}$

Wahrscheinlichkeit ist für alle Varianten mit $\sum Y_n = k$ identisch.

$\mathbb{P}(Y_1=i_1,\dots,Y_n=i_n) = \frac{K\cdots(K-k+1)(N-K)\cdots(N-K-(n-k)+1)}{N(N-1)\cdots(N-n+1)}$ für $\sum_{j=1}^{n} i_j = k$
$\Rightarrow$ gemeinsame Verteilung von $Y_1,\dots,Y_n$ ist vertauschbar,
d.h. $\mathbb{V}(Y_1,\dots,Y_n) = \mathbb{V}(Y_{\pi(1)},\dots,Y_{\pi(n)})$
$\Rightarrow$ Alle Randverteilungen sind identisch

$Y_i \sim \mathrm{B}(1, \frac{K}{N})$

$$\begin{aligned}
\mathbb{E}(\sum_{i=1}^n Y_i) &= \sum_{i=1}^n \mathbb{E}(Y_i) = n \cdot \frac{K}{N} \\
\mathbb{V}(\sum_{i=1}^n Y_i) &= \sum_{i=1}^n \mathbb{V}(Y_i) + \sum_{i \neq j, i<j} 2 \cdot \mathrm{Cov}(Y_i, Y_j)
\end{aligned}$$

Es gilt
$\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}(Y_1, Y_2) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1) \cdot \mathbb{E}(Y_2)$

$= \frac{K}{N} \cdot \frac{K-1}{N-1} - \left(\frac{K}{N}\right)^2 = \frac{K}{N} \cdot \left(\frac{K-1}{N-1} - \frac{K}{N}\right)$

(weil $Y_1 Y_2 = 1 \iff Y_1 = 1$ und $Y_2 = 1$ und damit $\mathbb{E}(Y_1 Y_2) = \mathbb{P}(Y_1 = 1, Y_2 = 1) = \frac{K}{N} \cdot \frac{K-1}{N-1}$)

$\mathbb{V}(\sum_{i=1}^n Y_i) = n \cdot \frac{K}{N} \left(1 - \frac{K}{N}\right) + \frac{n(n-1)}{2} \cdot 2 \frac{K}{N} \left(\frac{K-1}{N-1} - \frac{K}{N}\right)$
$= n \cdot \frac{K}{N} \left(1 - \frac{K}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$

□

Satz 12.4 (Zusammenhang zwischen $\mathrm{B}(n,p)$ und $\mathrm{Hyp}(n,K,N)$)**.** *Seien $K_m, N_m \in \mathbb{N}$ mit $0 \leq K_m \leq N_m$ Folgen ($m \in \mathbb{N}$) mit $K_m \to \infty, N_m \to \infty$ und $p \in [0,1], n \in \mathbb{N}$ fest. Dann gilt mit $\dfrac{K_m}{N_m} \to p$ für $m \to \infty$*

$$\lim_{m \to \infty} \mathbb{P}(H_m = k) = \mathbb{P}(B = k) \quad \forall k \in \mathbb{N}$$

für $H_m \sim Hyp(n, K_m, N_m)$ und $B \sim B(n,p)$.

Beweis: Dichte

$$\begin{aligned}
\mathbb{P}(H_m = k) &= h(k, n, K_m, N_m) = \frac{\binom{K_m}{k}\binom{N_m - K_m}{n-k}}{\binom{N_m}{n}} \\
&= \frac{n!}{k!\,(n-k)!} \cdot \frac{\frac{K_m!}{(K_m-k)!} \cdot \frac{(N_m-K_m)!}{((N_m-K_m)-(n-k))!}}{\frac{N_m!}{(N_m-n)!}} \\
&= \binom{n}{k} \underbrace{\frac{K_m}{N_m}}_{\to p} \cdot \underbrace{\frac{K_m - 1}{N_m - 1}}_{\to p} \cdots \underbrace{\frac{K_m - (k-1)}{N_m - (k-1)}}_{\to p} \cdot \\
&\quad \underbrace{\frac{(N_m - K_m)}{N_m - k}}_{\to 1-p} \cdot \underbrace{\frac{N_m - K_m - 1}{N_m - k - 1}}_{\to 1-p} \cdots \underbrace{\frac{N_m - K_m - (n-k-1)}{N_m - (n-1)}}_{\to 1-p}
\end{aligned}$$

$$\overset{m\to\infty}{\Rightarrow} \binom{n}{k} p^k (1-p)^{n-k} = b(k,n,p) = \mathbb{P}(B=k) \qquad \square$$

Satz 12.5 (Zusammenhang zwischen B(n,p) und Po(λ))**.** *Sei $p_n \in [0,1], n \in \mathbb{N}$, eine Folge und $\lambda_n = n \cdot p_n \to \lambda > 0$ für $n \to \infty$. Dann gilt*

$$b(k,n,p_n) \overset{n\to\infty}{\to} e^{-\lambda} \frac{\lambda^k}{k!} \quad \forall k \in \mathbb{N}.$$

Beweis:

$$\begin{aligned}
b(k,n,p_n) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\
&= \frac{n!}{k!\,(n-k)!} \left(\frac{\lambda_n}{n}\right)^k \left(1-\frac{\lambda_n}{n}\right)^{n-k} \\
&= \frac{\lambda_n^k}{k!} \left(1-\frac{\lambda_n}{n}\right)^n \frac{n\,(n-1)\cdots(n-k+1)}{n^k} \left(1-\frac{\lambda_n}{n}\right)^{-k} \\
&= \frac{\lambda_n^k}{k!} \left(1-\frac{\lambda_n}{n}\right)^n \frac{\frac{n}{n}\,\frac{n-1}{n}\,\frac{n-2}{n}\cdots\frac{n-k+1}{n}}{\left(1-\frac{\lambda_n}{n}\right)^k} \\
&\overset{n\to\infty}{\to} \frac{\lambda^k}{k!} e^{-\lambda} \underbrace{\frac{1\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda_n}{n}\right)^k}}_{\to 1 \text{ da } \frac{1}{n}\ldots\frac{\lambda_n}{n}\to 0} \\
&= \frac{\lambda^k}{k!} e^{-\lambda} = f(k) \quad \widehat{=} \quad \text{Dichte der Poissonverteilung mit Parameter } \lambda.
\end{aligned}$$

$\square$

Definition 12.2 (Geometrische Verteilung)**.** *Die diskrete Verteilung mit Parameter $p \in [0,1]$ und Dichte*

$$\mathbb{P}(X=x) = (1-p)^{x-1}\,p \quad x \in \mathbb{N}$$

heißt <u>geometrische Verteilung</u>, kurz $X \sim Geom(p)$, und beschreibt die Wahrscheinlichkeit der Anzahl von Versuchen bis zum ersten Erfolg.

Satz 12.6 (Eigenschaften der Geometrischen Verteilung)**.** $X \sim Geom(p)$

(a) $\mathbb{E}(X) = \frac{1}{p}$ *und* $\mathbb{V}(X) = \frac{1-p}{p^2}$

(b) *Gedächtnislosigkeit* $\mathbb{P}(X = x + x_0 | X > x_0) = \mathbb{P}(X = x)$

Definition 12.3 (Negative Binomialverteilung)**.** *Die diskrete Verteilung mit Parametern $p \in [0,1]$ und $n \in \mathbb{N}$ sowie der Dichte*

$$\mathbb{P}(X=x) = \binom{x-1}{n-1} p^n \, (1-p)^{x-n} \quad \textit{für } x \in \mathbb{N}, x \geq n$$

heißt <u>*negative Binomialverteilung,*</u> *kurz $X \sim NegBin(n,p)$.*

Bemerkung: X beschreibt die Anzahl von Versuchen, die zur Erreichung von n Erfolgen notwendig sind, wenn die Wahrscheinlichkeit für einen Erfolg gleich p ist. Die Dichte kann folgendermaßen hergeleitet werden. Sei Y die Anzahl von Erfolgen bei $x-1$ Versuchen; $Y \sim \mathrm{B}(x-1,p)$. Die Wahrscheinlichkeit, $n-1$ Erfolge bei $x-1$ Versuchen erzielt zu haben, ist

$$\begin{aligned}
\mathbb{P}(Y=n-1) &= b(x-1,n-1,p) = \binom{x-1}{n-1} p^{n-1}(1-p)^{x-1-(n-1)} \\
&= \binom{x-1}{n-1} p^{n-1}(1-p)^{x-n}.
\end{aligned}$$

Die Wahrscheinlichkeit, im $x$ten Versuch einen Erfolg und damit insgesamt n Erfolge zu erzielen, ist p, also

$$\mathbb{P}(X=x) = \mathbb{P}(Y=n-1)p = \binom{x-1}{n-1} p^n (1-p)^{x-n}.$$

Satz 12.7 (Zusammenhang zwischen Geom(p) und NegBin(n,p))**.** *Seien $X_1,\ldots,X_n$ stu mit $X_i \sim Geom(p), i=1,\ldots,n$. Dann gilt*

$$X = \sum_{i=1}^{n} X_i \sim NegBin(n,p).$$

Korollar 12.2. *$X \sim NegBin(n,p) \Rightarrow \mathbb{E}(X) = \frac{n}{p}$*
$X = \sum_{i=1}^{n} X_i \; X_i \sim Geom(p) \quad stu.$
$\mathbb{E}(X) = n \cdot \mathbb{E}(X_i) = \frac{n}{p}$ und $\mathbb{V}(X) = \frac{n(1-p)}{p^2}$
Negative Binomialverteilung ist nicht gedächtnislos.

Beweis: Zunächst $n=2$.

$$f_{X_1,X_2}(x_1,x_2) \overset{\text{stu}}{=} (1-p)^{x_1-1} \cdot p \cdot (1-p)^{x_2-1} \cdot p$$

$$
\begin{aligned}
\mathbb{P}(X_1 + X_2 = x) &= \mathbb{P}\left(\{\omega \in \Omega \mid X_1(\omega) + X_2(\omega) = x\}\right) \\
&= \mathbb{P}\left(\dot{\bigcup_{x_1 \in \mathbb{N}}} \{\omega \in \Omega \mid X_1(\omega) = x_1 \wedge X_2(\omega) = x - x_1\}\right) \\
&= \sum_{x_1 \in \mathbb{N}} \mathbb{P}\left(\{\omega \in \Omega \mid X_1(\omega) = x_1 \wedge X_2(\omega) = x - x_1\}\right) \\
&= \sum_{x_1 \in \mathbb{N}} \mathbb{P}(X_1 = x_1,\, X_2 = x - x_1) \\
&\overset{\text{alles diskret}}{=} \sum_{x_1 \in \mathbb{N}} f_{X_1,X_2}(x_1, x - x_1) \quad \text{"Faltung" im diskreten Fall} \\
&\overset{\text{stu}}{=} \sum_{x_1 \in \mathbb{N}} f_{X_1}(x_1) \cdot f_{X_2}(x - x_1) \\
&\overset{f_{X_2}(x-x_1)=0 \text{ für } x \le x_1}{=} \sum_{x_1=1}^{x-1} (1-p)^{x_1-1}\, p^2\, (1-p)^{x-x_1-1} \\
&= p^2 \sum_{x_1=1}^{x-1} (1-p)^{x_1-1+(x-x_1-1)} \\
&= p^2 \sum_{x_1=1}^{x-1} (1-p)^{x-2} \\
&= p^2\, (x-1) \cdot (1-p)^{x-2} \\
&\overset{n=2}{=} \binom{x-1}{n-1} p^n\, (1-p)^{x-n} \\
&\mathrel{\widehat{=}} \text{Dichte der NegBin}(2, p)
\end{aligned}
$$

Rest: VI □

Definition 12.4 (Multinomialverteilung)**.** *Sei $n \in \mathbb{N}$ und $p_1, \dots, p_k \in [0,1]$ mit $\sum_{j=1}^k p_j = 1$. Sei $Y = (Y_1, \dots, Y_k)$ eine k-dimensionale diskrete ZV mit Dichte*

$$
\begin{aligned}
f_Y(y) &= \mathbb{P}(Y_j = y_j \quad \forall j = 1, \dots, k) \\
&= \begin{cases} \dfrac{n!}{y_1! \cdots y_k!}\, p_1^{y_1} \cdots p_k^{y_k} & \textit{falls } \sum_{j=1}^k y_j = n \\ 0 & \textit{sonst.} \end{cases}
\end{aligned}
$$

Dann heißt die zugehörige Verteilung Multinomialverteilung mit Parametern n und $p_1, \dots, p_k$: $\quad Y \sim MN(n, p_1, \dots, p_k)$.

Bemerkung: Es gengt, $p_1, \ldots, p_{k-1}$ anzugeben, denn

$$p_k = 1 - \sum_{j=1}^{k-1} p_j.$$

Satz 12.8 (Eigenschaften der Multinomialverteilung)**.**

i) $Y_i \sim B(n, p_i)$

ii) $\mathbb{E}(Y) = n\,(p_1, \ldots, p_k)$

iii) $\mathbb{V}(Y) = n \begin{pmatrix} p_1\,(1-p_1) & -p_1\,p_2 & \cdots & -p_1\,p_k \\ -p_1\,p_2 & p_2\,(1-p_2) & \cdots & -p_2\,p_k \\ \vdots & & & \\ -p_1\,p_k & -p_2\,p_k & \cdots & p_k\,(1-p_k) \end{pmatrix}$

Beweis:

i)

$$\begin{aligned} \mathbb{P}(Y_i = y_i) &= \mathbb{P}(Y_i = y_i, \sum_{j \neq i} Y_j = n - y_i) \\ &\overset{\mathrm{MN}(n, p_i, 1-p_i)}{=} \frac{n!}{y_i!\,(n-y_i)!}\, p_i^{y_i}\,(1-p_i)^{n-y_i} \\ &= \binom{n}{y_i} p_i^{y_i}\,(1-p_i)^{n-y_i} \end{aligned}$$

$\Rightarrow Y_i \sim \mathrm{B}(n, p_i)$

ii)

$$\begin{aligned} \mathbb{E}(Y) &= (\mathbb{E}(Y_1), \mathbb{E}(Y_2), \ldots, \mathbb{E}(Y_n)) \\ &\overset{Y_i \sim \mathrm{B}(n, p_i)}{=} n(p_1, \ldots, p_k) \end{aligned}$$

iii) $Y_i \sim \mathrm{B}(n, p_i) \quad \Rightarrow \quad \mathbb{V}(Y_i) = n\,p_i\,(1-p_i)$

$$\begin{aligned} \mathbb{V}(Y_i + Y_j) &= \mathbb{V}(Y_i) + \mathbb{V}(Y_j) + 2\,\mathrm{Cov}(Y_i, Y_j) \\ &= n\,p_i\,(1-p_i) + n\,p_j\,(1-p_j) + 2\,\mathrm{Cov}(Y_i, Y_j) \end{aligned}$$

Es gilt: $Y_i + Y_j \sim B(n, p_i + p_j)$, denn

$$\left(Y_i, Y_j, Z = \sum_{k \neq i,j} Y_k\right) \sim M(n, p_i, p_j, (1 - (p_i + p_j)))$$

und die Randverteilung von Z ist $Z \sim B(n, 1 - (p_i + p_j))$. Somit ist die Verteilung von $n - Z = Y_i + Y_j \sim B(n, p_i + p_j)$ und $\mathbb{V}(Y_i + Y_j) = n\,(p_i + p_j)\,(1 - (p_i + p_j))$.

Somit gilt (nach Cov auflösen):

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_j) &= \frac{\mathbb{V}(Y_i + Y_j) - n\,(p_i\,(1 - p_i) + p_j\,(1 - p_j))}{2} \\
&= \frac{n\,(p_i + p_j)\,(1 - (p_i + p_j)) - n\,(p_i\,(1 - p_i) + p_j\,(1 - p_j)))}{2} \\
&= \frac{n}{2}\,(p_i\,(1 - (p_i + p_j)) + p_j\,(1 - (p_i + p_j)) - p_i\,(1 - p_i) - p_j\,(1 - p_j)) \\
&= \frac{n}{2}\,(p_i\,(1 - (p_i + p_j) - 1 + p_i) + p_j\,(1 - (p_i + p_j) - 1 + p_j)) \\
&= \frac{n}{2}(-p_i\,p_j - p_j\,p_i) = -\frac{n}{2}\,2\,p_i\,p_j = -n\,p_i\,p_j
\end{aligned}$$

□

12.2 Die Normalverteilung

12.2.1 Univariate Normalverteilung

Bekannt: $X \sim \mathrm{N}(\mu, \sigma^2) \quad \Rightarrow \quad \mathbb{E}(X) = \mu, \quad \mathbb{V}(X) = \sigma^2$

Satz 12.9 (Lineare Transformation)**.**

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Y = aX + b \sim N(a\mu + b, a^2\sigma^2) \quad \textit{für } a, b \in \mathbb{R}, a \neq 0.$$

Beweis: $y = g(x) = ax + b \quad \Rightarrow g^{-1}(y) = \dfrac{y - b}{a}$ bijektiv

$$\frac{\partial g^{-1}(y)}{\partial y} = \frac{1}{a}$$

$$\begin{aligned}
\overset{\text{Trafo}}{\Rightarrow} f_Y(y) &= f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|} \\
&= \frac{1}{\sqrt{2\pi}\,\sigma\,|a|} \exp\left(-\frac{1}{2}\left(\frac{\frac{y-b}{a}-\mu}{\sigma}\right)^2\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp\left(-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\,\sigma}\right)^2\right)
\end{aligned}$$

$\Rightarrow Y \sim \mathrm{N}(a\mu + b, a^2\,\sigma^2)$ □

Korollar 12.3.

$$X \sim N(\mu, \sigma^2) \quad \Rightarrow \quad Y = \frac{X-\mu}{\sigma} \sim N(0,1).$$

Satz 12.10. $X \sim N(\mu, \sigma^2) \Rightarrow$
$m_r^0(X) = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (r-1)\,\sigma^r & r = 2k, k \in \mathbb{N} \\ 0 & \textit{sonst} \end{cases}$

Beweis: $r = 2k+1 \overset{\text{Symmetrie Satz 9.1}}{\Rightarrow} m_r^0(X) = 0$

Sei $U \sim \mathrm{N}(0,1)$. Dann gilt

$$\begin{aligned}
\frac{\partial \varphi(x)}{\partial x} &= -x\,\varphi(x) \quad \text{und} \\
\mathbb{E}(U^r) &= \int \underbrace{x^{r-1}}_{v} \cdot \underbrace{x \cdot \varphi(x)}_{u'}\,dx \\
&\overset{PI}{=} \left[-x^{r-1}\varphi(x)\right]_{-\infty}^{\infty} - \int -\varphi(x)(r-1)x^{r-2}\,dx \\
&= 0 + (r-1)\underbrace{\int x^{r-2}\,\varphi(x)\,dx}_{\mathbb{E}(U^{r-2})}.
\end{aligned}$$

$\Rightarrow \mathbb{E}(U^r) = (r-1)\mathbb{E}(U^{r-2})$

Wegen $\mathbb{E}(U) = 0$ folgt $\mathbb{E}(U^r) = 0 \quad \forall\, r$ ungerade und wegen $\mathbb{E}(U^2) = \mathbb{V}(U) =$

1 folgt für r gerade, $r = 2k$,

$$\begin{aligned}
\mathbb{E}(U^r) &= (r-1)\,\mathbb{E}(U^{r-2}) \\
&\overset{r=2k}{=} (2k-1)\,\mathbb{E}(U^{2\,(k-1)}) \\
&= (2k-1)\,(2k-3)\,\mathbb{E}(U^{2\,(k-2)})\ldots \\
&= (2k-1)\,(2k-3)\cdots(3)\;\underbrace{\mathbb{E}(U^2)}_{=1} \\
&= 1\cdot 3\cdot 5\cdots(2k-1).
\end{aligned}$$

Wegen $\mathbb{E}((X-\mu)^r) = \mathbb{E}((\sigma\,U)^r) = \sigma^r\,\mathbb{E}(U^r)$ folgt

$$m_r^0(X) = \mathbb{E}((X-\mu)^r) = 1\cdot 3\cdot 5\cdots(2k-1)\,\sigma^r.$$

□

Satz 12.11 (Additivität der Normalverteilung)**.** *Seien* $X_1 \sim N(\mu_1, \sigma_1^2)$ *und* $X_2 \sim N(\mu_2, \sigma_2^2)$ *stu*

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

Beweis:

$$X_1 + X_2 \sim \mathrm{N}(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2) \quad \Leftrightarrow \quad \frac{X_1+X_2-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} \sim \mathrm{N}(0,1)$$

Weiterhin:

$$\frac{X_1+X_2-(\mu_1+\mu_2)}{\sqrt{\sigma_1^2+\sigma_2^2}} = \underbrace{\frac{X_1-\mu_1}{\sqrt{\sigma_1^2+\sigma_2^2}}}_{\widetilde{X}_1} + \underbrace{\frac{X_2-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}}}_{\widetilde{X}_2} \sim \mathrm{N}(0,1)$$

$\mathbb{E}(\widetilde{X}_1) = \mathbb{E}(\widetilde{X}_2) = 0$

$$\mathbb{V}(\widetilde{X}_1) = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} := \lambda^2$$

$$\mathbb{V}(\widetilde{X}_2) = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} = 1-\lambda^2$$

Es reicht also zu zeigen:

$$\widetilde{X}_1 \sim \mathrm{N}(0,\lambda^2) \text{ und } \widetilde{X}_2 \sim \mathrm{N}(0,1-\lambda^2) \text{ stu} \quad \Rightarrow \quad \widetilde{X}_1 + \widetilde{X}_2 \sim \mathrm{N}(0,1)$$

$$
\begin{aligned}
f_{\widetilde{X}_1+\widetilde{X}_2}(z) &\overset{\text{Faltung}}{=} \int f_{\widetilde{X}_1}(z-x)\, f_{\widetilde{X}_2}(x)\, dx \\
&= \int \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left(-\frac{1}{2}\left(\frac{z-x}{\lambda}\right)^2\right) \frac{1}{\sqrt{2\pi(1-\lambda^2)}} \exp\left(-\frac{1}{2}\,\frac{x^2}{1-\lambda^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\, z^2\right) \cdot \\
&\quad \underbrace{\int \frac{1}{\sqrt{2\pi}}\, \frac{1}{\sqrt{\lambda^2\,(1-\lambda^2)}} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{(z-x)^2}{\lambda^2} + \frac{x^2}{1-\lambda^2} - z^2\right)}_{A}\right) dx}_{B}
\end{aligned}
$$

$$
\begin{aligned}
A &= \frac{(1-\lambda^2)\, z^2 - 2\,(1-\lambda^2) z\, x + (1-\lambda^2)\, x^2 + \lambda^2\, x^2 - \lambda^2\,(1-\lambda^2)\, z^2}{\lambda^2\,(1-\lambda^2)} \\
&= \frac{x^2 - 2\,(1-\lambda^2)\, z\, x + [(1-\lambda^2)\, z]^2}{\lambda^2\,(1-\lambda^2)} = \frac{(x-(1-\lambda^2)\, z)^2}{\lambda^2\,(1-\lambda^2)}
\end{aligned}
$$

$\Rightarrow B(x)$ ist Dichte von $\mathrm{N}((1-\lambda^2)\, z, \lambda^2\,(1-\lambda^2))$

$$\Rightarrow \int B(x)\, dx = 1 \qquad \square$$

Korollar 12.4. *$X_1, \ldots, X_n$ stu, $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1, \ldots, n$. Dann gilt*

i) $\sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$

ii) $\mu_1 = \ldots = \mu_n =: \mu, \sigma_1^2 = \ldots = \sigma_n^2 =: \sigma^2$
$\Rightarrow \; \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Korollar 12.5. *$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ stu. $\Rightarrow \quad aX_1 + bX_2 + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$*

Bemerkung: Es gilt auch die Umkehrung:
Sind X und Y stu und $X+Y$ normalverteilt, so sind auch X und Y normalverteilt.

12.2.2 k-dimensionale Normalverteilung

Definition 12.5 (k-dimensionale Normalverteilung)**.** *Sei $X : \Omega \to \mathbb{R}^k$. $X = (X_1, \ldots, X_k)$ heißt k-dimensional (multivariat) standardnormalverteilt, wenn X eine stetige Verteilung mit Dichte*

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^k \exp\left(-\frac{1}{2}\sum_{i=1}^n x_i^2\right) \qquad x = (x_1, \ldots, x_k) \in \mathbb{R}^k$$

besitzt: $X \sim N_k(0, I_k)$.

Satz 12.12.

$$X \sim N_k(0, I_k) \quad \Leftrightarrow \quad X_1, \ldots, X_k \text{ stu}, X_i \sim N(0,1)\ \forall i = 1, \ldots, k.$$

Beweis:

$$
\begin{aligned}
X_1, \ldots, X_k \text{ stu} \quad \overset{10.8}{\Leftrightarrow} \quad f_X(x) &= \prod_{i=1}^{k} f_{X_i}(x_i) \\
&\overset{X_i \sim \mathrm{N}(0,1)}{=} \prod_{i=1}^{k} \varphi(x_i) \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^k \exp\left(-\frac{1}{2}\sum x_i^2\right)
\end{aligned}
$$

□

Definition 12.6. *Sei* $X \sim N_k(0, I_k)$ *und* $A \in \mathbb{R}^{p \times k}$ *sowie* $\mu \in \mathbb{R}^p$. *Dann heißt*

$$Y = A\,X + \mu$$

p-dimensional normalverteilt:

$$Y \sim N_p(\mu, \Sigma) \quad \textit{mit } \Sigma = A\,A^\top.$$

Satz 12.13.

$$
\begin{aligned}
Y \sim N_p(\mu, \Sigma) \quad \Rightarrow \quad &\mathbb{E}(Y) = \mu \\
&\mathbb{V}(Y) = \Sigma
\end{aligned}
$$

Beweis: Satz 10.12 □

Satz 12.14. *Wenn* $A \in \mathbb{R}^{k \times k}$ *invertierbar, dann hat* $Y \sim N_k(\mu, \Sigma)$ *die Dichte* $f_Y : \mathbb{R}^k \to \mathbb{R}$

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{(y-\mu)^\top \Sigma^{-1} (y-\mu)}{2}\right)$$

mit $\Sigma = AA^\top$.

Beweis: Mit Satz 10.6 gilt für allgemeine $Y = A\,X + \mu$

$$f_Y(y) = f_X(A^{-1}(y-\mu)) \cdot |\det(A)|^{-1} \quad \text{vgl. Übung}$$

Also wegen

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2} x^\top x\right)$$

gilt

$$\begin{aligned} f_Y(y) &= f_X(A^{-1}(y-\mu)) \cdot |\det(A)|^{-1} \\ &= \frac{|\det(A)|^{-1}}{\sqrt{(2\pi)^k}} \exp\left(-\frac{1}{2}(y-\mu)^\top A^{-1\top} A^{-1}(y-\mu)\right) \end{aligned}$$

und wegen $A^{-1\top}A^{-1} = \left(AA^\top\right)^{-1} = \Sigma^{-1}$ und $0 \overset{\text{psd}}{\leq} \det(\Sigma) = \det\left(A^\top A\right) = \det(A)^2$ folgt

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{(y-\mu)^\top \Sigma^{-1}(y-\mu)}{2}\right).$$

□

Satz 12.15.

a) $Y \sim N_k(\mu, \Sigma) \quad \Rightarrow \quad Y_i \sim N(\mu_i, \Sigma_{ii}) \quad \forall i = 1, \ldots, k$

b) $Y \sim N_k(\mu, \Sigma) \quad \Leftrightarrow \quad v^\top Y \sim N_1(v^\top \mu, v^\top \Sigma v) \quad \forall v \in \mathbb{R}^k$

c) $(Y_1, Y_2) \sim N_2(\mu, \Sigma)$ *mit* $\mu = (\mu_1, \mu_2)$ *und* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$$Y_1, Y_2 \text{ stu} \quad \Leftrightarrow \quad \Sigma_{12} = \Sigma_{21} = 0$$

Beweis:

a), b) Meintrup and Schäffler [2005] Seite 197ff

c) Übung.

□

12.3 Weitere stetige Verteilungen

12.3.1 Gamma-Verteilung

Definition 12.7 (Γ-Funktion).

$$\begin{aligned} \Gamma : \mathbb{R}^+ &\to \mathbb{R}^+ \\ \Gamma(z) &= \int_0^\infty t^{z-1} e^{-t}\, dt \end{aligned}$$

Satz 12.16 (Eigenschaften der Γ-Funktion)**.**

- $\Gamma(z+1) = z\,\Gamma(z)$ *für* $z > 0$.
- $\Gamma(1) = 1$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(m+1) = m! \quad m \in \mathbb{N}$
- $\Gamma\left(\frac{k+1}{2}\right) = \frac{k-1}{2} \cdot \frac{k-3}{2} \cdot \ldots \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi}$ *falls* k *gerade.*

Definition 12.8. *Eine ZV* X *mit stetiger Dichte*

$$f_X(x;k,\lambda) = \frac{\lambda}{\Gamma(k)}\,(\lambda x)^{k-1}\,\exp(-\lambda x) \cdot I_{(0,\infty)}(x)$$

mit $k > 0$ *und* $\lambda > 0$ *heißt Gamma-verteilt:* $\quad X \sim Ga\,(k,\lambda)$.

Satz 12.17.

$$Ga\,(k=1,\lambda) = Exp(\lambda)$$

Beweis: klar. □

Satz 12.18. $X \sim Ga\,(k,\lambda) \Rightarrow \mathbb{E}(X) = \dfrac{k}{\lambda}, \quad \mathbb{V}(X) = \dfrac{k}{\lambda^2}$ *und* $M_X(s) = \left(\dfrac{\lambda}{\lambda - s}\right)^k$ *für* $s < \lambda$.

Beweis:

$$\begin{aligned}
M_X(s) &= \mathbb{E}(\exp(sX)) \\
&= \int_0^\infty \frac{\lambda}{\Gamma(k)}\,(\lambda x)^{k-1}\,\exp(-\lambda x)\,\exp(sx)\,dx \\
&= \int_0^\infty \frac{\lambda^k}{\Gamma(k)}\,x^{k-1}\,\exp((-\lambda+s)\,x)\,dx \\
&= \lambda^k \int \frac{1}{\Gamma(k)}\,x^{k-1}\,\exp(-(\lambda-s)\,x)\,dx \\
&\overset{\text{Kreative 1}}{=} \frac{\lambda^k}{(\lambda-s)^k} \int \underbrace{\frac{(\lambda-s)^k}{\Gamma(k)}\,x^{k-1}\,\exp(-(\lambda-s)\,x)}_{\text{Dichte Ga}(k,\lambda-s)\,\equiv\,1}\,dx \\
&= \frac{\lambda^k}{(\lambda-s)^k}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \left.\frac{\partial M_X(s)}{\partial s}\right|_{s=0} &= \left. k \cdot \lambda^k \cdot (\lambda - s)^{-k-1}\right|_{s=0} \\
&= \frac{k \cdot \lambda^k}{\lambda^{k+1}} = \frac{k}{\lambda} = \mathbb{E}(X) \\
\left.\frac{\partial^2 M_X(s)}{\partial^2 s}\right|_{s=0} &= \left. k\,(k+1)\ \lambda^k\,(\lambda - s)^{-k-2}\right|_{s=0} \\
&= \frac{k\,(k+1)}{\lambda^2} = \mathbb{E}(X^2)
\end{aligned}$$

Zusammen: $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{k\,(k+1)}{\lambda^2} - \frac{k^2}{\lambda^2} = \frac{k}{\lambda^2}$ □

Bemerkung:

$$\begin{aligned}
\int_0^\infty f_X(x)\,dx &= \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \int_0^\infty \frac{x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}} \exp\left(-\frac{1}{2}\,x\right)\,dx \\
&= \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \int_0^\infty \frac{1}{2}\left(\frac{x}{2}\right)^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right)\,dx \\
&\overset{t=\frac{x}{2},\,dx=dt\cdot 2}{=} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \underbrace{\int_0^\infty t^{\frac{k}{2}-1}\exp(-t)\,dt}_{\Gamma\left(\frac{k}{2}\right)} = 1
\end{aligned}$$

12.3.2 Chi-Quadrat-Verteilung

Definition 12.9. *Eine ZV X mit stetiger Dichte*

$$f_X(x;k) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} \exp\left(-\frac{1}{2}\,x\right) \cdot I_{(0,\infty)}(x)$$

heißt χ^2-verteilt mit k Freiheitsgraden: $X \sim \chi^2\,(k)$.

Bemerkung:

$$\mathrm{Ga}\left(\frac{k}{2}, \frac{1}{2}\right) = \chi^2\,(k)$$

Satz 12.19. $X \sim \chi^2\,(k) \Rightarrow \mathbb{E}(X) = k, \quad \mathbb{V}(X) = 2k, \quad M_X(s) = \left(\frac{1}{1-2s}\right)^{\frac{k}{2}} \quad s <$
$\frac{1}{2}$

Beweis: Satz 12.18 mit $k = \frac{k}{2}$ und $\lambda = \frac{1}{2}$. □

Satz 12.20. *$X_1, \ldots, X_n$ stu mit $X_i \sim N(0,1)$. Dann*

$$Y = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n).$$

Beweis:

$$\begin{aligned} M_Y(s) &= \mathbb{E}(\exp(sY)) \\ &= \mathbb{E}\left(\exp\left(s\sum_{i=1}^{n} X_i^2\right)\right) \\ &= \mathbb{E}\left(\prod_{i=1}^{n} \exp(s\,X_i^2)\right) \\ &\overset{X \text{ stu}}{=} \prod_{i=1}^{n} \mathbb{E}\left(\exp\left(s\,X_i^2\right)\right) \end{aligned}$$

Aber

$$\begin{aligned} \mathbb{E}\left(\exp(sX_i^2)\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \cdot \exp(sx^2)\,dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(\left(-\frac{1}{2}+s\right)x^2\right)dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(1-2s)\,x^2\right)dx \\ &\overset{\text{Kreative 1}}{=} \frac{1}{\sqrt{1-2s}} \underbrace{\int_{-\infty}^{\infty} \frac{\sqrt{1-2s}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(1-2s)\,x^2\right)dx}_{\equiv 1,\text{ da der Integrand die Dichte von N}(0,(1-2s)^{-1})\text{ ist.}} \\ &= \frac{1}{\sqrt{1-2s}} \end{aligned}$$

Zusammen:

$$\begin{aligned} M_Y(s) &= \prod_{i=1}^{n} \mathbb{E}(\exp(s\, X_i^2)) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{1-2s}} = \underbrace{\left(\frac{1}{1-2s}\right)^{\frac{n}{2}}}_{\substack{\text{Momentengenerierende Fkt.} \\ \text{von } \chi^2(n) \text{ nach Satz 12.19}}} \end{aligned}$$

□

12.3.3 Fisher-Snedecor-Verteilung

Satz 12.21. *Sei $U \sim \chi^2\,(m)$ und $V \sim \chi^2\,(n)$ stu. Dann hat die Zufallsvariable*

$$X = \frac{U/m}{V/n}$$

eine stetige Verteilung mit Dichte

$$f_X(x; m, n) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1+\left(\frac{m}{n}\right)\cdot x\right)^{\frac{m+n}{2}}} \cdot I_{(0,\infty)}(x).$$

Beweis: $f_{U,V}(u,v) \overset{\text{stu}}{=} \frac{1}{\Gamma\left(\frac{m}{2}\right)\,\Gamma\left(\frac{n}{2}\right)\cdot 2^{\frac{m+n}{2}}}\, u^{\frac{m-2}{2}}\, v^{\frac{n-2}{2}} \cdot \exp\left(-\frac{1}{2}(u+v)\right) \cdot$
$I_{(0,\infty)}(u) \cdot I_{(0,\infty)}(v)$

$$g(u,v) = \left(\underbrace{\frac{u/m}{v/n}}_{x}, \underbrace{v}_{y}\right)$$

$$h(x,y) = \left(x\,y\,\frac{m}{n}, y\right)$$

$$J = \frac{\partial h(x,y)}{\partial x \partial y} = \begin{pmatrix} y\,\frac{m}{n} & 0 \\ x\,\frac{m}{n} & 1 \end{pmatrix} \Rightarrow |J| = y\,\frac{m}{n}$$

$$\overset{\text{Trafo}}{\Rightarrow}$$

$$\begin{aligned}
f_{X,Y}(x,y) &= \frac{m}{n} \cdot y \cdot \frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{\frac{m+n}{2}}} \left(\frac{m}{n}\, x\, y\right)^{\frac{m-2}{2}} \cdot \\
&\quad \cdot y^{\frac{n-2}{2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{m}{n}\, x\, y + y\right)\right) \\
f_X(x) &= \int_0^\infty f_{X,Y}(x,y)\, dy \\
&= \underbrace{\frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right) 2^{\frac{m+n}{2}}} \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m-2}{2}}}_{=:A} \cdot \\
&\quad \cdot \int_0^\infty y^{\frac{m+n-2}{2}} \exp\left(-\frac{1}{2}\left(\frac{m}{n}\, x + 1\right) y\right) dy \\
&= A \cdot \int_0^\infty y^{(m+n-2)} \exp\left(-\frac{1}{2}\left(\frac{m}{n}\, x + 1\right)\right) dy \\
&\overset{\text{div. Umformungen}}{=} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1 + \frac{m}{n} \cdot x\right)^{\frac{m+n}{2}}} \cdot I_{(0,\infty)}(x)
\end{aligned}$$

□

Definition 12.10 (F-Verteilung)**.** *Eine ZV X mit Dichte f_X wie in Satz 12.21 heißt F-verteilt mit Freiheitsgraden m und n:* $X \sim F(m,n)$.

Satz 12.22.

$$\begin{aligned}
X \sim F(m,n) \Rightarrow \mathbb{E}(X) &= \frac{n}{n-2} \text{ für } n > 2 \text{ und} \\
\mathbb{V}(X) &= \frac{2\, n^2\, (m+n-2)}{m\,(n-2)^2\,(n-4)} \text{ für } n > 4.
\end{aligned}$$

Beweis: $\mathbb{E}(X) = \mathbb{E}\left(\frac{U/m}{V/n}\right) \overset{\text{stu}}{=} \left(\frac{m}{n}\right)^{-1} \mathbb{E}(U) \cdot \mathbb{E}\left(\frac{1}{V}\right) = \left(\frac{m}{n}\right)^{-1} \cdot m \cdot \underbrace{\mathbb{E}\left(\frac{1}{V}\right)}_{=\frac{1}{n-2} \text{ o.B.}} = \frac{n}{n-2}$ □

Bemerkung:

$$X \sim \mathrm{F}\,(m,n) \quad \Rightarrow \quad \frac{1}{X} \sim \mathrm{F}\,(n,m)$$

12.3.4 Student-t-Verteilung

Satz 12.23. *Sei $Z \sim N(0,1)$ und $U \sim \chi^2(k)$ stu. Dann hat*

$$X = \frac{Z}{\sqrt{U/k}}$$

eine Verteilung mit Dichte

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1+\frac{x^2}{k}\right)^{\frac{k+1}{2}}}.$$

Beweis: $f_{Z,U}(z,u) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}} u^{\frac{k}{2}-1} \exp\left(-\frac{1}{2}u\right) \exp\left(-\frac{1}{2}z^2\right) I_{(0,\infty)}(u)$

$g(z,u) = (x,y) = \left(\frac{z}{\sqrt{u/k}}, u\right)$

$\Rightarrow |J| = \sqrt{y/k}$ und

$\int\limits_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = f_X(x)$ wie oben angegeben. □

Definition 12.11 (t-Verteilung)**.** *Eine ZV X mit Dichte f_X wie in Satz 12.23 heißt t-verteilt mit k Freiheitsgraden: $X \sim t(k)$.*

Satz 12.24. $X \sim t(k) \Rightarrow \mathbb{E}(X) = 0, \quad \mathbb{V}(X) = \frac{k}{k-2}.$

Beweis:

a) Symmetrie in x

b) geschenkt.

□

Literaturverzeichnis

E. Lehmann. *Elements of Large-Sample Theory.* Springer, Berlin, 2001.

D. Meintrup and S. Schäffler. *Stochastik: Theorie und Anwendungen (Statistik und ihre Anwendungen).* Springer, Heidelberg, 2005.

K. Schmidt. *Maß und Wahrscheinlichkeit.* Springer, 2011.

Lizenz

Basierend auf einem Skript von Prof. Dr. Torsten Hothorn (2007–2011) unter GFDL 1.2. LaTeX-Satz von Esther Herberich. Korrekturen von Dipl.-Math. Michael Kobl, MSc. Weitere Korrekturen von Studierenden des Bachelor Statistik seit 2007.

Das ursprüngliche Skript basiert auf Meintrup and Schäffler (2005).