Wahrscheinlichkeitstheorie
und Inferenz I

Prof. Dr. Volker Schmid

Wintersemester 2020/21

Institut für Statistik
Ludwig–Maximilians–Universität München

Stand: 26. Januar 2021

Inhaltsverzeichnis

0 Einführung **4**

I Maßtheorie **6**

1 Mengensysteme **6**
1.1 Mengen von Ereignissen 6
1.2 sigma-Algebra 8

2 Wahrscheinlichkeit als Maß **12**
2.1 Mathematisches Maß 12
2.2 Wahrscheinlichkeit als normiertes Maß 16

3 Abbildungen und Zufallsvariablen **19**
3.1 Bild- und Urbild 19
3.2 Meßbare Abbildungen 21
3.3 Zufallsvariablen 22

II Verteilungen und ihre Eigenschaften **27**

4 Verteilungsfunktion **27**

5 Lebesgue-Integral **31**
5.1 Definition 31
5.2 Eigenschaften des Integrals 35
5.3 Konvergenz des Lebesgue-Integrals 37
5.4 Riemann- und Lebesgue-Integral 39

6 Dichte **40**

7 Arten von Verteilungen **43**
7.1 Diskrete Verteilungen 43
7.2 Stetige Verteilungen 46
7.3 Gemischte Verteilungen 52
7.4 Mehrdimensionale Verteilungen 54

8 Momente **56**
8.1 Erwartungswert und Varianz 56
8.2 Allgemeine Momente 59
8.3 Ungleichungen 61

8.4 Weitere Lagemaße 65

9 Erzeugende Funktionen **67**
9.1 Momenterzeugende Funktion 67
9.2 Charakteristische Funktion 70

III Mehrdimensionale Zufallsvariablen **72**

10 Unabhängigkeit von Zufallsvariablen **72**

11 Transformationssatz für Dichten **77**

12 Mehrdimensionale Momente **81**

13 Die Normalverteilung **86**
13.1 Eindimensionale Normalverteilung 86
13.2 k-dimensionale Normalverteilung 90

14 Copulas **93**

15 Bedingte Verteilung **96**

IV Konvergenz **106**

16 Konvergenzarten **106**

17 Gesetz der großen Zahlen **111**

18 Zentraler Grenzwertsatz **113**
18.1 Konvergenz in Verteilung 113
18.2 Zentraler Grenzwertsatz 116
18.3 Konvergenzrate 119
18.4 Folgerungen 121

V Anhang **125**

A Spezielle diskrete Verteilungen **125**
A.1 Diskrete Gleichverteilung 125
A.2 Bernoulli- und Binomialverteilung 125
A.3 Poisson-Verteilung 126

A.4 Hypergeometrische Verteilung . . . 126
A.5 Geometrische und Negativ-Binomial-Verteilung . . . 128
A.6 Multinomialverteilung . . . 130

B Spezielle stetige Verteilungen **132**
B.1 Stetige Gleichverteilung . . . 132
B.2 Exponential- und Gammaverteilung . . . 133
B.3 Chi-Quadrat-Verteilung . . . 137
B.4 Beta-Verteilung . . . 140
B.5 Student-t-Verteilung . . . 141

Literaturverzeichnis **146**

Lizenz **147**

0 Einführung

Wahrscheinlichkeit

Beispiel 0.1 (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 0.2 (n-facher Münzwurf)
Wir werfen die Münze n*-mal. Ergebnisse sind* n*-Tupel der Form* $\omega = (1, 0, 0, \ldots, 1)$*. Die Ergebnismenge* Ω_n *besteht aus* 2^n *Elementen. Wir interessieren uns für die Wahrscheinlichkeit von Ereignissen, z.B.* $A := \{\text{genau 2 Würfe sind 0}\}$*.* A *enthält mehrere Ergebnisse, ist also eine Teilmenge von* Ω*. Was ist* $P(A)$*?*

Beispiel 0.3 (Einfacher Münzwurf II)
Wir werfen eine Münze auf eine Gerade der Länge 1. Die (Mitte der) Münze bleibt irgendwo zwischen 0 und 1 liegen. Mögliche Ergebnisse sind alle reellen 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.* $B := \{\text{Münze liegt links von 0.4}\}$*. Was ist* $P(B)$*?*

Was ist Wahrscheinlichkeit?

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).

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: $P(A_1 \cup A_2 \cup \cdots) = \sum P(A_i)$ (σ-Additivität).

Was passiert, wenn wir unendlich viele Ereignisse haben?
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.

Beispiel 0.4 (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 Radius, darauf einen zufälligen Punkt, lege Sehne orthogonal zum Radius durch den Punkt* $\to P(Y > X) = 1/2$
- *wähle einen zufälligen Punkt im Kreis. Dieser bildet den Mittelpunkt der Sehne* $\to P(Y > X) = 1/4$.

Teil I
Maßtheorie

1 Mengensysteme

1.1 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.

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

Bemerkung:

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

Definition 1.1 (Mengenoperationen)
Seien $A, B, A_i \subset \Omega, i \in I$.

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

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 $B \subseteq A \implies A = B$
Transivität:	$A \subseteq B$ und $B \subseteq C \implies A \subseteq C$
	$\implies \mathcal{P}$ 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:	$\lvert A\rvert = \lvert B\rvert :\iff \exists f : A \to B$ bijektiv
Addition von Mächtigkeiten:	$A \cap B = \emptyset \iff \lvert A\rvert + \lvert B\rvert = \lvert A \cup B\rvert$
Multiplikation von Mächtigkeiten:	$\lvert A \times B\rvert = \lvert A\rvert \cdot \lvert B\rvert$

1.2 σ-Algebra

Ziel: Wir wollen eine allgemeine Definition des Begriffs Wahrscheinlichkeit. Weg: Wahrscheinlichkeit 'messe' ein Ereignis. Wir definieren erstmal Räume von Ereignissen, die wir dann messen können.

Definition 1.2 (σ-Algebra)
Ein Mengensystem $\mathcal{F} \subset \mathcal{P}$ heißt <u>σ-Algebra über Ω</u> oder abgeschlossenes Mengensystem, falls gilt:

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

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

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

Bemerkungen:

- Die σ-Algebra definiert uns die Menge aller möglichen Ereignisse, denen wir Wahrscheinlichkeiten zuweisen
- (S3) lässt sich auch über Schnitte definieren. Benutze dazu (S2): $\bigcup_{i=1}^{\infty} A_i = \overline{\bigcap_{i=1}^{\infty} \bar{A}_i}$

Beispiel 1.1
$1\times$ *Würfeln* $\quad \Omega_1 = \{1,\dots,6\}$
$2\times$ *Würfeln* $\quad \Omega_2 = \{(1,1),(1,2),\dots,(6,6)\}$
$\vdots$
$n\times$ *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 $A = \{(1,1),(1,2),\dots,(1,6)\} = \{\text{"1 im ersten Wurf"}\} \in \mathcal{F}_2$

Beispiel 1.2 (kleinste und größte σ-Algebra)
Kleinste σ-Algebra: $\{\emptyset,\Omega\}$ Größte σ-Algebra: $\mathcal{P}$

Beispiel 1.3
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 1.3 (π-System)
Ein Mengensystem $\mathcal{F} \subset \mathcal{P}$ heißt durchschnittsstabil, wenn gilt:

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

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

Definition 1.4 (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 1.1 (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. 1.2:

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

(S1) analog

(S3) analog

$\square$

Definition 1.5 (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 1.4 (Erzeugte 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 $\Omega = \{1, 2, \ldots, 7\}$ 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 1.6 (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 1.2 (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 $\overset{(S1)}{\Longrightarrow} \quad \Omega = \mathbb{R} \in \mathcal{B}$
$\overset{(S2)}{\Longrightarrow} \quad \bar{\Omega} = \emptyset \in \mathcal{B}$

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

Bemerkungen
Definition von $\mathcal{B}$ ist auch über geschlossene oder halboffene Intervalle möglich. Wir verwenden im Folgenden diese Definitionen analog.

Beispiel 1.5 (Boule/Boccia)
1 *Wurf* $\Omega_1 = \mathbb{R}^2$
2 *Würfe* $\Omega_2 = \mathbb{R}^2 \times \mathbb{R}^2$
$\vdots$
n *Würfe* $\Omega_n = \prod_{i=1}^{n} \mathbb{R}^2$
$\mathcal{F}_n = (\mathcal{B}^2)^n$ *Borelsche σ-Algebra*
Interessierendes Ereignis: $A = \{(x, y)|x^2 + y^2 \leq r^2\} = \{$1. *Wurf höchstens* r *cm von* $(0, 0)$ *entfernt*$\} \subset \Omega_1$
$A \in \mathcal{B}^2$.

In der Wahrscheinlichkeitstheorie brauchen wir in der Regel

- für abzählbare Mengen (z.B. $\Omega = \mathbb{N}$) die Potenzmenge
- für überabzählbare Mengen (z.B. $\Omega = \mathbb{R}$) die Borelsche σ-Algebra

als σ-Algebra. Für überabzählbare Mengen wäre die Potenzmenge zu groß (enthält z.B. Vitali-Mengen).

2 Wahrscheinlichkeit als Maß

2.1 Mathematisches Maß

Definition 2.1 (meßbar)
Sei $\mathcal{F}$ eine σ-Algebra über Ω und $A \in \mathcal{F}$. Dann heißt A <u>meßbar</u> bezüglich $\mathcal{F}$.

Definition 2.2 (Meßraum)
Sei $\mathcal{F}$ eine σ-Algebra über Ω. Das Tupel $(\Omega, \mathcal{F})$ heißt <u>Meßraum</u>.

Definition 2.3 ((σ-endliches, endliches, normiertes) Maß)
Sei $(\Omega, \mathcal{F})$ ein Meßraum. Eine Funktion $\mu : \mathcal{F} \to \overline{\mathbb{R}}$ heißt <u>Maß</u> 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*

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

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

Notation:
$\bar{\mathbb{R}} = \mathbb{R} \cup [-\infty, \infty]$ bezeichnet die <u>erweiterten reellen Zahlen</u>.
$\dot{\bigcup}$ bezeichnet die Vereinigung *disjunkter* Mengen.

Beispiel 2.1 (Zählmaß)

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

$\Omega = \mathbb{N}$, $A = \{1, 2, 3\}$, $\mu_Z(A) = 3$

Beispiel 2.2 (Lebesgue-Maß)

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

$\Omega = \mathbb{R}$, $A = [2, 4]$, $\lambda(A) = 4 - 2 = 2$; $B = 3$, $\lambda(B) = 0$;

Definition 2.4 (Indikatorfunktion)
Sei $A \subset \Omega$*. Dann versteht man unter der Indikatorfunktion* I_A *von* A *die Abbildung* $I_A : \Omega \to \mathbb{R}$*,*

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

Satz 2.1
Seien A *und* B *zwei Teilmengen von* Ω*.*

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

b) Für den Schnitt 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 $I_{B \setminus A} = I_B - I_A$.

Beispiel 2.3 (Dirac-Maß)

$$\delta_\omega : \mathcal{F} \to \mathbb{R}, \quad \delta_\omega(A) := I_A(\omega) = \begin{cases} 1 \text{ falls } \omega \in A \\ 0 \text{ sonst} \end{cases}$$

$\Omega = \mathbb{R}$, $A = [2,4]$, $\delta_3(A) = 1$; $\delta_1(A) = 0$

Beispiel 2.4
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.

Definition 2.5 (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 2.5
Sei Ω überabzählbar und $A \in \mathcal{F}$ genau dann, wenn A oder $\bar{A}$ abzählbar ist $\Longrightarrow$ $\mathcal{F}$ ist σ-Algebra (Beispiel 1.3)

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

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 2.2 (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: $A \cap B = \emptyset \Longrightarrow \mu(A \cup B) = \mu(A) + \mu(B)$

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

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

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) $B = A \dot{\cup} (B \setminus A) \overset{\text{(M3)}}{\Rightarrow}$

$$\mu(B) = \underbrace{\mu(A)}_{\geq 0} + \underbrace{\mu(B \setminus A)}_{\geq 0 \text{ nach (M2)}} \geq \mu(A) \iff \text{iii)}$$

und

$$\mu(B \setminus A) = \mu(B) - \mu(A) \iff \text{ii) falls } \mu(A) < \infty$$

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

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

□

Im Folgenden:
Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge von Mengen.
$A_n \uparrow A :\iff A_1 \subset A_2 \subset A_3 \subset \ldots$ und $\bigcup\limits_{i=1}^{\infty} A_i = A$
$A_n \downarrow A :\iff A_1 \supset A_2 \supset A_3 \supset \ldots$ und $\bigcap\limits_{i=1}^{\infty} A_i = A$

Satz 2.3 (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: $A_n \uparrow A \implies \mu(A_n) \uparrow \mu(A)$

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

Beweis:

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

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

ii) $A \subset A_n \subset A_1 \overset{\text{Monotonie}}{\Longrightarrow} \begin{array}{l} \mu(A) < \infty \\ \mu(A_n) < \infty \quad \forall n \in \mathbb{N} \end{array}$

Wegen $A_n \downarrow A$ gilt $A_1 \setminus A_n \uparrow A_1 \setminus A$ und somit auch $\mu(A_1 \setminus A_n) \uparrow \mu(A_1 \setminus A)$. Dann:

$$\begin{aligned} \mu(A_1) - \mu(A) &= \mu(A_1 \setminus A) = \lim_{n\to\infty} \mu(A_1 \setminus A_n) = \\ \lim_{n\to\infty} (\mu(A_1) - \mu(A_n)) &= \mu(A_1) - \lim_{n\to\infty} (\mu(A_n)) \\ &\Longrightarrow \mu(A_n) \downarrow \mu(A) \end{aligned}$$

□

2.2 Wahrscheinlichkeit als normiertes Maß

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

Definition 2.6 (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 2.7 (Wahrscheinlichkeitsraum, Verteilung)
$(\Omega, \mathcal{F}, \mathbb{P})$ heißt Wahrscheinlichkeitsraum.
$\mathbb{P}$ heißt auch Verteilung.

Satz 2.4 (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) $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

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: $A_n \uparrow A \implies \mathbb{P}(A_n) \uparrow \mathbb{P}(A)$

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

Beweis:

i) ii) iv) $\Longleftarrow$ Satz 2.2 v) vi) $\Longleftarrow$ Satz 2.3 iii) Übung. $\square$

Beispiel 2.6
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 $\mathcal{F} = \sigma(A) = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}, \Omega\}$. Wir folgen dem ***Prinzip vom unzureichenden Grund*** *und nehmen an, dass alle $\omega \in \Omega$ 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 2.8 (Nullmenge)
Ist $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $A \in \mathcal{F}$ mit $\mu(A) = 0$, so heißt A (μ-)Nullmenge.

Definition 2.9 (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 2.10 ($\mathbb{P}$-fast sicher)

$$\mathbb{P}\text{-}\textit{fast überall} \iff \mathbb{P}\text{-}\textit{fast sicher}$$

Beispiel 2.7
Situation von Bsp. 0.3 (Münzwurf auf eine Gerade). Das Ereignis $A = \{\textit{Münze liegt genau auf } 0.4\}$ ist eine Nullmenge;

$$P(A) = 0.$$

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

Definition 2.11 (Bedingte Wahrscheinlichkeit)
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'keitsverteilung bzw. <u>bedingte Wahrscheinlichkeit</u> von A gegeben B.

Satz 2.5
$\mathbb{P}(\cdot \,|\, B)$ *ist W'keitsmaß auf* $(\Omega, \mathcal{F})$.

Beweis:

M1) $\checkmark$
M2) $\mathbb{P}(\cdot \,|\, B) \geq 0 \quad \checkmark$
M3) Sei $A_i, i \in \mathbb{N} \in \mathcal{F}$ disjunkt, zu zeigen: $\mathbb{P}\left(\dot{\bigcup}_{i=1}^{\infty} A_i \middle| B\right) \overset{!}{=} \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$

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

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

Definition 2.12 (Unabhängigkeit von Ereignissen)
Ereignisse $A_i \in \mathcal{F}, i \in I \neq \emptyset$, eines Wahrscheinlichkeitsraumes $(\Omega, \mathcal{F}, \mathbb{P})$ heißen <u>stochastisch unabhängig</u> (stu), wenn für jede endliche Teilmenge $\emptyset \neq J \subset I$ gilt

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

Bemerkung:

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

Satz 2.6

a) A, B *stu*, $\mathbb{P}(B) > 0 \implies \mathbb{P}(A \mid B) = \mathbb{P}(A)$

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

Beweis:

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

b)

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

□

3 Abbildungen und Zufallsvariablen

3.1 Bild- und Urbild

Ziel: Wir wollen Verteilungen auf Zufallsvariablen definieren.
Weg: Wir definieren Zufallsvariablen als Abbildung von einem Wahrscheinlichkeitsraum (mit Maß, also Verteilung) auf den (meßbaren) Wertebereich der Zufallsvariable.
Wir definieren im Folgenden allgemein Abbildungen $f : \Omega_1 \to \Omega_2$ über $f(\omega_1) = \omega_2$, wobei $\omega_1 \in \Omega_1$ und $\omega_2 \in \Omega_2$.

Definition 3.1
Sei $f : \Omega_1 \to \Omega_2$ *eine Abbildung, so ist das* *Bild* *einer Menge* $A \subset \Omega_1$

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

.

Definition 3.2
Sei $f : \Omega_1 \to \Omega_2$ eine Abbildung, so ist das <u>Urbild</u> einer Menge $B \subset \Omega_2$

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

Dabei muss f nicht bijektiv sein, aber die <u>Operationstreue</u> erfüllen: Für beliebige $B, B_i \in \Omega_2, i \in I$:

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

$$f^{-1}\left(\bigcap_{i \in I} B_i\right) = \bigcap_{i \in I} f^{-1}(B_i)$$

$$\begin{aligned} f^{-1}(\bar{B}) &= \{\omega \in \Omega_1 | f(\omega) \in \bar{B}\} \\ &= \Omega_1 \setminus \{\omega \in \Omega_1 | f(\omega) \in B\} \\ &= \Omega_1 \setminus f^{-1}(B) = \overline{f^{-1}(B)} \end{aligned}$$

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

Definition 3.3
Sei $f : \Omega_1 \to \Omega_2$ und $\mathcal{F} \subset \mathcal{P}(\Omega_2)$ ein Mengensystem. Dann ist

$$f^{-1}(\mathcal{F}) = \{f^{-1}(B) | B \in \mathcal{F}\} \subset \mathcal{P}(\Omega_1)$$

Satz 3.1
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.

$$\text{(S2) } A \in f^{-1}(\mathcal{F}_2) \implies A = f^{-1}(B), B \in \mathcal{F}_2$$
$$\implies \bar{B} \in \mathcal{F}_2$$
$$\implies \bar{A} = \overline{f^{-1}(B)} = f^{-1}(\bar{B}) \in f^{-1}(\mathcal{F}_2)$$

□

3.2 Meßbare Abbildungen

Definition 3.4 (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

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

Man spricht abgekürzt von einer "meßbaren" Abbildung f falls die involvierten σ-Algebren eindeutig bekannt sind.

Satz 3.2
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:

"$\Longrightarrow$" f $\mathcal{F}_1$-$\mathcal{F}_2$-meßbar $\implies f^{-1}(\mathcal{F}_2) = \{f^{-1}(B) | B \in \mathcal{F}_2\} \subset \mathcal{F}_1$
$\underset{\mathcal{E} \subset \sigma(\mathcal{E}) = \mathcal{F}_2}{\Longrightarrow} f^{-1}(\mathcal{E}) = \{f^{-1}(B) | B \in \mathcal{E}\} \subset \mathcal{F}_1$

"$\Longleftarrow$" $f(\mathcal{F}_1) = \{B \subset \Omega_2 | f^{-1}(B) \in \mathcal{F}_1\}$ ist σ-Algebra über Ω_2 (Satz 3.1). $f^{-1}(\mathcal{E}) \subset \mathcal{F}_1 \implies \mathcal{E} = f(f^{-1}(\mathcal{E})) \subset f(\mathcal{F}_1)$. $f(\mathcal{F}_1)$ ist also eine σ-Algebra die $\mathcal{E}$ enthält. Damit auch $\sigma(\mathcal{E}) \subset f(\mathcal{F}_1)$, denn $\sigma(\mathcal{E})$ ist der Schnitt aller σ-Algebren, die $\mathcal{E}$ enthalten.

□

Beispiel 3.2
Sei $(\Omega, \mathcal{F})$ Meßraum, dann ist $f : \Omega \to \mathbb{R}$ genau dann meßbar, wenn

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

denn $]-\infty, c]$ bildet ein Erzeugendensystem von $\mathcal{B}$ (folgt aus Satz 1.2).

Beispiel 3.3 (stetige Abbildungen)

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

$\mathcal{O}$ ist Erzeuger von $\mathcal{B} = \sigma(\mathcal{O})$ (Def. 1.6) und somit ist f $\mathcal{B}$-$\mathcal{B}$-meßbar.

Beispiel 3.4 (Indikatorfunktion)
Sei $(\Omega, \mathcal{F})$ Meßraum und $A \subset \Omega$.

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

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. 2.1)

Bemerkung:

Eine Funktion $f : \Omega \to \mathbb{R}$ heißt reellwertig.
Eine Funktion $f : \Omega \to \bar{\mathbb{R}}$ heißt numerisch.
Ist $(\Omega, \mathcal{F})$ ein Meßraum, so heißt f $\mathcal{F}$-$\bar{\mathcal{B}}$-meßbar, wenn

$$\{\omega \in \Omega | f(\omega) \leq c\} \in \mathcal{F} \quad \forall c \in \mathbb{R} \text{ bei } \bar{\mathcal{B}} = \sigma(\mathcal{B} \cup \{\infty\} \cup \{-\infty\})$$

Schreibweise:

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

3.3 Zufallsvariablen

Definition 3.5 (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

$$X : \Omega_1 \to \Omega_2$$

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.

Definition 3.6 (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 <u>*Bildmaß*</u> μ_f *von* μ *unter* f *und* $(\Omega_2, \mathcal{F}_2, \mu_f)$ *ist ein Maßraum.*

Beispiel 3.5
$(\mathbb{R}, \mathcal{B}, \lambda)$ *und* $(\mathbb{R}, \mathcal{B})$ $\quad f : \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + a \quad \lambda_f(A) = \lambda(f^{-1}(A)) = \lambda(A - a) = \lambda(A)\ \forall A \in \mathcal{B}$ *und* λ_f *ist Maß auf* $(\mathbb{R}, \mathcal{B})$.

Definition 3.7 (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 <u>*Verteilung von* X</u>.

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.

Beispiel 3.6 (Binomialverteilung)
Wir betrachten die n*-fache Wiederholung eines Zufallsexperiment mit Ereignissen* $\Omega_0 = \{0, 1\}$. *Sei* p *die Wahrscheinlichkeit für das Eintreten der* 1.
Dann ist $\Omega = \prod_{i=1}^{n} \Omega_0$. *Für* k *Anzahl der* 1 *in* ω *gilt* $\mathbb{P}(\{\omega\}) = p^k \cdot (1-p)^{n-k}$
Sei X *die messbare Abbildung (Zufallsvariable)* $X(\omega) = \sum_{i=1}^{n} \omega_i$ *mit* $\omega = (\omega_1, \ldots, \omega_n)$ *(Anzahl der Erfolge).*
Uns interessiert die Verteilung von X. *Es gilt*

$$\begin{aligned} \mathbb{P}_X(\{k\}) &= \mathbb{P}(X^{-1}(\{k\})) = \sum_{\omega \in X^{-1}(\{k\})} \mathbb{P}(\{\omega\}) \\ &= \sum_{\omega \in X^{-1}(\{k\})} p^k \cdot (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k} \end{aligned}$$

Definition 3.8
Die Verteilung $\mathbb{P}_X$ *in obigem Beispiel nennen wir* <u>*Binomialverteilung*</u> $B(n, p)$.

Bemerkung:

Schreibweise:

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

Satz 3.3 (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 $g^{-1}(\mathcal{F}_3) \subset \mathcal{F}_2 \implies$
$f^{-1}(\underbrace{g^{-1}(\mathcal{F}_3)}_{\subset \mathcal{F}_2}) \subset \mathcal{F}_1 \implies (g \circ f)^{-1}(\mathcal{F}_3) \subset \mathcal{F}_1$ □

Bemerkung:

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

Satz 3.4 (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 3.3) und somit Zufallsvariable. □

Satz 3.5
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:

$$\begin{array}{rcll} h : \Omega & \to & \mathbb{R}^2 & \\ \omega & \mapsto & h(\omega) = (f(\omega), g(\omega)) & \text{meßbar (siehe letzte Bemerkung)} \\ \ell : \mathbb{R}^2 & \to & \mathbb{R} & \\ (x, y) & \mapsto & \ell(x, y) & \end{array}$$

a) $\ell(x, y) = \alpha x + \beta y$ ist stetig und somit meßbar, also ist $\ell \circ h = \alpha f + \beta g$ mit Satz 3.3 meßbar.

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

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

□

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 3.5

Beispiel 3.7
Sei X *eine Zufallsvariable mit* $\mathbb{P}_X = B(n, p)$ *(wie in Beispiel 3.6). Dann ist auch*

$$X^2 : \Omega \to \mathbb{R}$$
$$X^2(\omega) = (X(\omega))^2$$

eine Zufallsvariable

Bemerkung:

Schreibweise:

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

Beispiel 3.8

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

Teil II

Verteilungen und ihre Eigenschaften

4 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) $a \le b \implies F_{\mathbb{P}}(b) - F_{\mathbb{P}}(a) = \mathbb{P}(]a,b]) \ge 0$

ii) $x_n \downarrow x \implies$

$$\lim_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz 2.3}}{=} \mathbb{P}(]-\infty, x]) = F_{\mathbb{P}}(x)$$

iii) $x_n \downarrow -\infty \implies$

$$\lim_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz 2.3}}{=} \mathbb{P}(\emptyset) = 0$$

$x_n \uparrow \infty \implies$

$$\lim_{n\to\infty} F_{\mathbb{P}}(x_n) = \lim_{n\to\infty} \mathbb{P}(]-\infty, x_n]) \overset{\text{Satz 2.3}}{=} \mathbb{P}(]-\infty, \infty[) = \mathbb{P}(\mathbb{R}) = 1$$

$\square$

Definition 4.2 (Verteilungsfunktion)
Jede monoton wachsende, rechtsstetige Funktion

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

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

heißt Verteilungsfunktion.

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] □

Bemerkung:

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

$$\lambda(\{x\}) \overset{\text{Satz 2.3}}{=} \lim_{n \to \infty} \lambda\left(\left]x - \frac{1}{n}, x\right]\right) = \lim_{n \to \infty} x - \left(x - \frac{1}{n}\right) = \lim_{n \to \infty} \frac{1}{n} = 0.$$

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

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

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: $\mu = \nu$.

Beweis:

siehe Meintrup and Schäffler [2005], Theorem 1.38. □

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:

$$\lambda_F(\mathbb{R}) = \lim_{n\to\infty} \lambda_F(]-\infty, n]) = \lim_{n\to\infty} F(n) = 1$$

$$\implies \lambda_F \text{ ist Wahrscheinlichkeitsmaß.}$$

Bemerkung:

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

Beweis:

„$\Longrightarrow$": F Verteilungsfunktion, $\mu = \lambda_F$

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

„$\Longleftarrow$": $\mathbb{P}$ W'keitsmaß, $G = F_\mathbb{P}$

$$\begin{aligned}\lambda_G(]a,b]) &= G(b) - G(a)\\ &= \mathbb{P}(]-\infty,b]) - \mathbb{P}(]-\infty,a])\\ &= \mathbb{P}(]a,b]) \quad \forall \underbrace{]a,b], a<b \in \mathbb{R}}_{\pi\text{-System Erzeuger von } \mathcal{B}}\end{aligned}$$

$$\Longrightarrow \mathbb{P} = \lambda_G \text{ auf durchschnittsstabilem Erzeuger von } \mathcal{B} \overset{\text{Satz 4.4}}{\Longrightarrow} \mathbb{P} = \lambda_G$$

$\square$

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

Beweis:

$$\left]x - \frac{1}{n}, x\right] \downarrow \{x\} \overset{2.3}{\Longrightarrow} \mathbb{P}(\{x\}) = \lim_{n\to\infty} \mathbb{P}\left(\left]x-\frac{1}{n},x\right]\right)$$
$$= \lim_{n\to\infty}\left(F_\mathbb{P}(x) - F_\mathbb{P}\left(x - \frac{1}{n}\right)\right)$$
$$= F_\mathbb{P}(x) - F_\mathbb{P}(x^-)$$

$\square$

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

5 Lebesgue-Integral

Ziel: Wir brauchen eine andere Art des Integrals für beliebige Messräume. Damit lassen sich Dichte, Erwartungswert, etc. allgemein definieren.

5.1 Definition

Sei im Folgenden

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

Definition 5.1 (Lebesgue-Integral für Indikatorfunktionen)
Sei A eine meßbare Menge bezüglich $(\Omega, \mathcal{F})$.
Das Lebesgue-Integral für die Indikatorfunktion I_A ist definiert als

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

Beispiel 5.1
Sei $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}, \lambda)$, $A =]1, 2[$ und $f = I_A$.
Es gilt:

$$\int f d\lambda = \lambda(A) = 2 - 1 = 1.$$

Definition 5.2 (Treppenfunktion)
Sei $(\Omega, \mathcal{F}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{R}$ eine $\mathcal{F}$-$\mathcal{B}$-meßbare Funktion mit endlichem Bild $f(\Omega) = \{y_1, \ldots, y_n\}$.
Dann gibt es $A_i \in \mathcal{F}, i = 1, \ldots, n$ so dass sich f darstellen lässt als

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

f heißt dann 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 5.3 (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 5.2
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 Einser in ω'.
Darstellung als Treppenfunktion:

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

Es gilt

$$\int X d\mathbb{P} = 0 \cdot \mathbb{P}(A_1) + 1 \cdot \mathbb{P}(A_2) + 2 \cdot \mathbb{P}(A_3) = 1$$

Satz 5.1 (Eigenschaften des Integrals)
Es gilt:

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) Linearität:

$$\begin{aligned}
\int (\alpha f + \beta g)\, d\mu &= \int \alpha \sum_{i=1}^{n} y_i I_{A_i} + \beta \sum_{i=1}^{n} z_i I_{A_i} d\mu \\
&= \int \sum_{i=1}^{n} (\alpha y_i + \beta z_i) I_{A_i} d\mu \overset{\text{Def. 5.3}}{=} \sum_{i=1}^{n} (\alpha y_i + \beta z_i)\mu(A_i) \\
&= \alpha \sum_{i=1}^{n} y_i \mu(A_i) + \beta \sum_{i=1}^{n} z_i \mu(A_i) = \alpha \int f\, d\mu + \beta \int g\, d\mu
\end{aligned}$$

ii) Monotonie:

$$\begin{array}{rc}
 & f \leq g \\
\Longleftrightarrow & f(x) \leq g(x) \quad \forall\, x \in \Omega \\
\Longleftrightarrow & y_i \leq z_i \qquad \forall\, i = 1, \ldots, n \\
\Rightarrow & \displaystyle\sum_{i=1}^{n} y_i \mu(A_i) \leq \sum_{i=1}^{n} z_i \mu(A_i) \\
\Longleftrightarrow & \displaystyle\int f\, d\mu \leq \int g\, d\mu
\end{array}$$

□

Lemma 5.1
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 $f = \sin : [0, \pi] \to \mathbb{R}$ (s. Abbildung 1). □

Definition 5.4 (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

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

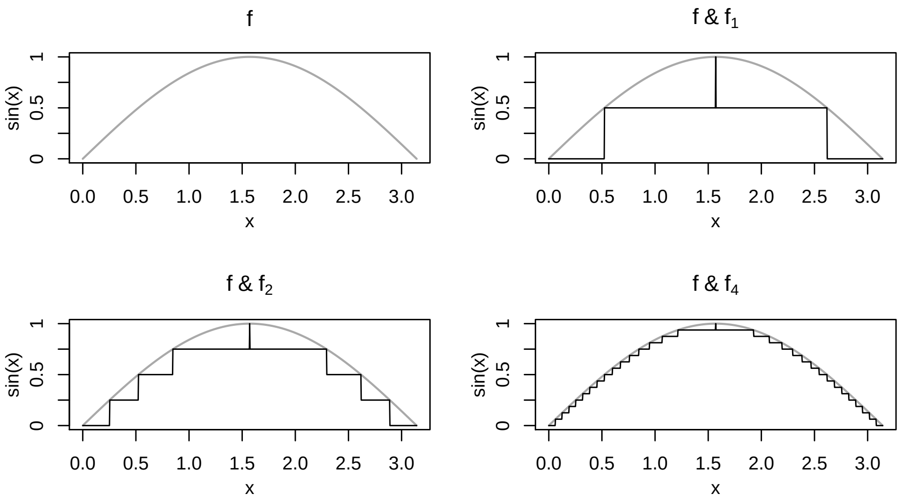


Abbildung 1: Approximation von $f = \sin \in M^+$ (auf $[0, \pi]$) durch $f_n \in T^+$ für $n = 1, 2, 4$.

Lemma 5.2
Sei $f \in M^+$ und $g \in T^+$ mit $g \leq f$. Dann gilt für eine Folge $(f_n)_{n\in\mathbb{N}} \in T^+$:

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

Beweis:

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

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

Und damit

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

Für $\epsilon \to 0$ folgt die Behauptung. $\square$
Jetzt allgemein $f : \Omega \to \bar{\mathbb{R}}$ meßbar.
Zerlegung in Positiv- und Negativteil: $f = \underbrace{f^+}_{\in M^+} - \underbrace{f^-}_{\in M^+}$

Definition 5.5 (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 (μ-) quasi-integrierbar, 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 $\iff$ $|f|$ integrierbar.

Bemerkung:

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

5.2 Eigenschaften des Integrals

Satz 5.2
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:

Erst: $f, g \in M^+$:
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 5.1 (i) für f_n, g_n

ii) $f_n \leq \max(f_n, g_n) \uparrow g$ für $f \leq g$ und aus Satz 5.1 (ii) folgt ii)

Für $f, g \in M$ folgen i) und ii) nach Definition.

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

$\square$

Satz 5.3
Für $f \in M^+$ gilt

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

Bemerkung:
Also für $f \in M^+$:

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

Beweis:

$$\text{„} \Longrightarrow : \text{“} \quad \left.\begin{array}{l} A := \{\omega \in \Omega | f(\omega) > 0\} \\ A_n := \left\{\omega \in \Omega | f(\omega) > \frac{1}{n}\right\} \quad \forall n \in \mathbb{N} \end{array}\right\} \Longrightarrow A_n \uparrow A$$

$$\frac{1}{n} I_{A_n} \leq f \;\forall\, n \in \mathbb{N}$$

$$0 \overset{\text{Maß}>0}{\leq} \frac{1}{n}\mu(A_n) = \int \frac{1}{n} I_{A_n}\, d\mu \overset{\text{Monotonie}}{\leq} \int f\, d\mu = 0$$

$$\Longrightarrow \mu(A_n) = 0 \,\forall n \in \mathbb{N} \xrightarrow{A_n \uparrow A} \mu(A_n) \uparrow \mu(A) = 0 \text{ (Stetigkeit von unten)}$$

$$
\text{“} \Longleftarrow : \text{“} \qquad \mu(A) = 0\,;\; f_n \in T^+, f_n \uparrow f \in M^+; m_n = \max_\Omega(f_n(x)) \in \mathbb{R}
$$

$$
\begin{aligned}
&\Longrightarrow \; \{\omega \in \Omega | f_n(\omega) > 0\} \subset A \\
&0 \leq \int f_n \, d\mu = \int_A f_n \, d\mu \leq \int m_n I_A \, d\mu = m_n \mu(A) = 0 \\
&\Longrightarrow \; \int f_n \, d\mu = 0 \; \forall \, n \in \mathbb{N} \\
&\Longrightarrow \; \int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu = 0
\end{aligned}
$$

□

Satz 5.4
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

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

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

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

5.3 Konvergenz des Lebesgue-Integrals

Ziel: Zufallsvariablen sind Abbildungen. Später sind die Realisierungen der Zufallsvariablen unsere beobachteten Daten. Wir wollen Aussagen darüber

machen, was passiert, wenn wir (unendlich) viele Daten, damit Zufallsvariablen haben (Grenzwertsätze).
Weg: Wir sehen uns Folgen von Abbildungen bzw. Funktionen an.

Lemma 5.3
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 $\underbrace{g_k}_{\in T^+} \leq \lim_{n\to\infty} f_n \quad \forall k\in\mathbb{N}$ (Monotonie).

Also $\int g_k\,d\mu \overset{\text{Lemma 5.2}}{\leq} \lim_{n\to\infty}\int f_n\,d\mu \quad \forall k\in\mathbb{N}$,
und somit $\lim_{k\to\infty}\int g_k\,d\mu \leq \lim_{n\to\infty}\int f_n\,d\mu$.

Umgekehrt: $f_k \leq \lim_{n\to\infty} g_n$ und also $\lim_{k\to\infty}\int f_k\,d\mu \leq \lim_{n\to\infty}\int g_n\,d\mu$.

Zusammen: $\lim_{n\to\infty}\int f_n\,d\mu = \lim_{n\to\infty}\int g_n\,d\mu$ □

Satz 5.5 (Monotone 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:

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

"$\geq$": Sei $(e_n)_{n\in\mathbb{N}} \in T^+$, $e_n \uparrow f$;
$A_n := \{\omega\in\Omega \,|\, c\,f_n(\omega) \geq e_k(\omega)\} \uparrow \Omega, k\in\mathbb{N}$ fest, $c>1$.
($c=1$ reicht nicht bei $f\in T^+, e_k = f\;\forall\, k\in\mathbb{N},\; f_n \leq f = e_k$)

Dann gilt: $c\, f_n(\omega) \geq e_k(\omega)\, I_{A_n}(\omega) \uparrow e_k(\omega)$.
$\Longrightarrow \int e_k\, d\mu = \lim\limits_{n\to\infty} \int e_k I_{A_n}\, d\mu \leq c \lim\limits_{n\to\infty} \int f_n\, d\mu.$
$\overset{c \text{ beliebig}}{\Longrightarrow} \int e_k\, d\mu \leq \lim\limits_{n\to\infty} \int f_n\, d\mu \quad \forall k \in N$
(sonst wäre $\int e_k\, d\mu > \lim_n \int f_n\, d\mu$ und es gäbe $c > 1$ mit $\int e_k\, d\mu > c \lim_n \int f_n\, d\mu$ $\to$ Widerspruch)
$\Longrightarrow \int \lim\limits_{n\to\infty} f_n\, d\mu = \int f\, d\mu = \lim\limits_{k\to\infty} \int e_k\, d\mu \leq \lim\limits_{n\to\infty} \int f_n\, d\mu$

□

Korollar 5.1 (zu Satz 5.5)
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 5.5 auf $g_n = \sum_{i=1}^{n} f_i$ (Partialsummen) an. □

Satz 5.6 (Dominierte 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 $|f_n| \leq g_n\ \forall\, n \in \mathbb{N}$. Sind g und $g_n\ \forall\, n \in \mathbb{N}$ 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:

Siehe Meintrup and Schäffler [2005], Theorem 2.14. □

5.4 Riemann- und Lebesgue-Integral

Satz 5.7 (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

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

Beweis:

Satz 2.17 Meintrup and Schäffler [2005].
Skizze: Stelle Ober- und Untersumme jeweils als Integral einer Treppenfunktion dar, dann benutze Satz 5.5 (monotone Konvergenz). □

Bemerkung:

Es gilt:
f Riemann-integrierbar $\implies$ f Lebesgue-integrierbar, aber
f Lebesque-integrierbar $\not\Longrightarrow$ f Riemann-integrierbar.

Bemerkung:

Riemann-Integrale sind nur für bestimmte $\Omega \subset \mathbb{R}^n$ definiert, Lebesque-Integrale auf beliebigen Messräumen $(\Omega, \mathcal{F})$ bzgl. beliebiger Maße μ.

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].

6 Dichte

Lemma 6.1
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 $\dot{\bigcup}_{n=1}^{\infty} A_n = A$ und damit $f\, I_A = \sum_{n=1}^{\infty} f\, I_{A_n}$.

$$\begin{aligned}
(f \odot \mu)(A) &= \int f\, I_A\, d\mu = \int \left(\sum_{n=1}^{\infty} f\, I_{A_n}\right) d\mu \\
&= \sum_{n=1}^{\infty} \int f\, I_{A_n}\, d\mu \quad (\text{da } f\, I_{A_n} \in M^+ \;\forall\, n \in \mathbb{N}) \\
&= \sum_{n=1}^{\infty} (f \odot \mu)(A_n)
\end{aligned}$$

Die zweite Behauptung folgt für $g = I_A$ aus der Definition, und damit für allgemeine $g \in M^+$ via Standardprozedur. □

Definition 6.1 (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 $f \odot \mu$.

Satz 6.1
Ist $f \in M^+$ und $A \in \mathcal{F}$ mit $\mu(A) = 0$, so folgt:

$$(f \odot \mu)(A) = 0.$$

Beweis:

Ist $A \in \mathcal{F}$ eine μ-Nullmenge, so ist auch $\{\omega : f(\omega)\, I_A(\omega) > 0\}$ eine μ-Nullmenge, d.h. $f\, I_A = 0$ μ-fast überall. Daraus folgt nach Satz 5.3

$$(f \odot \mu)(A) = \int f\, I_A d\mu = 0.$$

□

Definition 6.2 (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:

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

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

Satz 6.2 (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

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

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

Beweis:

" $\Longleftarrow$ ": Sei $A \in \mathcal{F}$ μ-Nullmenge, d.h. $\mu(A) = 0$ und $Z := \{\omega : (f\, I_A)(\omega) > 0\}$. $Z \subset A \implies \mu(Z) \leq \mu(A) = 0 \implies \mu(Z) = 0$. Damit: $\nu(A) = (f \odot \mu)(A) = \int_A f \, d\mu = \int f \, I_A \, d\mu \overset{\text{Satz 5.3}}{=} 0$

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

□

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.

Otto 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.

Beispiel 6.1
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei μ ein Maß, welches $\mathbb{P}$ dominiert, also aus $\mu(A) = 0 \implies \mathbb{P}(A) = 0$. Dann gibt es eine Dichte f mit $\mathbb{P} = f \odot \mu$, die Dichte der Verteilung $\mathbb{P}$.
Bei $\mathcal{F} = \mathcal{B}$ gilt für die Verteilungsfunktion

$$F(x) = \mathbb{P}(]-\infty, x]) = \int_{]-\infty,x]} f d\mu$$

- *Für $\mu = \mu_Z$: $\int_A f d\mu_Z = \sum_{\omega \in A} f(\omega)$*
- *Für $\mu = \lambda$: $\int_{[u,o]} f d\lambda = \int_u^o f(x)dx$, falls f Riemann-integrierbar.*

7 Arten von Verteilungen

7.1 Diskrete Verteilungen

Definition 7.1 (Diskreter Wahrscheinlichkeitsraum)
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Gibt es eine (endlich oder unendlich) abzählbare 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 (Zähldichte)
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

$$\mathbb{P}(A) = \int_A f \, d\mu_Z = \sum_{\omega \in A \cap T} f(\omega) \quad \forall A \in \mathcal{F}. \tag{1}$$

*Umgekehrt gibt es zu jeder Funktion f der obigen Gestalt **genau** ein diskretes Wahrscheinlichkeitsmaß $\mathbb{P}$ auf $(\Omega, \mathcal{F})$, so dass (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!
<u>Schritt I</u>: Sei $g = I_B$, $B \subset \Omega$:

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

<u>Schritt II</u>: Sei $g = \sum_{i=1}^n c_i I_{B_i} \in T^+$

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

<u>Schritt III</u>: Sei $g \in M^+$ (Träger immer noch abzählbar.)

Mit Lemma 5.2 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 $\int_A f\,d\mu_Z = \sum_{\omega\in A} f(\omega)\ \forall A \in \mathcal{F}$.

$$\begin{aligned}\text{Desweiteren } \mathbb{P}(A) = \mathbb{P}(A\cap T) &\overset{\mathbb{P}\text{ ist diskret}}{=} \sum_{\omega\in A\cap T}\mathbb{P}(\{\omega\}) \overset{\text{Def. } f}{=} \sum_{\omega\in A} f(\omega) \\ &= \int_A f\,d\mu_Z \overset{\text{Lemma 6.1}}{=} (f\odot\mu_Z)(A)\end{aligned}$$

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

Umgekehrt: Sei f Zähldichte und $\mathbb{P} = f \odot \mu_Z$. Per Definition:

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

$\square$

Bemerkung:

Jedes diskrete Wahrscheinlichkeitsmaß $\mathbb{P}$ hat genau eine Zähldichte f bezüglich μ_Z. Dies folgt aus Satz 6.2.

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

Beispiel 7.2 (Bernoulli-Verteilung)
$\Omega = \{\underbrace{\text{"Kopf"}}_{\omega_1}, \underbrace{\text{"Zahl"}}_{\omega_2}\}$ *bzw.* $\Omega = \{\text{"defekt"}, \text{"OK"}\}$
Zähldichte $f(\omega) = \begin{cases} p & \omega = \omega_1 \\ 1-p & \omega = \omega_2 \end{cases}$
Parameter $p \in [0,1]$*: "Erfolgswahrscheinlichkeit"*

$$\mathbb{P}(A) = \int_A f \, d\mu_Z = \sum_{\omega \in A} f(\omega) \; \forall A \in \mathcal{P}(\Omega) = \{\{\omega_1\}, \{\omega_2\}, \Omega, \emptyset\}$$

$$B(1,p) := \mathbb{P}$$

Beispiel 7.3 (Binomial-Verteilung)
$\Omega = \{0, 1, 2, \ldots, n\}$ *endlich.*
Zähldichte $f(\omega) = \binom{n}{\omega} p^\omega (1-p)^{n-\omega} \qquad \forall \omega \in \Omega,\ p \in [0,1], n \in \mathbb{N}$

$$\mathbb{P}(A) = \int_A f \, d\mu_Z = \sum_{\omega \in A} \binom{n}{\omega} p^\omega (1-p)^{n-\omega} \qquad \forall A \in \mathcal{F} = \mathcal{P}(\Omega)$$

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \binom{n}{\omega} p^\omega (1-p)^{n-\omega} = \sum_{\omega=0}^{n} \binom{n}{\omega} p^\omega (1-p)^{n-\omega}$$
$$\overset{Bin.\ Formel}{=} (p + (1-p))^n = 1^n = 1 \quad \checkmark$$

Parameter: n, p.
Definiere $B(n,p) := \mathbb{P}$

Beispiel 7.4 (Bernoulli-Verteilung II)
Wahrscheinlichkeitsraum $(\Omega_1 = \{\text{"Kopf"}, \text{"Zahl"}\}, \mathcal{P}(\Omega_1), B(1,p))$,
Meßraum $(\Omega_2 = \{0,1\}, \mathcal{P}(\Omega_2))$.

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

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

Weiterhin: $\mathbb{P}_X = f_X \odot \mu_Z$ *mit* $f_X : \underbrace{\{0,1\}}_{\Omega_2!} \to [0,1];\quad f_X(x) = \begin{cases} p^x(1-p)^{1-x} & x \in \Omega_2 \\ 0 & \textit{sonst} \end{cases}$.
Also ist $(\Omega_2, \mathcal{P}(\Omega_2), \mathbb{P}_X)$ *ein Wahrscheinlichkeitsraum.*
$X \sim B(1,p)$*: "X ist Bernoulli-verteilt".*
Äquivalent für $X \sim B(n,p)$.

Beispiel 7.5 (Poisson-Verteilung)
$\Omega = \mathbb{N}_0$
Zähldichte $f(\omega) = \dfrac{\lambda^\omega e^{-\lambda}}{\omega!}\ \forall\, \omega \in \Omega$
Parameter: $\lambda \in \mathbb{R}^+$ *("Rate")*

$$\begin{aligned} \mathbb{P}(A) &= \int_A f\, d\mu_Z = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A} \frac{\lambda^\omega e^{-\lambda}}{\omega!} \\ \mathbb{P}(\Omega) &= \sum_{\omega \in \Omega} \frac{\lambda^\omega e^{-\lambda}}{\omega!} = \sum_{\omega=0}^{\infty} \frac{\lambda^\omega e^{-\lambda}}{\omega!} \\ &= \frac{1}{\exp(\lambda)} \sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!} = \frac{1}{\sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!}} \sum_{\omega=0}^{\infty} \frac{\lambda^\omega}{\omega!} \\ &= 1 \quad \checkmark \end{aligned}$$

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

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:

$$\mathbb{P}([a,b]) = \int_{[a,b]} f \, d\lambda \quad \in [0,1] \quad \forall a, b \text{ Insbesondere } \int_{\mathbb{R}} f \, d\lambda = 1. \qquad \square$$

Bemerkung:

Ist $g : \mathbb{R} \to \mathbb{R}$ stetig und $f = g \cdot I_{]a,b[}$ eine Dichte, so heißt f stetige Dichte, $\mathbb{P} = f \odot \lambda$ heißt stetige Verteilung.
$F_{\mathbb{P}}$ ist auf ganz $\mathbb{R}$ stetig.
Wir nennen $T = [a, b]$ Träger der stetigen Verteilung.
Allgemeiner: Sei $\mathbb{P}$ eine (beliebige) Verteilung mit Dichte $f(x)$, dann ist $T = \{x | f(x) > 0\}$ der Träger der Verteilung.

Beispiel 7.6 (Stetige Gleichverteilung)
$\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B},\ a < b \in \mathbb{R}$

$$\begin{aligned}
f(\omega) &= \frac{1}{b-a} I_{]a,b[}(\omega) \qquad \forall \omega \in \Omega \\
\mathbb{P}(A) &= \int_A f \, d\lambda = \frac{1}{b-a} \int I_{]a,b[\cap A} \, d\lambda = \frac{\lambda\left(]a,b[\cap A\right)}{b-a} \\
&= 1 \textit{ für } A =]a,b[\\
F(x) &= \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases} \quad \textit{Verteilungsfunktion} \\
U(a,b) &:= \mathbb{P}
\end{aligned}$$

Satz 7.3 (Inversionsmethode)
Sei X eine Zufallsvariable mit Verteilungsfunktion F und U eine auf $(0,1)$ gleichverteilte Zufallsvariable. Dann gilt

i) *$F^{-1}(U)$ hat die Verteilungsfunktion F.*

ii) *$\mathbb{P}(F(X) \le u) \le u$ und $\mathbb{P}(F(X) \ge u) \ge 1 - u$ für $u \in [0,1]$*

iii) *Ist F stetig in $F^{-1}(u)$, dann ist $\mathbb{P}(F(X) \le u) = u$ und $\mathbb{P}(F(X) \ge u) = 1 - u$.*

Beweis:

i) Aus Definition von F^{-1} und Monotonie von F folgt:
$\mathbb{P}(F^{-1}(U) > x) = \mathbb{P}(U > F(x)) = 1 - F(x)$.
Damit $\mathbb{P}(F^{-1}(U) \leq x) = F(x)$.

ii) Für $u = 1$ ist die Aussage trivial. Im Folgenden $u < 1$.

Sei $z = \sup\{x : F(x) \leq u\}$. Dann gilt $F(z-) \leq u$.

Ist $F(z) = u \implies \mathbb{P}(F(X) \leq u) = \mathbb{P}(X \leq z) = F(z) = u$.

Ist $F(z) < u \implies \mathbb{P}(F(X) \leq u) = \mathbb{P}(X < z) = F(z-) \leq u)$.

Für Ungleichheit muss $F(z-) < u$ gelten. Also $z = F^{-1}(u)$ und damit $\mathbb{P}(F(X) \leq u) = \mathbb{P}(X < z) = F(z-) < u$.

iii) Ist F stetig in $F^{-1}(u)$, so ist $u \leq F(F - 1(u)) = F(F - 1(u)-) \leq u$ und somit muss Gleichheit gelten.

Weiter gilt $\mathbb{P}(F(X) \geq u) = \mathbb{P}(X \geq F^{-1}(u)) = 1 - F(F^{-1}(u)-) \geq 1 - u$. Ist F stetig in $F^{-1}(u)$, so haben wir Gleichheit.

$\square$

Beispiel 7.7 (Normalverteilung)
$\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B};\ \mu, \sigma \in \mathbb{R};\ \sigma > 0$
Dichte der Normalverteilung:

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

Dichte der Standardnormalverteilung ($\mu = 0, \sigma = 1$)*:*

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\ \exp\left(-\frac{1}{2}x^2\right)$$

Verteilungsfunktion der Standardnormalverteilung:

$$\mathbb{P}(]-\infty, x]) = \int_{-\infty}^{x} \varphi(y)\, dy =: \Phi(x)$$

Verteilungsfunktion der Normalverteilung

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

Beweis:

z.Z. $F(\infty) = \Phi(\infty) = \int\limits_{-\infty}^{\infty} \varphi(x)dx = 1$

$$\underline{\text{Zunächst}}: \left(\int\limits_0^\infty \exp(-y^2)\,dy\right)^2 = \int\limits_0^\infty \exp(-y^2)\,dy \int\limits_0^\infty \exp(-t^2)\,dt$$

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

$$\overset{\text{Fubini}}{=} \int\limits_0^\infty \underbrace{\int\limits_0^\infty y \cdot \exp(-(1+x^2)y^2)\,dy}_{\left[\frac{e^{-(1+x^2)y^2}}{-2(x^2+1)}\right]_0^\infty = \frac{1}{2}\frac{1}{1+x^2}}\ dx = \frac{1}{2}\left[\arctan(x)\right]_0^\infty = \frac{1}{2}\left(\frac{\pi}{2} - 0\right) = \frac{\pi}{4}$$

$$\Longrightarrow \int\limits_0^\infty \exp(-y^2)\,dy = \frac{\sqrt{\pi}}{2} \overset{\textit{Symmetrie}}{\Longrightarrow} \int\limits_{-\infty}^\infty \exp(-y^2)\,dy = \sqrt{\pi}$$

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

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

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

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

□

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

Beispiel 7.8 (Exponentialverteilung)

$\Omega = \mathbb{R},\ \mathcal{F} = \mathcal{B}$

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

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

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

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \int_{-\infty}^{x} \lambda \exp(-\lambda y)\, dy = \lambda \int_{0}^{x} \exp(-\lambda y)\, dy$$

$$= \lambda \frac{1}{-\lambda} \left[\exp(-\lambda y)\right]_0^x = \begin{cases} 1 - \exp(-\lambda x) & x > 0 \\ 0 & \textit{sonst} \end{cases}$$

$$\to 1 \quad \textit{für } x \to \infty$$

Definition 7.3 (Γ-Funktion)

$$\begin{aligned} \Gamma : \mathbb{R}^+ &\to \mathbb{R}^+ \\ \Gamma(z) &= \int_0^\infty t^{z-1}\, e^{-t}\, dt \end{aligned}$$

Definition 7.4
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 7.4

$$Ga\,(k = 1, \lambda) = Exp(\lambda)$$

Beweis:

Setze $k = 1$ in Dichte der Gamma-Verteilung. □

Satz 7.5 (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,* $\frac{\partial g(x)}{\partial x} \neq 0$ *mit* $h = g^{-1}$.

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

Beweis:

Fallunterscheidung:

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

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

(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)

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

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

zusammen (A) und (B):

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

$\square$

Beispiel 7.9
Sei $X \sim Exp(\lambda)$, was ist $F(X^2)$?

$$\begin{aligned}
g(x) &= x^2 \\
h(y) &= g^{-1}(y) = \sqrt{y} \ \textit{(da } f_X(x) = 0 \textit{ für } x \leq 0 \textit{ keine Fallunterscheidung nötig)} \\
\frac{\partial h(y)}{\partial y} &= \frac{1}{2\sqrt{y}} \\
f_{X^2}(y) &= f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} = \lambda \exp\left(-\lambda\sqrt{y}\right) \cdot \frac{1}{2\sqrt{y}} \cdot I_{[0,\infty[}(\sqrt{y}) \\
F_{X^2}(x) &= \int_0^x f_{X^2}(y)\, dy = 1 - \exp\left(-\lambda\sqrt{y}\right)
\end{aligned}$$

7.3 Gemischte Verteilungen

Beispiel 7.10 (Gemischte Verteilung)
Ein Smartphone-Hersteller verkauft eine erweiterte Reparaturversicherung. Aus langjähriger Erfahrung ist ihm bekannt, dass der Anteil $\alpha > 0$ 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,
- $\mathbb{P}(0) = \alpha > 0$.

Wir suchen die Dichte und das dominierende Maß (siehe Satz von Radon-Nikodym, Satz 6.2).
$\mathbb{P}$ wird nicht vom Lebesguemaß dominiert: $\lambda(\{0\}) = 0$, aber $\mathbb{P}(0) > 0$.
μ_Z ist nicht geeignet, da kein σ-endliches Maß auf $(\mathbb{R}, \mathcal{B})$.
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

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

Sei dazu $B \in \mathcal{B}$ eine μ-Nullmenge, also $\mu(B) = 0$. Dann gilt:

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

und somit

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

Offensichtlich ist $\{0\} \notin B$. Damit folgt (mit $f(x)$ Dichte der Exponentialverteilung):

$$\mathbb{P}(B) = \int_B f(x)\,d\lambda(x) = \int f(x) I_B(x)\,d\lambda(x) \overset{\text{Bem. zu Satz 5.3}}{=} 0.$$

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

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

Setze nun

$$f^\star(x) := (1-\alpha) \cdot f(x) \qquad \forall\, x \in \mathbb{R} \setminus \{0\}$$

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, wir konstruieren so wie in dem Bsp. oben einfach ein entsprechendes dominierendes Maß über beliebige Messräume und eine Dichte dazu.

7.4 Mehrdimensionale Verteilungen

Definition 7.5 (Produktmaßraum)
Seien $(\Omega_i, \mathcal{F}_i, \mu_i), i = 1, \ldots, n$ *Maßräume.*
Ihr *Produktmaßraum* *ist der Maßraum*

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

mit Basismenge

$$\Omega_1 \times \Omega_2 \times \cdots \times \Omega_n,$$

Produkt-σ*-Algebra*

$$\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \cdots \otimes \mathcal{F}_n = \sigma \left(\left\{ \bigtimes_{i=1}^{n} A_i | A_i \in \mathcal{F}_i, i = 1, \ldots, n \right\} \right),$$

wobei $\left(\left\{\bigtimes_{i=1}^{n} A_i | A_i \in \mathcal{F}_i, i = 1, \ldots, n\right\}\right)$ *ein durchschnittsstabiler Erzeuger von* $\bigotimes_{i=1}^{n} \mathcal{F}_i$ *ist,*
und Produktmaß $\bigotimes_{i=1}^{n} \mu_i$ *mit*

$$\left(\bigotimes_{i=1}^{n} \mu_i \right) \left(\bigtimes_{i=1}^{n} A_i \right) = \prod_{i=1}^{n} \mu_i(A_i) \; \forall \, A_i \in \mathcal{F}_i \; \forall \, i = 1, \ldots, n \, .$$

Definition 7.6 (n-dimensionale ZV, Verteilungsfunktion)
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ *ein Wahrscheinlichkeitsraum und* $X : \Omega \to \mathbb{R}^n$ *eine* n*-dimensionale reelle ZV,* $X = (X_1, \ldots, X_n)$*. Dann heißt die Funktion*

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

die n*-dimensionale Verteilungsfunktion von* X*.*

Bemerkung:

Für $x \in \mathbb{R}^n$ und $h \in \mathbb{R}$ sei $(x + h) := (x_1 + h, \ldots, x_n + h)$. 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) $\lim\limits_{x_k \to -\infty} F_X(x) = 0$ für alle $k \in \{1, \ldots, n\}$, $\lim\limits_{h \to \infty} F_X(x+h) = 1$.

Bemerkung:

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

Satz 7.6 (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] □

Bemerkung:

$(\mathbb{R}^n, \mathcal{B}^n, \mathbb{P} = f \odot \lambda^n)$ mit ZV $X(\omega) = \omega \in \mathbb{R}^n$
Da f Riemann-integrierbar gilt mit dem Satz von Fubini:

$$\begin{aligned} \Longrightarrow \qquad \mathbb{P}_X(]a,b]) &= \int f \, I_{]a,b]} \, d\lambda^n \\ &= \int\limits_{a_1}^{b_1} \ldots \int\limits_{a_n}^{b_n} f(x_1, \ldots, x_n) \, dx_n \, dx_{n-1} \ldots dx_1 \end{aligned}$$

Satz 7.7
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 3.3 (Meßbarkeit der Komposition) □

Definition 7.7 (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 <u>Randverteilung</u> oder <u>marginale Verteilung</u> von X_j.

Bemerkung:

Besitzt X die Dichte f, so hat X_j die Dichte

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

mit z.B. $\nu = \lambda$ oder $\nu = \mu_Z$
f_j heißt Randdichte von X_j.
Vgl. Satz 7.5.

8 Momente

8.1 Erwartungswert und Varianz

Definition 8.1 (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 <u>Erwartungswert</u> 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:

$$
\begin{aligned}
\mathbb{E}(X) &= \int X\, d\mathbb{P} = \int g \circ X\, d\mathbb{P} \\
&\overset{\text{Satz 5.4}}{=} \int g\, d\mathbb{P}_X = \int g\, d(f_X \odot \mu) \overset{\text{Lemma 6.1}}{=} \int g \cdot f_X\, d\mu \\
&= \begin{cases} \sum\limits_{x \in T} x\; f_X(x) & \mu = \mu_Z \text{ (diskret)} \\ \int\limits_{-\infty}^{\infty} x\; f_X(x)\, dx & \mu = \lambda \text{ (stetig)} \end{cases}
\end{aligned}
$$

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

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

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

Beispiel 8.1 (Bernoulli-Verteilung)
Sei $X \sim B(1, p)$

Dann:

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

Beispiel 8.2 (Binomialverteilung)
Sei $(\Omega_1 = \{0,\dots,n\}, \mathcal{P}(\Omega_1), B(n,p))$; $\quad (\Omega_2 = \Omega_1, \mathcal{P}(\Omega_2))$;
$X : \Omega_1 \to \Omega_2$; $\quad \omega \mapsto X(\omega) = \omega$ *(Identität).*

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

Definition 8.2 (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:

$$\begin{aligned}\mathbb{P}_X = f_X \odot \lambda &\implies \mathbb{V}(X) = \int (x-\mathbb{E}(X))^2 f(x)\, dx\\ \mathbb{P}_X = f_X \odot \mu_Z &\implies \mathbb{V}(X) = \sum_{x\in T}(x-\mathbb{E}(X))^2 f(x)\end{aligned}$$

Beispiel 8.3

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

Beispiel 8.4 (Bernoulli-Verteilung)

Sei $X \sim B(1,p)$.

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

Beispiel 8.5 (Binomialverteilung)
$X \sim B(n,p)$

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

Beispiel 8.6 (Gleichverteilung)

$$\begin{aligned}\mathbb{E}(X) &= \frac{1}{b-a} \int_a^b x \, dx = \frac{a+b}{2}, \\ \mathbb{V}(X) &= \frac{1}{b-a} \int_a^b x^2 \, dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a} \left[\frac{1}{3} x^3\right]_a^b - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{4(b^3 - a^3) - 3(b-a)(a^2 + 2ab + b^2)}{12(b-a)} = \frac{b^3 - a^3 - 3a^2 b + 3ab^2}{12(b-a)} \\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

8.2 Allgemeine Momente

Definition 8.3 (Momente)
Sei X eine integrierbare Zufallsvariable und $n \in \mathbb{N}$, so daß X^n quasi-integrierbar

ist. Dann heißt

$$\begin{aligned} m_n(X) &= \mathbb{E}(X^n) && n\text{-tes } \underline{\text{Moment}}, \\ m_{(n)}(X) &= \mathbb{E}(|X|^n) && n\text{-tes } \underline{\text{absolutes Moment}}, \text{ und} \\ m_n^0(X) &= \mathbb{E}((X - \mathbb{E}(X))^n) && n\text{-tes } \underline{\text{zentriertes Moment}} \text{ von } X. \end{aligned}$$

Bemerkung:

$$\begin{aligned} \mathbb{E}(X) &= m_1(X) && \text{heißt erstes Moment.} \\ m_1^0(X) &= \mathbb{E}(X - \mathbb{E}(X)) \equiv 0 \\ \mathbb{V}(X) &= m_2^0(X) && \text{ist das zweite zentrierte Moment.} \\ Skew(X) &:= \frac{m_3^0(X)}{m_2^0(X)^{\frac{3}{2}}} = \frac{\mathbb{E}((X - \mathbb{E}(X))^3)}{\sqrt{\mathbb{V}(X)}^3} && \text{ist die } \underline{\text{Schiefe}} \text{ von } X. \\ K(X) &:= \frac{m_4^0(X)}{(m_2^0(X))^2} && \text{ist die } \underline{\text{Kurtosis}} \text{ oder } \underline{\text{Wölbung}} \text{ von } X. \end{aligned}$$

Definition 8.4 (symmetrische Verteilung)
Sei $\mathbb{P}$ Verteilung auf $(\mathbb{R}, \mathcal{B})$.
$\mathbb{P}$ heißt symmetrisch um a $\in \mathbb{R} \iff \mathbb{P}(]-\infty, a-x]) = \mathbb{P}([a+x, \infty[)$

Satz 8.1
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)
$$\int g \, d\mathbb{P}_X = \int g(x) \, d\mathbb{P}_X(x) = \int g\,(2a - x) \, d\mathbb{P}_X(x)$$

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)

$$\begin{aligned}
\Longrightarrow \mathbb{P}((X-a) \leq t) &= \mathbb{P}(X \leq a+t) \\
&= \mathbb{P}(X \geq a-t) = \mathbb{P}(-X+a \leq t) && \text{(Symmetrie)} \\
\Longrightarrow \int g \circ X \, d\mathbb{P} &= \int g \, d\mathbb{P}_X = \mathbb{E}(g(X)) \\
&= \mathbb{E}(g(a+(X-a))) = \mathbb{E}(g(a-(X-a))) && \text{(Symmetrie)} \\
&= \mathbb{E}(g(2a-X)) = \int g(2a-x) \, d\mathbb{P}_X(x)
\end{aligned}$$

ii)

$$m_1(X) = \int X \, d\mathbb{P} \overset{i)}{=} \frac{1}{2}\left[\int X \, d\mathbb{P} + \int (2a-X) \, d\mathbb{P}\right] = \frac{1}{2} \, 2\, a = a$$

iii)

$$\begin{aligned}
m_k^0(X) &= \int (X - m_1(X))^k \, d\mathbb{P} = \int (X-a)^k \, d\mathbb{P} \\
&\overset{i)}{=} \frac{1}{2}\left[\int (X-a)^k \, d\mathbb{P} + \int (2a-X-a)^k \, d\mathbb{P}\right] \\
&= \frac{1}{2}\left[\int (X-a)^k \, d\mathbb{P} + \int (-1)^k (X-a)^k \, d\mathbb{P}\right] \\
&= \frac{1}{2}\left[\int (X-a)^k \, d\mathbb{P} - \int (X-a)^k \, d\mathbb{P}\right] = 0 \qquad (k \text{ ungerade})
\end{aligned}$$

$\square$

Beispiel 8.7
Dichte der Normalverteilung: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{\sigma^2}(x-\mu)^2\right)$
$X \sim N(\mu, \sigma^2)$ *symmetrisch um* μ. $\Longrightarrow m_1(X) = \mathbb{E}(X) = \mu$

8.3 Ungleichungen

Satz 8.2 (Markov- und Chebyshev-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 für $n = 1$

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

und für $\mathbb{E}(|X|) < \infty$

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

Beweis:

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

$$\begin{aligned}
& \alpha \cdot I_{\{Y \geq \alpha\}} \leq Y \cdot I_{\{Y \geq \alpha\}} \qquad \leq Y. \\
\Longrightarrow & \int \alpha \cdot I_{\{Y \geq \alpha\}}\, d\mathbb{P} = \alpha \cdot \int I_{\{Y \geq \alpha\}}\, d\mathbb{P} = \alpha \cdot \mathbb{P}(\{Y \geq \alpha\}) \\
& \leq \int Y\, I_{\{Y \geq \alpha\}}\, d\mathbb{P} \;\leq \int Y\, d\mathbb{P} = \mathbb{E}(Y) \\
\Longrightarrow & \quad \mathbb{P}(Y \geq \alpha) \leq \frac{1}{\alpha} \int_{\{Y \geq \alpha\}} Y\, d\mathbb{P} \;\leq \frac{1}{\alpha}\,\mathbb{E}(Y)
\end{aligned}$$

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 $n = 1$.
Chebyshev: $n = 2$ und $X = X - \mathbb{E}(X)$. □

Satz 8.3 (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:

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

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

$$
\begin{aligned}
\implies \qquad f(x) &= \sup_{v \in V} v(x) \\
\implies \mathbb{E}(f(X)) &= \mathbb{E}(\sup_{v \in V} v(X)) \overset{\forall\, v_0 \in V}{\geq} \mathbb{E}(v_0(X)) = v_0(\mathbb{E}(X)) \\
\implies \mathbb{E}(f(X)) &\geq \sup_{v_0 \in V} v_0\,(\mathbb{E}(X)) \\
\mathbb{E}(f(X)) &\geq f(\mathbb{E}(X))
\end{aligned}
$$

$\square$

Beispiel 8.8

a) $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$ *mit* $f(x) = x^2$ *konvex*

b) $\mathbb{P}(X > 0) = 1 \implies \mathbb{E}\left(\frac{1}{X}\right) \geq \frac{1}{\mathbb{E}(X)}$ *mit* $f(x) = \frac{1}{x}$

Definition 8.5 (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 <u>*p-Norm*</u> *von* f.

Bemerkung:

Für eine Zufallsvariable X ist also $\|X\|_p$ die p-te Wurzel aus dem p-ten absoluten Moment von X:

$$\mathbb{E}(|X|^p) = \|X\|_p^p$$

Definition 8.6 (L^p-Raum)
Für $p \geq 1$ *sei*

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

Satz 8.4 (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 $\|g\|_q = 0 \implies f \cdot g = 0$ μ-f.ü. $\quad\surd$
- ($\|f\|_p = \infty$ oder $\|g\|_q = \infty$) <u>und</u> $\|f\|_p \|g\|_q > 0 \quad\surd$
- $0 < \|f\|_p, \|g\|_q < \infty$:
  Sei $x, y \geq 0,\ \alpha, \beta \geq 0$ mit $\alpha + \beta = 1$.

$$\begin{aligned}
&\ln(x^\alpha y^\beta) = \alpha \ln x + \beta \ln y \leq \ln(\alpha x + \beta y) \qquad \text{(Konkavität)} \\
&\implies x^\alpha y^\beta \leq \alpha x + \beta y
\end{aligned}$$

Mit $x := \dfrac{|f|^p}{\|f\|_p^p}; \quad y := \dfrac{|g|^q}{\|g\|_q^q}; \quad \alpha = \dfrac{1}{p},\ \beta = \dfrac{1}{q}$

$$\begin{aligned}
\frac{|f|\ |g|}{\|f\|_p \|g\|_q} &\leq \frac{1}{p} \frac{|f|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g|^q}{\|g\|_q^q} \\
\int \frac{|f|\ |g|\, d\mu}{\|f\|_p \|g\|_q} &\leq \frac{1}{p} \cdot \frac{1}{\|f\|_p^p} \underbrace{\int |f|^p\, d\mu}_{=\|f\|_p^p} + \frac{1}{q} \cdot \frac{1}{\|g\|_q^q} \underbrace{\int |g|^q\, d\mu}_{=\|g\|_q^q} \\
\frac{\int |f|\ |g|\, d\mu}{\|f\|_p \|g\|_q} &\leq \frac{1}{p} + \frac{1}{q} \overset{\text{n.V.}}{=} 1 \\
\int |fg|\, d\mu &\leq \|f\|_p \|g\|_q \\
\|fg\|_1 &\leq \|f\|_p \|g\|_q
\end{aligned}$$

□

Satz 8.5
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 $\left(\frac{1}{r} + \frac{1}{s}\right) = 1$.

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

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

Beispiel 8.9
$m_{(k)}(X) < \infty \implies m_j(X) < \infty \quad \forall j \leq k$

Satz 8.6 (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:
Für $p = 1$ oder $\|f\|_p = \infty$ oder $\|g\|_p = \infty$ oder $\|f + g\|_p = 0 \quad \checkmark$

Jetzt $p > 1, \|f||_p < \infty, \|g\|_p < \infty, \|f+g\|_p > 0, q := (1-\frac{1}{p})^{-1} \implies \frac{1}{p}+\frac{1}{q} = 1$:

$$\begin{aligned}
\|f + g\|_p^p &= \int |f + g|^p \, d\mu \\
&= \int |f + g| \, |f + g|^{p-1} \, d\mu \\
&\leq \int (|f| + |g|) \, |f + g|^{p-1} \, d\mu && \text{(Dreiecksungleichung)} \\
&= \int |f| \, |f + g|^{p-1} \, d\mu + \int |g| \, |f + g|^{p-1} \, d\mu \\
&\leq (\|f\|_p + \|g\|_p) \| \, |f + g|^{p-1} \, \|_q && (2 \times \text{Hölder})
\end{aligned}$$

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

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

Also da $p - \frac{p}{q} = p - p\left(1 - \frac{1}{p}\right) = 1$:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

$\square$

8.4 Weitere Lagemaße

Neben dem Erwartungwert gibt es andere (robustere) Lagemaße einer Verteilung.

Definition 8.7
Gegeben sei eine eindimensionale Zufallsvariable X mit Verteilung $\mathbb{P}$.
Eine reelle Zahl m für die gilt

$$\mathbb{P}(]-\infty, m]) \geq 0.5 \text{ und } \mathbb{P}([m, +\infty[) \geq 0.5$$

heißt Median von $\mathbb{P}$ bzw. X.

Beispiel 8.10
$X \sim B(2, 0.5)$ dann ist $m = 1$ Median, da $\mathbb{P}(X \leq 1) = 0.75$ und $\mathbb{P}(X \geq 1) = 0.75$.

Satz 8.7
Ist $\mathbb{P}$ symmetrisch um a, dann ist a der Median von $\mathbb{P}$.

Beweis:

Folgt direkt aus Def. 8.4 □

Definition 8.8
Gegeben sei eine eindimensionale Zufallsvariable X mit Verteilung $\mathbb{P}$.
Eine reelle Zahl q für die gilt

$$\mathbb{P}(]-\infty, q]) \geq p \text{ und } \mathbb{P}([q, +\infty[) \geq 1 - p$$

heißt p-Quantil von $\mathbb{P}$ bzw. X.

Definition 8.9
Gegeben sei eine eindimensionale Zufallsvariable X mit Verteilung $\mathbb{P}$ und Dichte $f(x)$.
Eine Zahl m heißt Modus von X bzw. $\mathbb{P}$ wenn $f(m)$ ein lokales Maximum von $f(x)$ ist. Das heißt, wenn gilt:

$$f(x) \leq f(m) \text{ für alle } x \in (m - \epsilon; m + \epsilon) \text{ für ein } \epsilon > 0.$$

Beispiel 8.11

- *Sei $X \sim Exp(1)$. Der Modus von X ist 0 und es gilt $\mathbb{P}(X > 0) = 1$.*
- *Der Modus der $B(n, 0.5)$-Verteilung ist $\frac{n}{2}$, wenn n gerade ist. Ist n ungerade, sind $\frac{n-1}{2}$ und $\frac{n+1}{2}$ Modi.*

Bemerkung:

- Die Definition lässt sich leicht auf mehrdimensionale Verteilungen erweitern.

- Da Dichten stetiger Zufallsvariablen nur fast sicher eindeutig sind, sind formal auch Modi nur fast sicher eindeutig.

Definition 8.10
Hat eine stetige Dichte nur einen Modus, so nennt man die Verteilung unimodal. Hat eine stetige Dichte mehrere Modi, so nennt man die Verteilung multimodal, bei zwei Modi speziell bimodal.

Beispiel 8.12
Mit $\phi(x)$ die Dichte der Standardnormalverteilung ist die Verteilung mit der Dichte $f(x) = 0.5\phi(x+2) + 0.5\phi(x-2)$ bimodal.

9 Erzeugende Funktionen

9.1 Momenterzeugende Funktion

Definition 9.1 (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 9.1
*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) $\left.\dfrac{\partial^n M(s)}{\partial^n s}\right|_{s=0} = \mathbb{E}(X^n)$.

Bemerkung:

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

Beweis:

i) $\mathbb{E}(\exp(sX)) < \infty \ \forall\, s \in]-a, a[\implies \exp(sX)$ ist $\mathbb{P}_X$- integrierbar
$\implies \exp(|sX|) = \sum_{n=0}^{\infty} \frac{|sX|^n}{n!}$ ist $\mathbb{P}_X$-integrierbar.
$\implies |X|^n$ und damit auch X^n ist $\mathbb{P}_X$-integrierbar $\forall\, n \quad \surd$

ii) Betrachte Folge von Partialsummen $f_k(X) = \displaystyle\sum_{n=0}^{k} \frac{(sX)^n}{n!} \xrightarrow{k\to\infty} \exp(sX)$.

$$\begin{aligned} M(s) = \mathbb{E}(\exp(sX)) &= \int \exp(sX)\, d\mathbb{P}_X = \int \lim_{k\to\infty} f_k\, d\mathbb{P}_X \\ &= \lim_{k\to\infty} \int f_k\, d\mathbb{P}_X = \lim_{k\to\infty} \mathbb{E}(f_k) && \text{(Satz 5.5)} \\ &= \sum_{n=0}^{\infty} \frac{s^n}{n!}\, \mathbb{E}(X^n) \quad s \in]-a, a[. \qquad \surd \end{aligned}$$

iii) Die Darstellung $M(s) = \sum_{n=0}^{\infty} \frac{s^n}{n!}\, \mathbb{E}(X^n) = 1 + \frac{s}{1!}\, \mathbb{E}(X) + \frac{s^2}{2!}\, \mathbb{E}(X^2) + \frac{s^3}{3!}\, \mathbb{E}(X^3) + \ldots$ ist eine eine Taylorreihe der Form $M(s) = \sum_{n=0}^{\infty} \frac{(s-a)^n}{n!} \left.\frac{\partial^n M(s)}{\partial^n s}\right|_{s=a}$ um den Aufpunkt a. Mit $a = 0$:
$\implies \left.\dfrac{\partial^n M(s)}{\partial^n s}\right|_{s=0} = \mathbb{E}(X^n) \quad \surd$

$\square$

Beispiel 9.1 (Standard-Normalverteilung)
$X \sim N(0,1)$

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

$$\left.\frac{\partial M(s)}{\partial s}\right|_{s=0} = s \cdot \exp\left(\frac{s^2}{2}\right)\Big|_{s=0} = 0 \implies \mathbb{E}(X) = 0$$
$$\left.\frac{\partial^2 M(s)}{\partial^2 s}\right|_{s=0} = s^2 \cdot \exp\left(\frac{s^2}{2}\right) + \exp\left(\frac{s^2}{2}\right)\Big|_{s=0} = 1 \implies \mathbb{E}(X^2) = 1$$
$$\implies \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1$$

Beispiel 9.2 (Normalverteilung)
Jetzt $Y \sim N(\mu, \sigma^2)$*:*
$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$
$F_X(x) = \Phi(x)$
$\iff$ *Y und $\sigma X + \mu$ haben dieselbe Verteilungsfunktion*
$\iff$ *Y und $\sigma X + \mu$ haben gleiche Verteilung: $Y = \sigma X + \mu$*

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

9.2 Charakteristische Funktion

Definition 9.2 (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:

Die charakteristische Funktion die Fourier-Transformierte der Dichte. Es gilt

$$\varphi_X(t) = \int f_X(x) \exp(\mathrm{i}tx) d\mathbb{P}$$

Beispiel 9.3 (Charakteristische Funktionen)

a) $X \sim B(n, p)$*:*

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

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

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

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

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

Satz 9.2 (Eindeutigkeit der Charakteristischen Funktion)
Seien φ_X und φ_Y charakteristische Funktionen und $a, b \in \mathbb{R}$. Dann gilt:

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

Beweis:

Beweis zu 16.4.7 in Schmidt [2011]. □

Satz 9.3 (Linearität der Charakteristischen Funktion)
Sei φ_X eine charakteristische Funktion und $a, b \in \mathbb{R}$. Dann gilt: $\varphi_{aX+b}(t) = \exp(itb)\varphi_X(at)$.

Beweis:

Linearität des Lebesgue-Integrals. □

Satz 9.4
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 9.5
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:

Beweis zu 16.4.6 in Schmidt [2011]. □

Teil III
Mehrdimensionale Zufallsvariablen

10 Unabhängigkeit von Zufallsvariablen

Definition 10.1 (Stochastisch unabhängige Familien) *i) Eine Familie von Mengensystemen $\mathcal{F}_i \subset \mathcal{F}$, $i \in I \neq \emptyset$, heißt stochastisch unabhängig (*stu*), wenn dies für jede Familie von Ereignissen $A_j \in \mathcal{F}_i$, $j \in J$, für alle $i \in I$ gilt. Vgl. Definition 2.12*

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:

$$\sigma(X_i^{-1}(\mathcal{F}_i)) \overset{\text{Satz 3.1}}{=} \underbrace{X_i^{-1}(\mathcal{F}_i) \subset \mathcal{F}}_{X_i \text{ ist ZV, also } \mathcal{F}\text{-}\mathcal{F}_i\text{-meßbar}}$$

Bemerkung:

$X_1, \ldots, X_n$ sind reelle ZV auf $(\mathbb{R}, \mathcal{B}, \mathbb{P})$. Dann gilt:

$$X_1, \ldots, X_n \text{ stu} \iff$$
$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) \quad \forall B_i \in \mathcal{B},$$

denn $\sigma(X_i) = X_i^{-1}(\mathcal{B}) \subset \mathcal{B}$ und nach Definition 10.1 muß obige Gleichung für alle möglichen Ereignisse B_i gelten.

Satz 10.1 (Unabhängigkeit n-dimensionaler ZV)

i) *Sei $X = (X_1, \dots, X_n)$ reelle n-dimensionale ZV.*

$$X_1, \dots, X_n \text{ stu} \iff \forall c \in \mathbb{R}^n :$$
$$\begin{aligned} \mathbb{P}(X \leq c) &= \mathbb{P}(X_1 \leq c_1, \dots, X_n \leq c_n) \\ &= \prod_{i=1}^{n} \mathbb{P}(X_i \leq c_i), \\ \text{bzw. } F_X(c) &= \prod_{i=1}^{n} F_{X_i}(c_i). \end{aligned}$$

ii) *Ist X diskret, so gilt:*

$$X_1, \dots, X_n \text{ stu} \iff \forall x \in T_1 \times T_2 \times \dots, \times T_n \quad (T_i \text{ Träger von } X_i) :$$
$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i).$$

Beweis:

X diskret:
Zweite Aussage folgt direkt aus Def. 10.1 ii). Noch zu zeigen:

$$\begin{aligned} \mathbb{P}(X_1 \leq c_1, \dots, X_n \leq c_n) &= \sum_{x_1 \in T_1, x_1 \leq c_1} \cdots \sum_{x_n \in T_n, x_n \leq c_n} \mathbb{P}(X_1 = x_1, \dots, X_n = x_n) \\ &= \sum_{x_1 \leq c_1} \cdots \sum_{x_n \leq c_n} \prod_{i=1}^{n} \mathbb{P}(X_i = x_i) \\ &= \left(\sum_{x_1 \leq c_1} \mathbb{P}(X_1 = x_1) \right) \cdots \left(\sum_{x_n \leq c_n} \mathbb{P}(X_n = x_n) \right) \\ &= \mathbb{P}(X_1 \leq c_1) \cdots \mathbb{P}(X_n \leq c_n) \end{aligned}$$

X stetig:
Weise die Unabhängigkeit der Mengensysteme $\{\omega \in \Omega \,|\, X_i(\omega) \leq c, \, c \in \mathbb{R}\}$ nach und nutze $\{] - \infty, c] | c \in \mathbb{R}\}$ als durchschnittsstabilen Erzeuger von $\mathcal{B}$. Details s. Satz 5.9/5.10, Meintrup and Schäffler [2005]. □

Satz 10.2
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:

<u>“X_i stu $\implies f_X = \prod f_{X_i}$”</u>:

Sei F_X Verteilungsfunktion von X, $c \in \mathbb{R}^n$. Nach Def.:

$$\begin{aligned}
F_X(c) &= \mathbb{P}_X(I_c) = \int_{I_c} f d\lambda^n \quad \text{für } I_c := (-\infty, c_1] \times \ldots \times (-\infty, c_n] \\
&= \mathbb{P}(X_1 \leq c_1, \ldots, X_n \leq c_n) \overset{\text{Satz 10.1}}{=} \prod_{i=1}^n \mathbb{P}(X_i \leq c_i) = \prod_{i=1}^n \int\limits_{(-\infty,c_i]} f_{X_i} d\lambda \\
&\overset{\text{Fubini}}{=} \int\limits_{(-\infty,c_1]} \cdots \int\limits_{(-\infty,c_n]} \prod_{i=1}^n f_{X_i} d\lambda \cdots d\lambda \overset{\text{Fubini}}{=} \int_{I_c} \prod_{i=1}^n f_{X_i} d\lambda^n \\
\implies f &= \prod_{i=1}^n f_{X_i} \ \lambda^n - \text{f.ü.}, \quad \text{(Satz 6.2 Radon-Nikodym)}
\end{aligned}$$

d.h. unter Unabhängigkeit ist die Dichte f_X von X bzgl. λ^n das Produkt der Einzeldichten.

Darstellung als Riemann-Integral:

$$\begin{aligned}
F_X(c) = \mathbb{P}(X_1 \leq c_1, \ldots, X_n \leq c_n) &= \prod_{i=1}^n \mathbb{P}(X_i \leq c_i) \\
&= \int_{-\infty}^{c_1} f_{X_1}(x_1)\, dx_1 \int_{-\infty}^{c_2} f_{X_2}(x_2)\, dx_2 \cdots \int_{-\infty}^{c_n} f_{X_n}(x_n)\, dx_n \\
&= \int_{-\infty}^{c_1} \cdots \int_{-\infty}^{c_n} f_{X_1}(x_1) \cdots f_{X_n}(x_n)\, dx_n \cdots dx_1 \\
&= \underbrace{\int_{-\infty}^{c} f_X(x)\, dx}_{n\text{-dimensionales Integral}}
\end{aligned}$$

<u>“$f_X = \prod f_{X_i} \implies X_i$ stu”</u>:

$$
\begin{aligned}
\mathbb{P}(X_1 \le c_1, \ldots, X_n \le c_n) = F_X(c) &= \int_{I_c} \prod_{i=1}^{n} f_{X_i} d\lambda^n \qquad \text{(nach Voraussetzung)} \\
&= \prod_{i=1}^{n} \int\limits_{(-\infty, c_i]} f_{X_i} d\lambda \qquad \text{(Fubini)} \\
&= \prod_{i=1}^{n} \mathbb{P}(X_i \le c_i)
\end{aligned}
$$

□

Satz 10.3
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.

$$\Longrightarrow Y_1 = h_1 \circ X_1 \textit{ und } Y_2 = h_2 \circ X_2 \textit{ stu.}$$

Beweis:

Nach Voraussetzung $\sigma(X_1) = X_1^{-1}(\mathcal{B}^{n_1})$ und $\sigma(X_2) = X_2^{-1}(\mathcal{B}^{n_2})$ stu.
z.Z.: $Y_1^{-1}(\mathcal{B}^{m_1})$ und $Y_2^{-1}(\mathcal{B}^{m_2})$ stu. h_i Baire-Funktion, also $\mathcal{B}^{n_i}-\mathcal{B}^{m_i}$-messbar $\Longrightarrow h_i^{-1}(\mathcal{B}^{m_i}) \subset \mathcal{B}^{n_i}$:

$$
\begin{aligned}
\Longrightarrow Y_i^{-1}(\mathcal{B}^{m_i}) &= (h_i \circ X_i)^{-1}(\mathcal{B}^{m_i})) \\
&= X_i^{-1}(h_i^{-1}(\mathcal{B}^{m_i})) \\
&\subset X_i^{-1}(\mathcal{B}^{n_i})
\end{aligned}
$$

$\Longrightarrow Y_1^{-1}(\mathcal{B}^{m_1}),\ Y_2^{-1}(\mathcal{B}^{m_2})$ stu weil $X_1^{-1}(\mathcal{B}^{n_1}),\ X_2^{-1}(\mathcal{B}^{n_2})$ stu. □

Satz 10.4 (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 $n = 2$

$$
\begin{aligned}
&\text{Sei } f_X = f_{X_1} \cdot f_{X_2} \text{ (stu) und} \\
&\quad \mathbb{P}_X = f_X \odot (\mu_1 \otimes \mu_2) = f_{X_1} \cdot f_{X_2} \odot (\mu_1 \otimes \mu_2)
\end{aligned}
$$

Betrachte $g : \mathbb{R}^2 \to \mathbb{R} \quad x \mapsto g(x) = x_1 \cdot x_2$

$$\begin{aligned}
\mathbb{E}(g \circ X) &= \int g \, d\mathbb{P}_X \\
&= \int g \, d(f_X \odot (\mu_1 \otimes \mu_2)) \\
&= \int g \cdot (f_{X_1} \cdot f_{X_2}) \, d(\mu_1 \otimes \mu_2) \qquad \text{(stu und 6.1)} \\
&= \int \int x_1 \, x_2 \, f_{X_1}(x_1) \cdot f_{X_2}(x_2) \, d\mu_1(x_1) \, d\mu_2(x_2) \\
&= \int x_1 \, f_{X_1}(x_1) d\mu_1(x_1) \int x_2 \, f_{X_2}(x_2) d\mu_2(x_2) \\
&= \mathbb{E}(X_1) \cdot \mathbb{E}(X_2)
\end{aligned}$$

□

Satz 10.5
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 10.3 sind $\exp(s \, X_1), \ldots, \exp(s \, X_n)$ als Komposition stochastisch unabhängiger ZV wieder stu und, nach Voraussetzung, auf $]-a, a[$ integrierbar. $\implies$ $\exp\left(s \cdot \sum_{i=1}^n X_i\right) = \exp(s \, X_1) \cdot \ldots \cdot \exp(s \, X_n)$ ist integrierbar $\implies M(s) = \mathbb{E}(\exp(s \, S_n)) \overset{\text{stu}}{=} \prod_{i=1}^n \mathbb{E}(\exp(s \cdot X_i)) = \prod_{i=1}^n M_i(s)$ □

Satz 10.6
Sind X und Y stochastisch unabhängig und seien φ_X und φ_Y ihre charakteristischen Funktionen. Dann gilt:

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

Beweis:

Wegen Unabhängigkeit gilt:

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

$\square$

11 Transformationssatz für Dichten

Satz 11.1 (Transformationssatz für surjektive Funktionen)
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 $\mathbb{R} \supset I = \dot{\bigcup}_{m \in M} I_m$ eine Vereinigung disjunkter offener Intervalle $I_m =]a_m, b_m[$ derart, daß gilt:

a) F_X ist stetig differenzierbar auf I mit $f_X(x) = \frac{\partial F_X(x)}{\partial x}$,

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

c) für die stetig differenzierbare und bijektive Einschränkung von g auf I_m $g_m = g|_{I_m}$; $g_m : I_m \to \mathbb{R}$ ist $\frac{\partial g_m(x)}{\partial x} \neq 0$.

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

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

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

$$V_m \text{ ist Stammfunktion von } v_m \iff v_m \text{ ist Dichte von } V_m$$

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)

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

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

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

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

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

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

$\square$

Beispiel 11.1
$X \sim N(0,1)$. *Gesucht:* f_{X^2}
$g(x) = x^2 \quad I_1 =]-\infty, 0[\quad I_2 =]0, \infty[$
$g_1 = g|_{I_1} \implies h_1(y) = -\sqrt{y} \quad g_2 = g|_{I_2} \implies h_2(y) = \sqrt{y}$
$\left|\frac{\partial h_i(y)}{\partial y}\right| = \frac{1}{2\sqrt{y}} \quad i = 1,2$

Fallunterscheidung:

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

Definition 11.1
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: $\quad X \sim \chi^2\,(k)$.

Satz 11.2 (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 $H_m(y) = \left(\frac{\partial h_{im}(y)}{\partial y_j}\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$ *Jacobi-Matrix für* $h_m = (h_{1m}, \ldots, h_{nm}) = g_m^{-1}$.
Dann gilt

$$f_{g \circ X}(y) = \sum_{m \in M} v_m(y) \tag{2}$$

$$\textit{mit } v_m(y) = \begin{cases} f_X(h_m(y)) \cdot |\det\left(H_m(y)\right)| & \textit{falls } h_m(y) \in G_m, \\ 0 & \textit{sonst.} \end{cases} \tag{3}$$

Beweis:

geschenkt. □

Beispiel 11.2
Sei $X = (X_1, X_2)$ eine zwei-dimensionale ZV mit Dichte f_X bzgl. λ^2. Gesucht ist $f_{X_1+X_2}$. Sei $g(x_1, x_2) = (X_1 + X_2, X_2) \in \mathbb{R}^2$.

$\Longrightarrow$ *g bijektiv und damit*

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

$$\overset{\text{Satz 11.2}}{\Longrightarrow} f_{Y,X_2}(y, x_2) = f_X(y - x_2, x_2) \cdot 1$$

$$\Longrightarrow f_Y(y) = \int f_{Y,X_2}(y, x_2)\, dx_2 \qquad \text{(Bem. zu Def. 7.7)}$$

$$= \int f_X(y - x_2, x_2)\, dx_2 \qquad \text{(Faltung)}$$

Falls X_1 und X_2 stu:

$$f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \qquad \text{(Satz 10.2)}$$

und damit

$$f_Y(y) = \int f_{X_1}(y - x_2) \cdot f_{X_2}(x_2)\, dx_2$$

Beispiel 11.3
Seien $X_1 \sim N(0,1),\ X_2 \sim N(0,1)$ stu.
Definiere $Y_1 := \frac{X_2}{X_1}$, $Y_2 := X_1$, also $g(x_1, x_2) = \left(\frac{x_2}{x_1}, x_1\right)$.
Gesucht: Dichte von Y_1.

$h := g^{-1}, h(y_1, y_2) = (y_2, y_1\, y_2)$, *also* $\det H = \det\begin{pmatrix} 0 & 1 \\ y_2 & y_1 \end{pmatrix} = -y_2$

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1}(y_2)\, f_{X_2}(y_1\, y_2)\, |\det H| \qquad (X_1, X_2 \text{ stu}) \\ &= \frac{|-y_2|}{2\pi} \exp\left(-\frac{1}{2}\left(y_2^2 + (y_1\, y_2)^2\right)\right) \\ &= \frac{|y_2|}{2\pi} \exp\left(-\frac{1}{2}\left(y_2^2 \cdot (1 + y_1^2)\right)\right) \end{aligned}$$

Randverteilung von Y_1:

$$\begin{aligned}
f_{Y_1}(y_1) &= \int_{-\infty}^{\infty} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \\
&= 2 \int_{0}^{\infty} \frac{y_2}{2\pi} \exp\left(-\frac{1}{2}\, y_2^2 (1+y_1^2)\right)\, dy_2 \qquad\qquad \textit{(Symmetrie)} \\
&= \frac{1}{\pi} \int_{0}^{\infty} y_2 \exp\left(-\frac{1}{2}\, y_2^2 (1+y_1^2)\right)\, dy_2 \\
&= \frac{1}{\pi} \left[\exp\left(-\frac{1}{2}\, y_2^2\, (1+y_1^2)\right) \frac{-1}{1+y_1^2}\right]_{y_2=0}^{\infty} = \frac{-1}{\pi(1+y_1^2)}\,(0-1) \\
&= \frac{1}{\pi\,(1+y_1^2)} \quad \cong \quad \textit{Cauchy-Verteilung}
\end{aligned}$$

Definition 11.2
Eine Zufallsvariable X mit Dichte

$$f(x) = \frac{1}{\pi(1+x^2)}$$

heißt Cauchy-verteilt.

12 Mehrdimensionale Momente

Definition 12.1
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.

$$\begin{aligned}
Cov(Y, Z) &:= \mathbb{V}((Y, Z))_{1,2} \\
&= \mathbb{V}((Z, Y))_{2,1} \\
&= \mathbb{E}((Y - \mathbb{E}(Y))(Z - \mathbb{E}(Z)))
\end{aligned}$$

heisst Kovarianz von zwei Zufallsvariablen $Y : \Omega \to \mathbb{R}^1$, $Z : \Omega \to \mathbb{R}^1$.

Bemerkung:

$$X \in \mathbb{R}^{n\times 1} \implies XX^\top \in \mathbb{R}^{n\times n}$$

Satz 12.1
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. $x^\top\,\mathbb{V}(X)\,x \geq 0 \quad \forall x \in \mathbb{R}^n$

$$\begin{aligned}
x^\top\,\mathbb{V}(X)\,x &= x^\top\,\mathbb{E}((X-\mathbb{E}(X))\,(X-\mathbb{E}(X))^\top)\,x \\
&= \mathbb{E}(\underbrace{(x^\top X - x^\top\,\mathbb{E}(X))}_{\in\mathbb{R}}\,(x^\top X - x^\top\,\mathbb{E}(X))^\top) \\
&= \mathbb{E}((x^\top X - x^\top\,\mathbb{E}(X))^2) \geq 0
\end{aligned}$$

□

Satz 12.2
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* $\implies$ $Cov(X,Y) = 0$

iii)

$$\mathbb{V}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$
$$\overset{X_1,\dots,X_n \ stu}{=} \sum_{i=1}^{n} \mathbb{V}(X_i)$$

Beweis:

i)

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}\left((X - \mathbb{E}(X))\,(Y - \mathbb{E}(Y))\right) \\
&= \mathbb{E}(XY) - \mathbb{E}(Y)\,\mathbb{E}(X) - \mathbb{E}(X)\,\mathbb{E}(Y) + \mathbb{E}(X)\,\mathbb{E}(Y) \\
&= \mathbb{E}(XY) - \mathbb{E}(X)\,\mathbb{E}(Y)
\end{aligned}$$

ii) X, Y stu $\implies$ $\mathbb{E}(X\,Y) \overset{\text{Satz 10.4}}{=} \mathbb{E}(X)\,\mathbb{E}(Y)$

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 12.2 (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 12.3

$$|\rho(X,Y)| \leq 1$$

Beweis:

$$\begin{aligned}|\mathrm{Cov}(X,Y)| &\leq \mathbb{E}(|(X-\mathbb{E}(X))\cdot(Y-\mathbb{E}(Y))|)\\ &\leq \mathbb{E}((X-\mathbb{E}(X))^2)^{\frac{1}{2}}\cdot \mathbb{E}((Y-\mathbb{E}(Y))^2)^{\frac{1}{2}} \quad \text{(Hölder-U., Satz 8.4)}\\ &= \mathbb{V}(X)^{\frac{1}{2}}\cdot \mathbb{V}(Y)^{\frac{1}{2}} = \sqrt{\mathbb{V}(X)\cdot\mathbb{V}(Y)}\end{aligned}$$

□

Satz 12.4 (Standardisierung)
Sei $X = (X_1, \dots, X_n)$ n-dimensionale ZV mit $\mathbb{E}(X) = \mu$ und positiv definiter Kovarianzmatrix $\mathbb{V}(X) = \Sigma$.
Dann existiert eine invertierbare Matrix $B \in \mathbb{R}^{n\times n}$ so dass gilt:

$$\textit{Für } Y = B^{-1}(X-\mu) \textit{ ist } \mathbb{E}(Y) = 0_n,\ \mathbb{V}(Y) = 1_n$$

Bemerkung:

Für $n = 1$ gilt: ($\mathbb{V}(X) = \sigma^2$) $Y = \dfrac{X-\mu}{\sigma}$ hat $\mathbb{E}(Y) = 0$ und $\mathbb{V}(Y) = 1$.

Beweis:

Σ ist nach Voraussetzung positiv definit. Damit existiert (z.B.) die Cholesky-Zerlegung

$$\Sigma = B\,B^\top \quad \text{mit } B \in \mathbb{R}^{n\times n},\ B^{-1} \text{ existiert.}$$

Also

$$\begin{aligned}\mathbb{E}(Y) &= \mathbb{E}(B^{-1}(X-\mu)) = B^{-1}(\mathbb{E}(X)-\mu)\\ &= B^{-1}0_n = 0_n\\ \mathbb{V}(Y) &= B^{-1}\,\mathbb{V}(X)\,{B^{-1}}^\top = B^{-1}\,B\,B^\top\,{B^{-1}}^\top\\ &= 1_n\,(B^{-1}B)^\top = 1_n\,1_n = 1_n\end{aligned}$$

□

Definition 12.3
Sei Σ die Kovarianzmatrix einer mehrdimensionalen Zufallsvariablen $X = (X_1, \ldots, X_n)$ und $diag(\Sigma)$ die Matrix der Diagonalelemente von Σ.
Dann heißt

$$R = Corr(X) := (diag(\Sigma))^{-1/2}\Sigma(diag(\Sigma))^{-1/2}$$

die <u>Korrelationsmatrix</u> von X. Es gilt:

- $r_{i,i} = 1$ *für alle* $i = 1, \ldots, n$
- $r_{i,j} = \rho(X_i, X_j)$

Definition 12.4

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$$

heißt <u>Beta-Funktion</u>. Sie kann als Verstetigung des Binomialkoeffizienten interpretiert werden: für $\alpha, \beta \in \mathbb{N}$ gilt:

$$B(\alpha, \beta) = \left((\alpha+\beta-1)\binom{\alpha+\beta-2}{\beta-1}\right)^{-1}.$$

Definition 12.5 (Beta-Verteilung)
Eine Zufallsvariable X mit stetiger Dichte

$$f_X(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \cdot I_{(0,1)}(x)$$

mit $\alpha > 0$ und $\beta > 0$ heißt <u>Beta-verteilt</u>: $\quad X \sim Beta(\alpha, \beta)$.

Definition 12.6 (Dirichlet-Verteilung)
Mit $x = (x_1, \ldots, x_n)$ mit $\sum_{i=1}^n x_i = 1$ und $x_i \geq 0$ für alle $i = 1, \ldots, n$,

$$f(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_n) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^n x_i^{\alpha_i - 1}$$

eine stetige Dichte bezüglich λ^{n-1}. Die Verteilung mit Dichte f heißt <u>Dirichlet-Verteilung</u> $Diri(\alpha_1, \ldots, \alpha_n)$.

Bemerkung:

Der Träger der Dirichlet-Verteilung ist ein $n-1$-Simplex.

Beispiel 12.1
Sei $X \sim Diri(\alpha_1, \ldots, \alpha_n)$ und $\alpha_0 = \sum_{i=1}^n \alpha_i$. Dann gilt:

- *Für $n = 2$ gilt $X \sim Beta(\alpha_1, \alpha_2)$,*
- $\mathbb{E}(X) = \left(\frac{\alpha_1}{\alpha_0}, \ldots, \frac{\alpha_n}{\alpha_0}\right)$,
- $\mathbb{V}(X_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0+1)}$,
- $Cov(X_i, X_j) = \frac{-\alpha_i \alpha_j}{\alpha_0^2(\alpha_0+1)}$,
- $X_i \sim Beta\ (\alpha_i, \alpha_0 - \alpha_i)$.

13 Die Normalverteilung

13.1 Eindimensionale Normalverteilung

$X \sim \mathrm{N}(\mu, \sigma^2)$
Bekannt:

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

$$\begin{aligned} \mathbb{E}(X) &= \mu \\ \mathbb{V}(X) &= \sigma^2 \end{aligned}$$

Satz 13.1 (Lineare Transformation)

$$X \sim N(\mu, \sigma^2) \implies Y = aX + b \sim N(a\mu + b, a^2\sigma^2) \quad \textit{für } a, b \in \mathbb{R}, a \neq 0.$$

Beweis:

$$
\begin{aligned}
y = g(x) = ax + b \implies g^{-1}(y) &= \frac{y-b}{a} \text{ bijektiv}, \frac{\partial g^{-1}(y)}{\partial y} = \frac{1}{a}.\\
\overset{\text{Satz 11.1}}{\implies} f_Y(y) &= f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{|a|}\\
&= \frac{1}{\sqrt{2\pi}\,\sigma\,|a|} \exp\left(-\frac{1}{2}\left(\frac{\frac{y-b}{a} - \mu}{\sigma}\right)^2\right)\\
&= \frac{1}{\sqrt{2\pi\sigma^2 a^2}} \exp\left(-\frac{1}{2}\left(\frac{y-(a\mu+b)}{a\,\sigma}\right)^2\right)\\
\implies \quad Y \quad &\sim \mathrm{N}(a\mu + b, a^2\,\sigma^2)
\end{aligned}
$$

□

Korollar 13.1

$$
X \sim N(\mu, \sigma^2) \implies Y = \frac{X-\mu}{\sigma} \sim N(0,1).
$$

Satz 13.2

$$
\begin{aligned}
&X \sim N(\mu, \sigma^2) \implies\\
&\qquad m_r^0(X) = \begin{cases} \left(\prod_{i=1}^{\frac{r}{2}} (1 + 2(i-1))\right) \sigma^r & r = 2k, k \in \mathbb{N}\\ 0 & sonst \end{cases}
\end{aligned}
$$

Bemerkung:

$\implies m_4^0(X) = 3\sigma^4 \implies K(X) = \frac{m_4^0(X)}{(m_2^0(X))^2} = 3$

Beweis:

$r = 2k+1 \overset{\text{Symmetrie Satz 8.1}}{\implies} m_r^0(X) = 0$

Sei $Z \sim \mathrm{N}(0,1)$. Dann gilt

$$
\frac{\partial \varphi(z)}{\partial z} = -z\,\varphi(z) \quad \text{und}
$$

$$\begin{aligned}
\mathbb{E}(Z^r) &= \int \underbrace{z^{r-1}}_{v} \cdot \underbrace{z \cdot \varphi(z)}_{u'} \, dx \\
&\overset{PI}{=} \left[-z^{r-1}\varphi(z)\right]_{-\infty}^{\infty} - \int -\varphi(x)(r-1)z^{r-2}\, dz \\
&= 0 + (r-1) \underbrace{\int z^{r-2}\, \varphi(z)\, dz}_{\mathbb{E}(Z^{r-2})}.
\end{aligned}$$

$\implies \mathbb{E}(Z^r) = (r-1)\mathbb{E}(Z^{r-2})$
Wegen $\mathbb{E}(Z) = 0$ folgt $\mathbb{E}(Z^r) = 0 \, \forall \, r$ ungerade.
Wegen $\mathbb{E}(Z^2) = \mathbb{V}(Z) = 1$ folgt für r gerade, $r = 2k$:

$$\begin{aligned}
\mathbb{E}(Z^r) &= (r-1)\, \mathbb{E}(Z^{r-2}) \\
&\overset{r=2k}{=} (2k-1)\, \mathbb{E}(Z^{2\,(k-1)}) \\
&= (2k-1)\,(2k-3)\, \mathbb{E}(Z^{2\,(k-2)}) \ldots \\
&= (2k-1)\,(2k-3) \cdots (3)\, \underbrace{\mathbb{E}(Z^2)}_{=1} \\
&= 1 \cdot 3 \cdot 5 \cdots (2k-1).
\end{aligned}$$

Da $\mathbb{E}((X-\mu)^r) = \mathbb{E}((\sigma\, Z)^r) = \sigma^r\, \mathbb{E}(Z^r)$ folgt die Behauptung. □

Satz 13.3 (Additivität der Normalverteilung)
Seien $X_1 \sim N(\mu_1, \sigma_1^2)$ und $X_2 \sim N(\mu_2, \sigma_2^2)$ stu

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

Beweis:

$$X_1 + X_2 \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \iff \frac{X_1 + X_2 - (\mu_1 + \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim \mathrm{N}(0,1)$$

Weiterhin:

$$\frac{X_1 + X_2 - (\mu_1 + \mu_2)}{\sqrt{\sigma_1^2 + \sigma_2^2}} = \underbrace{\frac{X_1 - \mu_1}{\sqrt{\sigma_1^2 + \sigma_2^2}}}_{\widetilde{X}_1} + \underbrace{\frac{X_2 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}}_{\widetilde{X}_2} \sim \mathrm{N}(0,1)$$

$$
\begin{aligned}
\mathbb{E}(\widetilde{X}_1) &= \mathbb{E}(\widetilde{X}_2) = 0 \\
\mathbb{V}(\widetilde{X}_1) &= \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} := \lambda^2 \\
\mathbb{V}(\widetilde{X}_2) &= \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} = 1-\lambda^2
\end{aligned}
$$

Es reicht also zu zeigen:

$$\widetilde{X}_1 \sim \mathrm{N}(0,\lambda^2) \text{ und } \widetilde{X}_2 \sim \mathrm{N}(0,1-\lambda^2) \text{ stu} \implies \widetilde{X}_1 + \widetilde{X}_2 \sim \mathrm{N}(0,1)$$

$$
\begin{aligned}
f_{\widetilde{X}_1+\widetilde{X}_2}(z) &\overset{\text{Faltung}}{=} \int f_{\widetilde{X}_1}(z-x)\, f_{\widetilde{X}_2}(x)\, dx \\
&= \int \frac{1}{\sqrt{2\pi\lambda^2}} \exp\left(-\frac{1}{2}\left(\frac{z-x}{\lambda}\right)^2\right) \frac{1}{\sqrt{2\pi(1-\lambda^2)}} \exp\left(-\frac{1}{2}\,\frac{x^2}{1-\lambda^2}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\, z^2\right) \cdot \\
&\quad \underbrace{\int \frac{1}{\sqrt{2\pi}}\, \frac{1}{\sqrt{\lambda^2\,(1-\lambda^2)}} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{(z-x)^2}{\lambda^2} + \frac{x^2}{1-\lambda^2} - z^2\right)}_{A}\right) dx}_{B}
\end{aligned}
$$

$$
\begin{aligned}
A &= \frac{(1-\lambda^2)\, z^2 - 2\,(1-\lambda^2) z\, x + (1-\lambda^2)\, x^2 + \lambda^2\, x^2 - \lambda^2\,(1-\lambda^2)\, z^2}{\lambda^2\,(1-\lambda^2)} \\
&= \frac{x^2 - 2\,(1-\lambda^2)\, z\, x + [(1-\lambda^2)\, z]^2}{\lambda^2\,(1-\lambda^2)} = \frac{(x-(1-\lambda^2)\, z)^2}{\lambda^2\,(1-\lambda^2)}
\end{aligned}
$$

$\implies B(x)$ ist Dichte von $\mathrm{N}((1-\lambda^2)\, z, \lambda^2\,(1-\lambda^2)) \implies \int B(x)\, dx = 1$ $\square$

Korollar 13.2
$X_1, \ldots, X_n$ 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) $\mu_1 = \ldots = \mu_n =: \mu, \sigma_1^2 = \ldots = \sigma_n^2 =: \sigma^2 \implies \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

Korollar 13.3
$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ stu. $\implies$ $aX_1 + bX_2 + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$

Bemerkung:

Es gilt auch die Umkehrung: Sind X und Y stu und $X+Y$ normalverteilt, so sind auch X und Y normalverteilt.

13.2 k-dimensionale Normalverteilung

Definition 13.1 (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

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^k \exp\left(-\frac{1}{2}\sum_{i=1}^{k} x_i^2\right) \qquad x = (x_1, \ldots, x_k) \in \mathbb{R}^k$$

besitzt: $X \sim N_k(0, I_k)$.

Satz 13.4

$$X \sim N_k(0, I_k) \qquad \iff X_1, \ldots, X_k \text{ stu}, X_i \sim N(0,1)\ \forall i = 1, \ldots, k.$$

Beweis:

$$\begin{aligned} X_1, \ldots, X_k \text{ stu} \overset{10.2}{\iff} f_X(x) &= \prod_{i=1}^{k} f_{X_i}(x_i) \\ &\overset{X_i \sim \mathrm{N}(0,1)}{=} \prod_{i=1}^{k} \varphi(x_i) \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^k \exp\left(-\frac{1}{2}\sum x_i^2\right) \end{aligned}$$

□

Definition 13.2
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 13.5

$$Y \sim N_p(\mu, \Sigma) \quad \Longrightarrow \quad \begin{aligned} \mathbb{E}(Y) &= \mu \\ \mathbb{V}(Y) &= \Sigma \end{aligned}$$

Beweis:

Satz 12.1 □

Satz 13.6
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 11.2 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

$$f_Y(y) = \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\left(-\frac{(y-\mu)^\top \Sigma^{-1} (y-\mu)}{2}\right).$$

$\square$

Bemerkung:

Ist $\Sigma = A^T A$ nicht invertierbar mit Rang $p < k$, hat $Y = AX + \mu$ keine Dichte bezüglich λ^k. Es lässt sich allerdings eine Restriktion des Lebesgue-Maßes auf den p-dimensionalen Unterraum von $\mathbb{R}^k$ konstruieren, dann hat Y die Dichte

$$f(x) = (2\pi)^{-\frac{p}{2}} \left(\det{}^*(\Sigma)\right)^{-\frac{1}{2}} \, exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^+ (x-\mu)\right)$$

mit Σ^+ eine verallgemeinerte Inverse (z.B. Moore–Penrose) und $\det^*$ die verallgemeinerte Determinante (Produkt der Eigenwerte ungleich Null).

Definition 13.3 (Kanonische Form der mehrdimensonalen Normalverteilung)
Sei $X \sim N_k(\mu, \Sigma)$. Dann heißen $Q = \Sigma^{-1}$ und $m = \Sigma^{-1}\mu$ die kanonischen Parameter der Normalverteilung. Die Dichte läßt sich schreiben als

$$f(x) = \frac{1}{\sqrt{(2\pi)^k}} \det(Q)^{1/2} \exp\left(-\frac{1}{2}(x^T Q x - 2x^T m - m^T Q^{-1} m)\right).$$

Satz 13.7

a) $Y \sim N_k(\mu, \Sigma) \implies Y_i \sim N(\mu_i, \Sigma_{ii}) \quad \forall i = 1, \ldots, k$

b) $Y \sim N_k(\mu, \Sigma) \iff v^\top Y \sim N_1(v^\top \mu, v^\top \Sigma v) \quad \forall v \in \mathbb{R}^k$

c) $(Y_1, Y_2) \sim N_2(\mu, \Sigma)$ *mit* $\mu = (\mu_1, \mu_2)$ *und* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$

$$Y_1, Y_2 \ stu \iff \Sigma_{12} = \Sigma_{21} = 0$$

Beweis:

a), b) Meintrup and Schäffler [2005] Seite 197ff

c) Übung.

$\square$

14 Copulas

Definition 14.1
Eine Copula $C : [0,1]^n \to [0,1]$ *ist eine* n*-dimensionale Verteilungsfunktion*

$$C(u_1, \dots, u_n) := \mathbb{P}(U_1 \le u_1, \dots, U_n \le u_n)$$

über $[0,1]^n$ *deren* n *Randverteilungen alle Standard-Gleichverteilungen* $U[0,1]$ *sind.*

Bemerkung:

- $C(u_1, \dots, u_n)$ ist monoton ansteigend in jeder Komponente u_j
- Die Randverteilung der j-ten Komponente bekommt man mit $u_i = 1 \, \forall \, i \neq j$, und da sie eine Gleichverteilung ist muss gelten:

$$C(1, \dots, 1, u_j, 1, \dots, 1) = u_j$$

Satz 14.1 (Verteilung durch Copula und Randverteilungen)
Sei $C(u_1, \dots, u_n)$ *eine Copula, und* $F_1, \dots F_n$ *beliebige eindimensionale Verteilungsfunktionen. Dann ist*

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \tag{4}$$

eine Verteilungsfunktion mit Randverteilungsfunktionen $F_1, \dots, F_n$*.*

Bemerkung:

$P(F(X) \le u) = u, u \in [0,1]$ für stetige Verteilungsfunktionen F, also $F(X) \sim U[0,1]$. Damit gilt auch:

$$C(u_1, \dots, u_n) = F\left(F_{X_1}^{-1}(u_1), \dots, F_{X_n}^{-1}(u_n)\right).$$

Satz 14.2 (Satz von Sklar)
Sei F *eine* n*-dimensionale Verteilungsfunktion mit den Randverteilungen* $F_1, \dots, F_n$*. Dann existiert eine Copula* C*, so dass Satz 14.1 gilt. Sind alle Randverteilungen stetig, so ist* C *eindeutig.*

Beweis:

Siehe Sklar [1997]. □

Bemerkung:

Das heißt, dass jede stetige multivariate Verteilung in ihre Randverteilungen (marginale Verteilungsfunktionen) und ihre Abhängigkeitsstruktur (Copula) zerlegt werden kann.

Satz 14.3 (Copula unter monotonen Transformationen)
Seien $X_1, \ldots, X_n$ rellwertige Zufallsvariablen mit stetigen Randverteilungen und einer Copula C_X. Seien $g_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$ strikt monoton steigende Funktionen und $Y_i = g_i(X_i)$. Dann gilt für die Copula C_Y von $Y_1, \ldots, Y_n$:

$$C_Y = C_X.$$

Beweis:

$$\begin{aligned} F_Y(y_1, \ldots, y_n) &= P\left(g_1(X_1) \leq y_1, \ldots, g_n(X_n) \leq y_n\right) \\ &= P\left(X_1 \leq g_1^{-1}(y_1), \ldots, X_n \leq g_n^{-1}(y_n)\right) \\ &= F_X(g_1^{-1}(y_1), \ldots, g_n^{-1}(y_n)) \end{aligned}$$

also mit $y_i \to \infty \, \forall \, i \neq j$: $F_{Y_j}(y_j) = F_{X_j}(g_1^{-1}(y_j))$, und damit:

$$\begin{aligned} C_Y(u_1, \ldots, u_n) &= F_Y\left(F_{Y_1}^{-1}(u_1), \ldots, F_{Y_n}^{-1}(u_n)\right) \\ &= F_X\left(g_1^{-1}\left(F_{Y_1}^{-1}(u_1)\right), \ldots, g_n^{-1}\left(F_{Y_n}^{-1}(u_n)\right)\right) \\ &= F_X\left(F_{X_1}^{-1}(u_1), \ldots, F_{X_n}^{-1}(u_n)\right) = C_X(u_1, \ldots, u_n) \end{aligned}$$

□

Satz 14.4 (Fréchet-Hoeffding Schranken)
Für eine gültige Copula $C(u_1, \ldots, u_n)$ muss gelten:

$$\max\left(1 - n + \sum_{i=1}^{n} u_i, 0\right) \leq C(u_1, \ldots, u_n) \leq \min(u_1, \ldots, u_n)$$

Bemerkung:

- Die obere Schranke bezeichnet man als ko-monotone Copula. Sie entsteht zum Beispiel wenn alle $X_i, i > 1$ monotone Transformationen von X_1 sind. ("vollständige gleichsinnige Abhängigkeit")
- Die untere Schranke ist für $n > 2$ keine gültige Copula. Für $n = 2$ entsteht sie durch die gemeinsame Verteilung von $(X, 1 - X)$ mit $X \sim U[0, 1]$. ("vollständige gegensinnige Abhängigkeit")

- Unter Unabhängigkeit ist die Copula $C(u_1, \ldots, u_n) = \prod_{i=1}^{n} u_i$.

Beispiel 14.1
Sei $X = (X_1, X_2) \sim N_2(0, R)$ *mit* $R = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$ *eine Korrelationsmatrix. Dann ist*

$$\begin{aligned} F_{X_1,X_2}(x_1, x_2) &= \Phi_R(x_1, x_2) \\ &= \Phi_R(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= C_R^{Ga}(u_1, u_2) \end{aligned}$$

Dabei ist Φ_R *die Verteilungsfunktion für* $N_2(0, R)$. C_R^{Ga} *heißt Gauss-Copula. Für* $\rho = -1, 0, 1$ *bekommt man entsprechend jeweils die Copula für vollständige gegensinnige Abhängigkeit, die Unabhängigkeitscopula, und die ko-monotone Copula.*

Definition 14.2 (Archimedische Copula)
Sei $\psi : [0, 1] \times \Theta \to \mathbb{R}_0^+$ *eine stetige, streng fallende und konvexe* <u>*Generator-Funktion*</u> *mit* $\psi(1; \theta) = 0$ *und Pseudo-Umkehrfunktion* $\psi^- = \begin{cases} \psi^{-1}(p, \theta) & 0 \le p \le \psi(0, \theta) \\ 0 & \textit{sonst} \end{cases}$.
Eine Copula heißt <u>*archimedische Copula*</u> *wenn sie sich darstellen läßt als*

$$C(u_1, \ldots, u_n) = \psi^- \left(\sum_{i=1}^{n} \psi(u_i; \theta); \theta \right)$$

Beispiel 14.2 (Archimedische Copulas)
Die <u>*Clayton-Copula*</u> *ist asymetrisch mit stärkerer Abhängigkeit für kleine u:* $\psi(u; \theta) = \frac{1}{\theta}(u^{-\theta} - 1);\ \theta \in [-1, \infty) \setminus \{0\}$ *ergibt*

$$C(u_1, u_2) = \left(max \left(u_1^{-\theta} + u_2^{-\theta} - 1, 0 \right) \right)^{-\frac{1}{\theta}}.$$

Die <u>*Gumbel-Copula*</u> *ist asymetrisch mit stärkerer Abhängigkeit für große u:* $\psi(u; \theta) = (-\log(u))^\theta;\ \theta \in [1, \infty)$ *ergibt*

$$C(u_1, u_2) = \exp\left(- \left((-\log(u_1))^\theta + (-\log(u_2))^\theta \right)^{1/\theta} \right).$$

Die <u>*Frank-Copula*</u> *ist symmetrisch:* $\psi(u; \theta) = -\log\left(\frac{\exp(-\theta u) - 1}{\exp(-\theta) - 1} \right);\ \theta \in \mathbb{R} \setminus \{0\}$ *ergibt*

$$C(u_1, u_2) = -\frac{1}{\theta} \log\left(1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right).$$

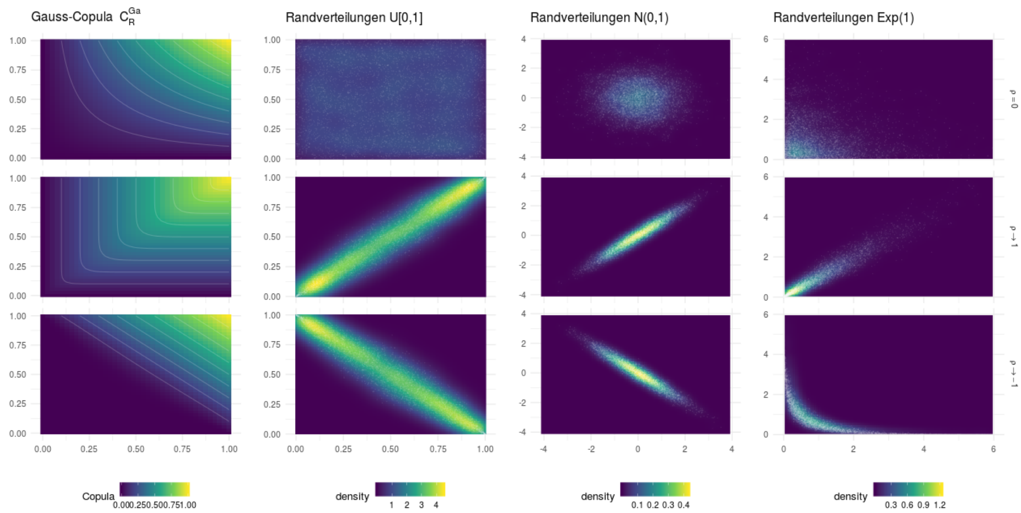


Abbildung 2: Gauss-Copula für extreme Parameter und Datenbeispiele mit verschiedenen Randverteilungen

15 Bedingte Verteilung

In Def. 2.11 haben wir die bedingte Wahrscheinlichkeit $\mathbb{P}(A|B)$ definiert. Die Definition überträgt 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 15.1 (Gedächtnislosigkeit der Exponentialverteilung)
$X \sim \mathit{Exp}(\lambda) \implies \mathbb{P}(X > t + s \,|\, X > t) = \mathbb{P}(X > s)$:
$\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = 1 - F_X(t) = 1 - (1 - \exp(-\lambda t)) = \exp(-\lambda t) \quad \forall t \geq 0$.

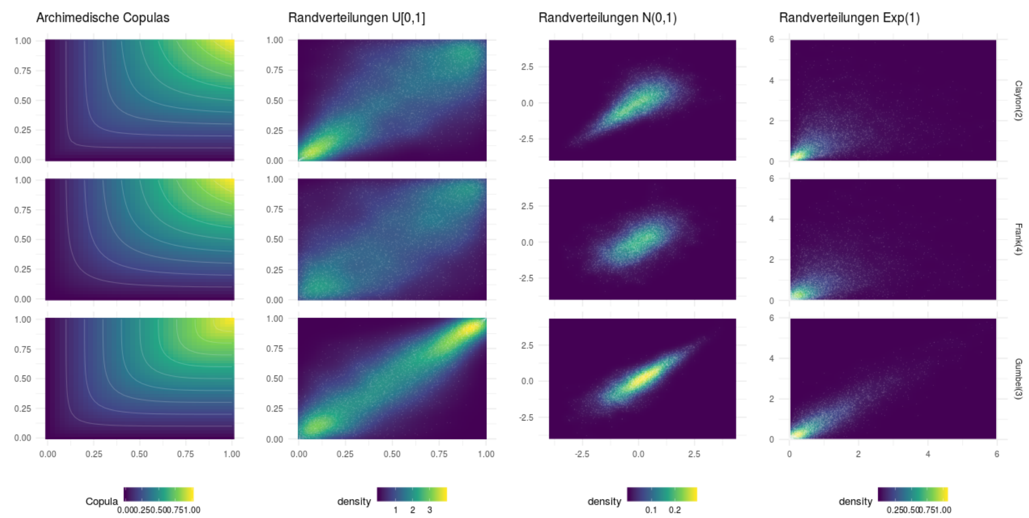


Abbildung 3: Archimedische Copulas und Datenbeispiele mit verschiedenen Randverteilungen

$\Longrightarrow \ \forall s, t \geq 0:$

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

Satz 15.1
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein W'keitsraum und $A_n \in \mathcal{F}, n \in \mathbb{N}$. Dann gilt

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

Beweis:

$\mathbb{P}(\bigcap_{i=1}^{n-1} A_i) > 0 \quad \Longrightarrow \quad \mathbb{P}(\bigcap_{i=1}^{j-1} A_i) > 0 \ \forall j \leq n-1$, 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. 2.11:

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

(IV) Für ein $n \in \mathbb{N}$ gelte

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

(IS) $n \to n+1$: Sei $\mathbb{P}\left(\bigcap\limits_{i=1}^{n} A_i\right) > 0$. 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 15.2 (Satz von der totalen Wahrscheinlichkeit)
Sei $(\Omega, \mathcal{F}, \mathbb{P})$ W'keitsraum. Sei $\dot{\bigcup\limits_{i=1}^{\infty}} A_i = \Omega$ 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:

$$A = \dot{\bigcup_{i=1}^{\infty}} A_i \cap A$$

$$\begin{aligned}\implies \mathbb{P}(A) = \mathbb{P}\left(\dot{\bigcup_{i=1}^{\infty}}(A_i \cap A)\right) &= \sum_{i=1}^{\infty}\mathbb{P}(A_i \cap A)\\ &= \sum_{i=1}^{\infty}\mathbb{P}(A_i)\cdot\mathbb{P}(A \mid A_i)\end{aligned}$$

□

Satz 15.3 (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

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

Beweis:

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

$\implies \mathbb{P}(A_i \,|\, B) = \dfrac{\mathbb{P}(B \,|\, A_i) \cdot \mathbb{P}(A_i)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(A_i \cap B)}{\underbrace{\mathbb{P}(B)}_{=\sum \mathbb{P}(A_i)\ldots}}$ (Def. 2.11) $\square$

Beispiel 15.2
Population von n Personen.
$(\Omega = \{1, \ldots, n\}, \mathcal{P}, U)$, *Gleichverteilung* $U(\{\omega\}) = \dfrac{1}{n}$
$K = \{1, \ldots, 0.02 \cdot n\}$ *Kranke* $G = \{0.02 \cdot n + 1, \ldots, n\}$ *Gesunde*
$\mathbb{P}(K) = 0.02$; $\mathbb{P}(G) = \mathbb{P}(\bar{K}) = 1 - 0.02 = 0.98$
$P =$ *"Test positiv", $N =$ "Test negativ"mit* $\mathbb{P}(P \,|\, K) = 0.95$
$\mathbb{P}(P \,|\, G) = 0.1$

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

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

Def. 2.11 lässt sich auch auf mehrdimensionale Räume anwenden.

Definition 15.1
Gegeben sei ein Produktmassraum $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ mit Verteilung $\mathbb{P}$ und Randverteilungen $\mathbb{P}_1$ und $\mathbb{P}_2$. Sei $A \in \Omega_1$, $B \in \Omega_2$ und $\mathbb{P}_2(B) > 0$. Dann ist

$$\begin{aligned} \mathbb{P}(\cdot \,|\, B) : \mathcal{F}_1 &\to [0,1] \\ A &\mapsto \mathbb{P}(A \,|\, B) = \frac{\mathbb{P}((A \times \Omega_2) \cap (\Omega_1 \times B))}{\mathbb{P}_2(B)} \end{aligned}$$

die bedingte Verteilung auf $(\Omega_1, \mathcal{F}_1)$ gegeben B.

Satz 15.4
Seien (X, Y) Zufallsvariablen mit gemeinsamer Verteilung, die von $\lambda \otimes \mu$

dominiert ist. Sei

$$f_{X|Y}(x|y) := \begin{cases} \frac{f(x,y)}{f_y(y)} & \textit{falls } f_y(y) > 0 \\ 0 & \textit{sonst} \end{cases}$$

Dann gilt:

$$\mathbb{P}(X \in B | Y = y) = \int_B f_{X|Y}(x|y) d\lambda$$

Definition 15.2
$f_{X|Y}(x|y)$ in Satz 15.4 heißt bedingte Dichte.

Beweis:

Es gilt für die Dichte $f(x,y)$

$$\mathbb{P}((x,y) \in A) = \int_A f(x,y) d(\lambda(x) \otimes \mu(y)), \quad A \in \mathcal{B}^2.$$

Dann ist $f_y(y) = \int_{\mathbb{R}} f(x,y) d\lambda(x)$ die (Rand-)Dichte von y bzgl. μ. Aus der Definition von bedingten Wahrscheinlichkeiten folgt:

$$\begin{aligned}
\mathbb{P}(X \in B, Y \in C) &= \mathbb{P}(X \in B | Y \in C) \mathbb{P}(Y \in C) \\
&= \int_C \mathbb{P}(X \in B | Y = y) d\mathbb{P}_Y(y) \\
&= \int_C \left(\int_B f_{X|Y}(x|y) d\lambda(x) \right) f_Y(y) d\mu(y) \\
&= \int_B \left(\int_C f_{X|Y}(x|y) f_Y(y) d\mu(y) \right) d\lambda(x) \\
&= \int_{B \times C} f(x,y) d(\lambda \otimes \mu) \\
&= \mathbb{P}((X,Y) \in B \times C) = \mathbb{P}(X \in B, Y \in C)
\end{aligned}$$

□

Satz 15.5 (Satz von Bayes für allgemeine Verteilungen)
Seien X und Y Zufallsvariablen mit Dichten f_X (bezüglich μ) und f_Y und gemeinsamer Dichte $f_{X,Y}$. Dann ist

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\int f_{Y|X}(y|x) f_X(x) d\mu(x)}$$

Beweis:

$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y) = f_{Y|X}(y|x) f_X(x)$
$f_Y(y) = \int f_{X,Y}(x,y) d\mu(x) = \int f_{Y|X}(y|x) f_X(x) d\mu(x)$ □

Thomas Bayes (1701 bis 7. April 1761)

Englischer Mathematiker, Philosoph und presbyterianischer Pfarrer. Der nach ihm benannte Satz wurde 1763 in einem posthum erschienenen Artikel bewiesen.

Beispiel 15.3 (Beispiel von Bayes)
Wir rollen eine weiße Billiardkugel auf einem Billiardtisch der Länge 1. Die Zufallsvariable X beschreibe den Punkt auf $[0, 1]$, auf dem die Kugel zu liegen kommt. Annahme: $X \sim U[0, 1]$.
Wir rollen weitere n rote Kugeln. Beschreibe die Zufallsvariable Y_i ob die rote Kugel i links von der weißen Kugel zu liegen kommt (also auf einem Punkt kleiner X). Offensichtlich gilt $\mathbb{P}(Y_i = 1|X = x) = x$.
Wir kennen die Anzahl y der Kugeln, die insgesamt links von $X = x$ zu liegen kommen. Mit $Y = \sum_{i=1}^n Y_i$ gilt $Y|X = x \sim B(n, x)$. Können wir eine Aussage über die Verteilung von X gegeben $Y = y$ treffen?

Satz von Bayes:

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{Y|X}(y|x)f_X(x)}{\int f_{Y|X}(y|x)f_X(x)dx} \\ &= \frac{\binom{n}{y}x^y(1-x)^{n-y} \cdot 1_{[0,1]}(x)}{\int \binom{n}{y}x^y(1-x)^{n-y} \cdot 1_{[0,1]}(x)dx} \\ &= \frac{1}{B(y+1, n-y+1)}x^{y+1-1}(1-x)^{n-y+1-1} \end{aligned}$$

Also ist $X|Y = y \sim \mathrm{Beta}(y + 1, n - y + 1)$. (vgl. Definition 12.5)

Definition 15.3
Seien X und Y reelwertige Zufallsvariablen. Ist $f_{X|Y}(x|y)$ die bedingte Dichte von X gegeben $Y = y$, dann ist

$$\mathbb{E}(X|Y = y)) = \int x f_{X|Y}(x, y) d\lambda(x)$$

der bedingte Erwartungswert von X gegeben $Y = y$.
Damit ist

$$\begin{aligned}\mathbb{E}(X|Y) : \mathbb{R} &\rightarrow \mathbb{R} \\ \text{mit } \mathbb{E}(X|Y)(y) &:= \mathbb{E}(X|Y = y),\end{aligned}$$

der bedingte Erwartungswert von X gegeben Y, eine Zufallsvariable.

Satz 15.6 (Satz vom iterierten Erwartungswert)

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$

Satz 15.7 (Varianzzerlegungssatz)
Sei $X|Y$ eine stetige bedingte Zufallsvariable. Dann gilt:

$$\mathbb{V}(X) = \mathbb{E}(\mathbb{V}(X|Y)) + \mathbb{V}(\mathbb{E}(X|Y)$$

Beweis:

Siehe Wahrscheinlichkeitstheorie und Inferenz II. □

Beispiel 15.4
Seien X und Y Zufallsvariablen mit gemeinsamer Dichte

$$f_{X,Y}(x, y) = \begin{cases} 1/x & \text{für } 0 < y \leq x < 1 \\ 0 & \text{sonst} \end{cases}$$

Die Dichte der Randverteilung von X ist

$$f_X(x) = \int_0^x \frac{1}{x} dy = I_{(0,1)}(x).$$

Die Dichte der Randverteilung von Y ist

$$f_Y(y) = \int_y^1 \frac{1}{x} dx = \log\left(\frac{1}{y}\right) I_{(0,1)}(y).$$

Die Dichte der bedingten Verteilung von $Y|X$ ist

$$f_{Y|X}(y|x) = \frac{1}{x} I_{(0,x)}(y).$$

Die Dichte der bedingten Verteilung von $X|Y$ ist

$$f_{X|Y}(x|y) = \frac{1/x}{\log(1/y)} I_{[y,1)}(x).$$

$X \sim U(0,1) \implies \mathbb{E}(X) = 1/2; \mathbb{V}(X) = 1/12$
$Y \sim ?; Y|X \sim U(0,x)!$

$$\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}\left(\frac{X}{2}\right) = \frac{1}{4} \\
\mathbb{V}(Y) &= \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) = \tfrac{1}{12}\mathbb{E}(X^2) + \tfrac{1}{4}\mathbb{V}(X) = 7/144 \\
\mathbb{E}(X \cdot Y) &= \mathbb{E}(\mathbb{E}(X \cdot Y|X) = \mathbb{E}(X\mathbb{E}(Y|X)) \\
&= \mathbb{E}\left(X\frac{X}{2}\right) = \frac{1}{2}\mathbb{E}(X^2) \\
&= \frac{1}{2}\left(\mathbb{V}(X) + \mathbb{E}(X)^2\right) = \frac{1}{2}\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{1}{6} \\
\mathrm{Cov}(X,Y) &= \frac{1}{6} - \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{24}
\end{aligned}$$

Beispiel 15.5
Normal-Gamma-Verteilung
Gegeben sei eine zweidimensionale Zufallsvariable (X,T) mit gemeinsamer Dichte

$$f(x,t) = \frac{\delta^\delta}{\Gamma(\delta)\sqrt{2\pi}} t^{\delta-\frac{1}{2}} \exp\left(-\delta t - \frac{t}{2}x^2\right)$$

Es gilt :

- *Randverteilung $T \sim Ga(\delta,\delta)$*
- *Randverteilung $X \sim t(2\delta)$*
- *Bedingte Verteilung $X|T = t \sim N(0, 1/t)$*

Definition 15.4
Seien Z_t unabhängig und identisch verteilt für $t = 1,\ldots,T$. Dann heißt $X = (X_0,\ldots,X_T)$ mit $X_t = X_{t-1} + Z_t$ und $X_0 = 0$ (einfache) Irrfahrt (Random Walk).

Bemerkung:

Allgemein betrachtet man auch den Fall $T \to \infty$.

Beispiel 15.6
Man betrachte zwei Spieler A und B, die in jeder Spielrunde mit Wahrscheinlichkeit p bzw. q (bzw. r) gewinnen (bzw. unentschieden spielen). X_t bezeichne den Gewinn von A nach t Spielrunden.

Beispiel 15.7
Sei $Z_t \sim N(0, \sigma^2)$. *Dann bezeichnet man* $X = (X_1, \ldots, X_T)$ *als Gauss-Irrfahrt. Es gilt*

- $X_t = \sum_{i=1}^{t} Z_i \sim N(0, t\sigma^2)$
- $X_{t+1}|X_t = x_t \sim N(x_t, \sigma^2)$
- $X \sim N_T(0, \Sigma)$ *mit*

$$Cov(X_t, X_s) = min(t, s)$$

bzw.

$$\Sigma^{-1} = Q = \begin{pmatrix} 2 & 1 & 0 & 0 & \cdots \\ 1 & 2 & 1 & 0 & \cdots \\ 0 & 1 & 2 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Definition 15.5
Sei X, Y *und* Z *Zufallsvariablen, so dass*

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z),$$

dann sind X *und* Y *bedingt unabhängig gegeben* Z.

Definition 15.6
Sei $X = (X_1, \ldots, X_T)$. *Sind* $(X_1, \ldots, X_{t-1})$ *und* $(X_t + 1, X_T)$ *unabhängig gegeben* X_t *für alle* t, *dann erfüllt* X *die Markov-Eigenschaft. Man nennt* X *eine Markov-Kette.*

Bemerkung:

Die Markov-Eigenschaft lässt sich auch auf stetige Zeit ($t \in R$) verallgemeinern.

Definition 15.7
Sei

$$X_t : \ (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (S, \mathcal{S})$$

für alle t *Zufallsvariablen auf* $(\Omega, \mathcal{F}, \mathbb{P})$. *Dann heißt*

$$X = \{\Omega, \mathcal{F}, P, \{X_t, \ t \in T\}\}$$

stochastischer Prozess. T *heißt Indexmenge,* S *heißt Zustandsraum.*

Beispiel 15.8
$T = \mathbb{R}_0^+$.

(a) *X ist ein Gauß-Prozess, wenn gilt* $(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \sim N_k(\mu, \Sigma)$ *für alle* $t_i \in T$ *für* $i = 1, \ldots, k$ *und für alle* k*. Speziell beim stationären, isotropen Gauß-Prozess gilt* $\mu = 0$ *und* $Cov(X_t, X_s) = \sigma^2 \rho(|t - s|)$*, wobei* ρ *eine Korrelationsfunktion ist.*

(b) *Beim (stationären, isotropen) Poisson-Prozess gilt für* $t > s$*:*

$$X_t - X_s \sim Po(\lambda(t - s)).$$

Teil IV
Konvergenz

16 Konvergenzarten

Beispiel 16.1
$X_i \sim B(1, \frac{1}{2})$; $\quad i = 1, 2, \ldots$ *stu.*
$Y_n = \sum_{i=1}^{n} X_i \sim B\left(n, \frac{1}{2}\right)$

- *Wie* groß *ist der Abstand* $\frac{1}{n}Y_n - \frac{1}{2}$ *?*
- *Wie ändert sich dieser Abstand mit* n*?*
- *Sei* $Z_i = 1 - X_i$. *Offensichtlich ist* $Z_i \neq X_i$, *aber* Z_i *und* X_i *haben die gleiche Verteilung!*

Definition 16.1
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:

$$X_n \overset{f.s.}{\to} X,\ X_n \overset{a.s.}{\to} X,\ X_n \overset{a.e.}{\to} X,\ X_n \to X \text{ wp } 1.$$

b) *im* r*-ten Moment,* $r \geq 1$*, wenn*

$$\mathbb{E}(|X_n|^r) < \infty\ \forall\, n \quad \text{und} \quad \mathbb{E}(|X_n - X|^r) \to 0 \quad \text{für } n \to \infty;$$

Schreibweise:

$$X_n \overset{r}{\to} X$$

c) *in Wahrscheinlichkeit, falls*

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0 \quad \text{für } n \to \infty\ \forall \epsilon > 0;$$

Schreibweise:

$$X_n \overset{\mathbb{P}}{\to} X$$

d) *in Verteilung, falls*

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x)$$

an allen Stetigkeitsstellen von $F_X(x) = \mathbb{P}(X \leq x)$; *Schreibweise:*

$$X_n \overset{\mathcal{D}}{\to} X$$

Bemerkung:

a) "Analogie: Punktweise Konvergenz"
$A := \{\omega \in \Omega \,|\, \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ ist uninteressant, da keine W'keit vorkommt. Konvergenz fast sicher meint, daß $\bar{A}$ ("Menge auf der X_n nicht gegen X konvergiert") eine Nullmenge bzgl. $\mathbb{P}$ ist. Allgemein für Folgen meßbarer Funktionen: Konvergenz μ-fast überall
b) Schreibweisen: $X_n \overset{2}{\to} X$ heißt $\mathbb{E}(|X_n - X|^2) \to 0$.
Allgemeiner: Eine Folge $(f_n)_{n\in\mathbb{N}} \in L^p$ heißt in $\underline{L^p \text{ konvergent}}$, wenn es ein $f \in L^p$ gibt, so daß

$$\|f_n - f\|_p \to 0 \quad \text{für } n \to \infty$$

Kurz: $f_n \overset{L^p}{\to} f$.
Zur Erinnerung: $(\mathbb{E}(|X|^p)) = \|X\|_p^p$
c) "Wahrscheinlichkeit großer Abweichungen zwischen X_n und X wird immer kleiner mit wachsendem n". $\mathbb{P}(|X_n - X| > \epsilon) = \displaystyle\int I_{A_n}\, d\mathbb{P}$ mit $A_n = \{\omega \in \Omega \,|\, |X_n(\omega) - X(\omega)| > \epsilon\}$
d) "Verteilung von X wird immer besseres Modell für X_n mit wachsendem n." $X_n(\omega) \neq X(\omega) \quad \forall n, \omega \in \Omega$, und dennoch $X_n \overset{D}{\to} X$ ist erlaubt.

Beispiel 16.2 (Fortsetzung)
$Z_i \overset{\mathcal{D}}{\to} X_1$ *denn* $Z_i \sim B\left(1, \frac{1}{2}\right), X_1 \sim B\left(1, \frac{1}{2}\right)$
aber $Z_i \neq X_1$

$$\mathbb{E}\left(\frac{1}{n}\, Y_n\right) = \frac{1}{n}\, n\, p = p$$

$$\Longrightarrow \mathbb{E}\left(\left(\frac{1}{n}\, Y_n - p\right)^2\right) = \mathbb{V}\left(\frac{1}{n} Y_n\right) = \frac{1}{n^2}\, n\, p\,(1-p) = \frac{1}{n}\, p\,(1-p) \to 0$$

und damit $\frac{1}{n}\, Y_n \xrightarrow{r=2} p$

$$\mathbb{E}\left(\frac{1}{n}\, Y_n\right) = \frac{1}{n}\, \mathbb{E}(Y_n) = \frac{1}{n}\, n\, p = p$$

$$\mathbb{P}\left(\left|\frac{1}{n}\, Y_n - \mathbb{E}\left(\frac{1}{n}\, Y_n\right)\right| > \epsilon\right) \overset{\textit{Tschebyschev}}{\leq} \frac{\mathbb{V}\left(\frac{1}{n} Y_n\right)}{\epsilon^2} = \frac{\frac{1}{n^2}\, n\, p\,(1-p)}{\epsilon^2}$$

$$= \frac{\frac{1}{n}\, p\,(1-p)}{\epsilon^2} \to 0$$

für $n \to \infty$ *und damit* $\frac{1}{n} Y_n \overset{\mathbb{P}}{\to} p = \frac{1}{2}$.

Satz 16.1
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) $X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\mathcal{D}} X$

b) $X_n \xrightarrow{f.s.} X \implies X_n \xrightarrow{\mathbb{P}} X$

c) $X_n \xrightarrow{r} X \implies X_n \xrightarrow{\mathbb{P}} X \quad$ *für* $r \geq 1$

d) $X_n \xrightarrow{r} X \implies X_n \xrightarrow{s} X \quad r > s \geq 1$

Beweis:
a) $X_n \xrightarrow{\mathbb{P}} X; \quad F_n(x) = \mathbb{P}(X_n \leq x); \quad F(x) = \mathbb{P}(X \leq x)$. Es gilt für $\epsilon > 0$:

$$\begin{aligned}
F_n(x) = \mathbb{P}(X_n \leq x) &= \mathbb{P}(X_n \leq x, X \leq x+\epsilon) + \mathbb{P}(X_n \leq x, X > x+\epsilon) \\
&\leq \mathbb{P}(X \leq x+\epsilon) + \mathbb{P}(X_n - X \leq x - X, x - X < -\epsilon) \\
&\leq F(x+\epsilon) + \mathbb{P}(X_n - X < -\epsilon) \\
&\leq F(x+\epsilon) + \mathbb{P}(X_n - X < -\epsilon) + \mathbb{P}(X_n - X > \epsilon) \\
&\leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon) \\
\text{und } F(x-\epsilon) &= \mathbb{P}(X \leq x-\epsilon) = \mathbb{P}(X \leq x-\epsilon, X_n \leq x) + \mathbb{P}(X \leq x-\epsilon, X_n > x) \\
&\leq F_n(x) + \mathbb{P}(|X_n - X| > \epsilon) \qquad \text{(Umformung s.o.)}
\end{aligned}$$

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 $n \to \infty$:

$$\begin{aligned}
F(x-\epsilon) - \underbrace{\mathbb{P}(|X_n - X| > \epsilon)}_{\to 0} \leq F_n(x) \leq F(x+\epsilon) + \underbrace{\mathbb{P}(|X_n - X| > \epsilon)}_{\to 0} \\
\implies F(x-\epsilon) \leq \lim_{n\to\infty} F_n(x) \leq F(x+\epsilon)
\end{aligned}$$

Falls F in x stetig:

$$\begin{aligned}
F(x) \leq \lim_{n\to\infty} F_n(x) \leq F(x) \\
\implies F(x) = \lim_{n\to\infty} F_n(x) \implies X_n \xrightarrow{\mathcal{D}} X
\end{aligned}$$

d) Vorraussetzung ist:

$$\mathbb{E}(|X_n - X|^r) = \|X_n - X\|_r^r \to 0.$$

Nach Satz 8.5 gilt für $r > s > 1$ mit $c \geq 0$:

$$\|X_n - X\|_r^r \geq c\|X_n - X\|_s^s \to 0$$

c) $r = 1$ reicht wegen d). Dann

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}(|X_n - X|)}{\epsilon} \quad \text{(Markov für } X_n - X)$$

b)

Sei $C = \{\omega \in \Omega \mid X_n(\omega) \not\to X(\omega)\}$. Nach Def. $X_n \xrightarrow{\text{f.s.}} X \iff \mathbb{P}(C) = 0$.

Sei $B_n = \bigcup_{m \geq n} \{\omega \in \Omega \mid |X_m(\omega) - X(\omega)| > \epsilon\}$ für beliebiges $\epsilon > 0$. Die Folge B_n ist monoton absteigend: $B_n \supseteq B_{n+1} \supseteq \ldots$, und es gilt für

$$B = \bigcap_{n \geq 1} B_n : \quad B_n \downarrow B, \text{ also auch } \lim_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(B).$$

Zeige zunächst dass $\mathbb{P}(B) = 0$:

$$\begin{aligned}
&\forall \omega \in \bar{C} : X_n(\omega) \to X(\omega) \\
\Longrightarrow\, &\forall \omega \in \bar{C} : \exists n : |X_m(\omega) - X(\omega)| < \epsilon \, \forall m \geq n \\
\Longrightarrow\, &\forall \omega \in \bar{C} : \exists n : \omega \notin B_m \, \forall m \geq n \\
\Longrightarrow\, &\forall \omega \in \bar{C} : \omega \notin B \\
\Longrightarrow\, &B \cap \bar{C} = \emptyset \implies B \subseteq C \implies P(B) \leq P(C) = 0
\end{aligned}$$

Da $\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \epsilon\} \subseteq B_n$ und $\lim_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}(B) = 0$:

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \mathbb{P}(B_n) \to 0 \implies X_n \xrightarrow{\mathbb{P}} X.$$

□

Bemerkung:

Andere Beziehungen gelten so allgemein nicht.

- $X_n \xrightarrow{s} X \quad \not\Rightarrow \quad X_n \xrightarrow{r} X \quad r > s \geq 1$

  Betrachte

  $$X_n = \begin{cases} n & \text{mit Wahrscheinlichkeit } n^{-\frac{1}{2}(r+s)} \\ 0 & \text{mit Wahrscheinlichkeit } 1 - n^{-\frac{1}{2}(r+s)} \end{cases}$$

  $$\begin{aligned} \mathbb{E}(|X_n|^s) &= n^s \cdot n^{-\frac{1}{2}(r+s)} \\ &= n^{\frac{1}{2}(s-r)} \to 0 \quad \text{weil } s < r \end{aligned}$$

$$\begin{aligned}\mathbb{E}(|X_n|^r) &= n^r \cdot n^{-\frac{1}{2}(r+s)} \\ &= n^{\frac{1}{2}(r-s)} \to \infty \quad \text{weil } r > s\end{aligned}$$

und somit $\quad X_n \xrightarrow{s} 0 \quad$ aber nicht $\quad X_n \xrightarrow{r} 0$

- $X_n \xrightarrow{\mathbb{P}} X \quad \not\Rightarrow \quad X_n \xrightarrow{1} X$

  Betrachte

  $X_n = \begin{cases} n^3 & \text{mit Wahrscheinlichkeit } n^{-2} \\ 0 & \text{mit Wahrscheinlichkeit } 1 - n^{-2} \end{cases}$

  $\mathbb{P}(|X_n| > \epsilon) = n^{-2} \to 0 \iff X_n \xrightarrow{\mathbb{P}} 0$

  Aber $\quad \mathbb{E}(|X_n|) = n^3 \cdot n^{-2} = n \to \infty$

Satz 16.2

a) $X_n \xrightarrow{\mathcal{D}} c$ *konstant* $\implies X_n \xrightarrow{\mathbb{P}} c$

b) $X_n \xrightarrow{\mathbb{P}} X$, $\mathbb{P}(|X_n| \le k) = 1 \, \forall n$ *und ein* $k \in \mathbb{R}^+ \implies X_n \xrightarrow{r} X \quad \forall r \ge 1$

c) $P_n(\epsilon) = \mathbb{P}(|X_n - X| > \epsilon) \quad$ *mit* $\sum_n P_n(\epsilon) < \infty \quad \forall \epsilon > 0 \implies X_n \xrightarrow{f.s.} X$

Beweis:

a)

$$\begin{aligned}\mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n < c - \epsilon) + \mathbb{P}(X_n > c + \epsilon) \\ &\le \mathbb{P}(X_n \le c - \epsilon) + (1 - \mathbb{P}(X_n \le c + \epsilon)) \\ &\to \mathbb{P}(X \le c - \epsilon) + (1 - \mathbb{P}(X \le c + \epsilon)) \\ &= 0 + (1 - 1) = 0\end{aligned}$$

b) $X_n \xrightarrow{\mathbb{P}} X, \quad \mathbb{P}(|X_n| \le k) = 1 \implies \mathbb{P}(|X| \le k) = 1$

da $\mathbb{P}(|X| \le k + \epsilon) = \lim_{n\to\infty} \mathbb{P}(|X_n| \le k + \epsilon) = 1 \, \forall \, \epsilon > 0$.

Sei $\quad A_n(\epsilon) = \{\omega \in \Omega \,|\, |X_n(\omega) - X(\omega)| > \epsilon\}$, dann gilt:

$$|X_n(\omega) - X(\omega)|^r \le \epsilon^r \quad \omega \in \overline{A_n(\epsilon)}$$

bzw. $|X_n(\omega) - X(\omega)|^r \le (2k)^r \quad \omega \in A_n(\epsilon)$ jeweils mit W'keit 1.

Also $|X_n - X|^r \le \epsilon^r \, I_{\overline{A_n(\epsilon)}} + (2k)^r \, I_{A_n(\epsilon)}$

$$\begin{aligned}\mathbb{E}(|X_n - X|^r) &\le \epsilon^r \, \mathbb{P}(\overline{A_n(\epsilon)}) + (2k)^r \, \mathbb{P}(A_n(\epsilon)) \\ &= ((2k)^r - \epsilon^r) \, \underbrace{\mathbb{P}(A_n(\epsilon))}_{\to 0} + \epsilon^r \\ &\to \epsilon^r \quad \text{für } n \to \infty \\ &\to 0 \quad \text{für } \epsilon \downarrow 0 \\ &\implies X_n \xrightarrow{r} X\end{aligned}$$

c) Ohne Beweis.

$\square$

17 Gesetz der großen Zahlen

Satz 17.1 (Schwaches Gesetz der großen Zahlen)
Seien $X_1, \dots, X_n$ unabhängig und identisch verteilt mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 < \infty \quad \forall i = 1, \dots, n$. 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

$$\mathbb{E}(|\bar{X} - \mu|^2) = \mathbb{V}(\bar{X}) = \frac{1}{n^2}\, n\, \sigma^2 = \frac{1}{n}\, \sigma^2 \to 0$$

$\implies \bar{X} \xrightarrow{\mathbb{P}} \mu$ wegen Satz 16.1 c). $\square$

Beispiel 17.1

$X_i \sim Po(\lambda)$ stu. $\implies \mathbb{E}(X_i) = \lambda$, $\mathbb{V}(X_i) = \lambda, i = 1, \dots, n$ 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 17.2
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) $\frac{A_n}{B_n} \xrightarrow{\mathbb{P}} \frac{a}{b} \quad \forall b \neq 0$

Beweis:

Vorarbeit: Seien E_n und F_n Folgen von Ereignissen ($E_n \in \mathcal{F}, F_n \in \mathcal{F}$), dann gilt: $\mathbb{P}(E_n) \to 1$ und $\mathbb{P}(F_n) \to 1 \implies \mathbb{P}(E_n \cap F_n) \to 1$ 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: $\mathbb{P}(|A_n - a| < \epsilon) \to 1; \mathbb{P}(|B_n - b| < \epsilon) \to 1$

$$\begin{aligned}\implies \quad & \mathbb{P}(|A_n - a| < \epsilon \wedge |B_n - b| < \epsilon) \\ \leq \quad & \mathbb{P}(|A_n - a + B_n - b| < \epsilon) = \mathbb{P}(|(A_n + B_n) - (a+b)| < \epsilon) \to 1\end{aligned}$$

□

Satz 17.3
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

$$\implies f(X_n) \xrightarrow{\mathbb{P}} f(c).$$

Beweis:

f stetig in $c \iff |f(x) - f(c)| < a,\ a > 0 \quad$ für ein b mit $|x - c| < b$.
$\mathbb{P}(|f(X_n) - f(c)| < a) \geq \mathbb{P}(|X_n - c| < b) \to 1 \quad \forall a > 0$
$\implies f(X_n) \xrightarrow{\mathbb{P}} f(c)$ □

Bemerkung:

Die Ergebnisse lassen sich im Wesentlichen auf den k-dimensionalen Fall übertragen, siehe Lehmann [2001].

Bemerkung:

$X_n \xrightarrow{\mathbb{P}} c \quad \text{konstant} \quad \not\Rightarrow \quad \mathbb{E}(X_n) \to c$

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 $\mathbb{P}(|X_n - 1| > \epsilon) \to 0 \implies X_n \xrightarrow{\mathbb{P}} 1$
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 17.2
$X_n \sim N(\mu, \sigma_n^2), n = 1, \ldots\ X_n \xrightarrow{\mathbb{P}} \mu \iff \sigma_n^2 \to 0.$

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* $\sigma_n \to 0$.

Satz 17.4 (Starkes Gesetz der großen Zahlen nach Kolmogorov)
Seien $X_1, \dots, X_n$ unabhängige, identische verteilte Zufallsvariablen. Dann gilt: Existiert der Erwartungswert $\mathbb{E}(X_1)$ und ist er endlich (d.h. $\mathbb{E}(|X_1|) < \infty$), so konvergiert $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$ $\mathbb{P}$-fast sicher mit

$$\lim_{n\to\infty} \bar{X} = \mathbb{E}(X_1) \qquad \mathbb{P} - f.s.$$

Ist $\mathbb{E}(X_1) = \infty$, so konvergiert $\bar{X}$ $\mathbb{P}$-fast sicher nicht gegen einen endlichen Grenzwert.

Beweis:

Schmidt [2011], Beweis zu Satz 15.2.7 □

Beispiel 17.3 (Monte-Carlo-Methode)
Seien $(X_i, Y_i)_{i=1,\dots,n}$ unabhängige, gleichverteilte Zufallsvariablen auf $\Omega = [0,1) \times [0,1)$. Sei B eine Borelmenge aus Ω und $Z_i = 1_B(X_i, Y_i)$. Es gilt

$$p := P(Z_i = 1) = P((X_i, Y_i) \in B) = \int_B 1 d\lambda^2$$

Z ist unabhängig identisch Bernoulli-verteilt mit $p = P(Z_1 = 1)$, daher $\mathbb{E}(Z_1) = p$, also Fläche von B.

Monte-Carlo-Integration:

- Ziehe n unabhängige gleichverteilte Zufallszahlen (X_i, Y_i) auf $[0,1)^2$
- Stelle fest, ob (X_i, Y_i) in B
- Gesetz der großen Zahlen sagt:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n Z_i = p$$

18 Zentraler Grenzwertsatz

18.1 Konvergenz in Verteilung

Beispiel 18.1
$X_n \sim N(0, \sigma_n^2)$ mit Verteilungsfunktion H_n mit $\sigma_n^2 \to \infty$
$H_n(x) = \mathbb{P}\left(\frac{X_n}{\sigma_n} \le \frac{x}{\sigma_n}\right) = \Phi\left(\frac{x}{\sigma_n}\right)$

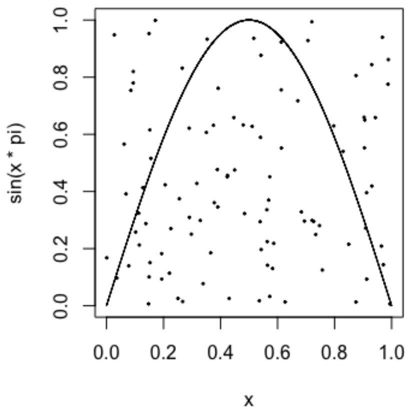


Abbildung 4: Beispiel zur Monte Carlo-Integration

Mit $\sigma_n^2 \to \infty$ *folgt* $\dfrac{x}{\sigma_n} \to 0$ *und somit* $H_n(x) \to \Phi(0) = \dfrac{1}{2} \quad \forall x \in \mathbb{R}$. *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 18.1
Sei $(X_n)_{n\in\mathbb{N}}$ *eine Folge von ZV mit Verteilungsfunktion* $X_n \sim F_n$. *Dann existiert* $X \sim F$ *mit* $X_n \xrightarrow{\mathcal{D}} X$ *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 $\quad \surd$

- $\lim\limits_{x\to-\infty} F(x) = 0,\ \lim\limits_{x\to\infty} F(x) = 1$

$F_n(x) = \mathbb{P}(X_n \le x) \ge 1-\epsilon \quad$ für $x \ge k, n > n_0$ und $\forall \epsilon > 0$
also $\lim\limits_{n\to\infty} F_n(x) = F(x) \ge 1-\epsilon \quad$ für $x \ge k$ und $\forall \epsilon > 0$
und $\lim\limits_{x\to\infty} F(x) = 1 \quad$ für $\epsilon \downarrow 0$ $\square$

Beispiel 18.2
$X \sim U(0,1)$
$X_n = \begin{cases} X & n \text{ gerade} \\ 1-X & n \text{ ungerade} \end{cases}$
Dann
$X_n \overset{\mathcal{D}}{\to} X$ *und*
$X_n \overset{\mathcal{D}}{\to} 1-X$ *nämlich*
$X_n \overset{\mathcal{D}}{\to} U(0,1) \quad$ *(d.h. jede ZV, welche gleichverteilt auf* $[0,1]$ *ist).*
$Y_n \sim N(0,1) \overset{\mathcal{D}}{\to} Y \sim N(0,1)$
$\overset{\mathcal{D}}{\to} -Y \sim N(0,1)$ *Grenzwert nicht eindeutig!*

Beispiel 18.3
$X_n \sim N(\mu_n, \sigma_n^2) \quad \mu_n \to 0, \sigma_n^2 \to 1$
$\implies X_n \overset{\mathcal{D}}{\to} N(0,1)$ *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 $\frac{x-\mu_n}{\sigma_n} \to x$ *folgt aus der Stetigkeit von* Φ *auch* $H_n(x) \to \Phi(x)$ *und somit* $X_n \overset{\mathcal{D}}{\to} N(0,1)$

Satz 18.2

$$X_n \overset{\mathcal{D}}{\to} X,\ a, b \ne 0 \in \mathbb{R} \textit{ konstant} \implies b\,X_n + a \overset{\mathcal{D}}{\to} b\,X + a.$$

Beweis:

Sei o.B.d.A $b > 0$.
$\mathbb{P}(b\,X_n + a \le x) = \mathbb{P}\left(X_n \le \frac{x-a}{b}\right) = F_n\left(\frac{x-a}{b}\right)$
$\to F\left(\frac{x-a}{b}\right) = \mathbb{P}\left(X \le \frac{x-a}{b}\right)$ $\square$

Satz 18.3
$X_n \overset{\mathcal{D}}{\to} X \iff \mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$

$\forall$ *beschränkten und stetigen Funktionen* $g : \mathbb{R} \to \mathbb{R}$.

Beweis:

ohne Beweis $\square$

Lemma 18.1

$$X_n \xrightarrow{\mathcal{D}} X \iff \varphi_{X_n}(t) \to \varphi_X(t) \quad \forall t$$

Beweis:

i.W. Konvergenz der $\mathbb{E}$-Werte $\quad \mathbb{E}(X_n^r) \to \mathbb{E}(X^r)$ $\square$

Satz 18.4 (Slutsky's Theorem)

$$X_n \xrightarrow{\mathcal{D}} X,\ A_n \xrightarrow{\mathbb{P}} a \text{ und } B_n \xrightarrow{\mathbb{P}} b \implies A_n + B_n\, X_n \xrightarrow{\mathcal{D}} a + b\, X$$

Beweis:

(Skizze)
Sei $Y_n \xrightarrow{\mathcal{D}} Y$ unabhängig von X_n, X. Dann
$\varphi_{(X_n,Y_n)}(s) \overset{\text{stu}}{=} \varphi_{X_n}(s) \cdot \varphi_{Y_n}(s) \overset{\text{Lemma 18.1}}{\Longrightarrow} \varphi_X(t) \cdot \varphi_Y(t) = \varphi_{(X,Y)}(t)$
Mit $Y_n \xrightarrow{\mathbb{P}} a$ und Faltung folgt $X_n + Y_n \to X + a$, analog für $Y_n \xrightarrow{\mathbb{P}} b$. $\square$

18.2 Zentraler Grenzwertsatz

Beispiel 18.4
$X_n \sim Po(\lambda) \quad \mathbb{E}(X_n) = \lambda = \mathbb{V}(X_n)$
$\bar{X}_n = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \lambda$
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* $\mathbb{V}(\bar{X}_n - \lambda) = \mathbb{V}(\bar{X}_n) = \dfrac{1}{n^2}\mathbb{V}\left(\sum X_i\right) = \dfrac{1}{n^2}\, n\, \lambda = \dfrac{\lambda}{n} \to 0$
Aber: $\dfrac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}}$ *hat Varianz* 1.
Was ist die Verteilung von $\dfrac{\bar{X}_n - \lambda}{\sqrt{\lambda/n}}$?

Satz 18.5 (Zentraler Grenzwertsatz)
Seien $X_i, i = 1, \dots$ unabhängig und identisch verteilt (u.i.v.) mit $\mathbb{E}(X_i) = \mu$ und $\mathbb{V}(X_i) = \sigma^2 < \infty \; \forall i = 1, \dots$. Dann gilt

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\mathcal{D}} N(0,1)$$

bzw.

$$\sqrt{n}\,(\bar{X} - \mu) \xrightarrow{\mathcal{D}} N(0, \sigma^2)$$

Bemerkung:

Statt u.i.v. häufig auch i.i.d. (*independent identically distributed*).

Beweis:

Sei $Y_i = \dfrac{X_i - \mu}{\sigma}$, also $\mathbb{E}(Y_i) = 0, \mathbb{V}(Y_i) = 1$ und $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$. Sei $K_X(s) = \log M_X(s)$. Benutze $\left.\frac{\partial^n K_X(s)}{\partial^n s}\right|_{s=0} = \left.\frac{\partial^n \log(M_X(s))}{\partial^n s}\right|_{s=0} = \frac{1}{M_X(s)^n} \left.\frac{\partial^n M_X(s)}{\partial^n s}\right|_{s=0} = \mathbb{E}(X^n)$ für $n \in \{1,2\}$ da $M(0) \equiv 1$.

$$\begin{aligned}
M_{Z_n}(t) &= \mathbb{E}\left(\exp(tZ_n)\right) = \mathbb{E}\left(\exp\left(\frac{t}{\sqrt{n}} \sum_{i=1}^n Y_i\right)\right) \\
&\overset{stu}{=} \prod_{i=1}^n \mathbb{E}\left(\exp\left(\frac{t}{\sqrt{n}} Y_i\right)\right) \overset{u.i.v.}{=} M_Y\left(\frac{t}{\sqrt{n}}\right)^n
\end{aligned}$$

$$\text{also } K_{Z_n}(t) = \log M_{Z_n}(t) = n \log M_Y\left(\frac{t}{\sqrt{n}}\right) = nK_Y\left(\frac{t}{\sqrt{n}}\right)$$

$$\begin{aligned}
\lim_{n\to\infty} K_{Z_n}(t) &= \lim_{n\to\infty} nK_Y\left(\frac{t}{\sqrt{n}}\right) \\
&= \lim_{\epsilon\to 0} \frac{K_Y(\epsilon t)}{\epsilon^2} \qquad (\epsilon := \tfrac{1}{\sqrt{n}}) \\
&= \lim_{\epsilon\to 0} \frac{tK_Y'(\epsilon t)}{2\epsilon} \qquad (\text{L'Hospital da ``}\tfrac{0}{0}\text{''}) \\
&= \lim_{\epsilon\to 0} \frac{t^2 K_Y''(\epsilon t)}{2} = \frac{t^2}{2} \qquad \text{da } K_Y''(0) = \mathbb{V}(Y) = 1.
\end{aligned}$$

Behauptung folgt aus $M_{\mathrm{N}(0,1)}(t) = \exp\left(\dfrac{t^2}{2}\right)$. $\square$

Bemerkung:

Der ZGWS gilt für alle X_i mit beliebiger Verteilung!

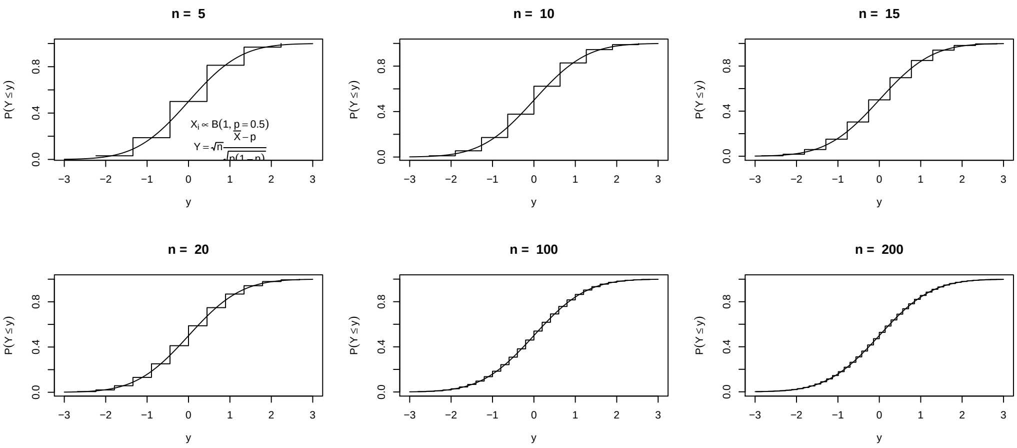


Abbildung 5: Approximation der Binomialverteilung durch die Normalverteilung.

Beispiel 18.5 (Bernoulliverteilung)
$X_i \sim B(1,p), i = 1,\ldots,n$ mit $p = 0.5$. Die Verteilung des standardisierten Mittelwertes konvergiert nach dem ZGWS gegen eine Normalverteilung: $\sqrt{n}\frac{n^{-1}\sum_{i=1}^n X_i - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{D}} N(0,1)$

Beispiel 18.6 (Poissonverteilung)
$X_i \sim Po(\lambda), i = 1,\ldots,n$ mit $\lambda = 5$. Die Verteilung des standardisierten Mittelwertes konvergiert nach dem Zentralen Grenzwertsatz gegen eine Normalverteilung $\sqrt{n}\frac{n^{-1}\sum_{i=1}^n X_i - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} N(0,1)$

Beispiel 18.7 (χ^2-Verteilung)
$X \sim \chi_n^2$ 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 $y = g(x) = \frac{x-n}{\sqrt{2n}}$. Mit dem Transformationssatz folgt

$$f_Y(y) = f_{\chi_n^2}(y\sqrt{2n} + n)\sqrt{2n}$$

Da X die Summe von unabhängigen identisch verteilten quadrierten normalverteilten ZV ist, folgt mit dem ZGWS, daß die zugehörigen Verteilungsfunktionen konvergieren, und zwar $F_Y(y) \to \Phi(y) \iff Y \xrightarrow{\mathcal{D}} N(0,1)$.

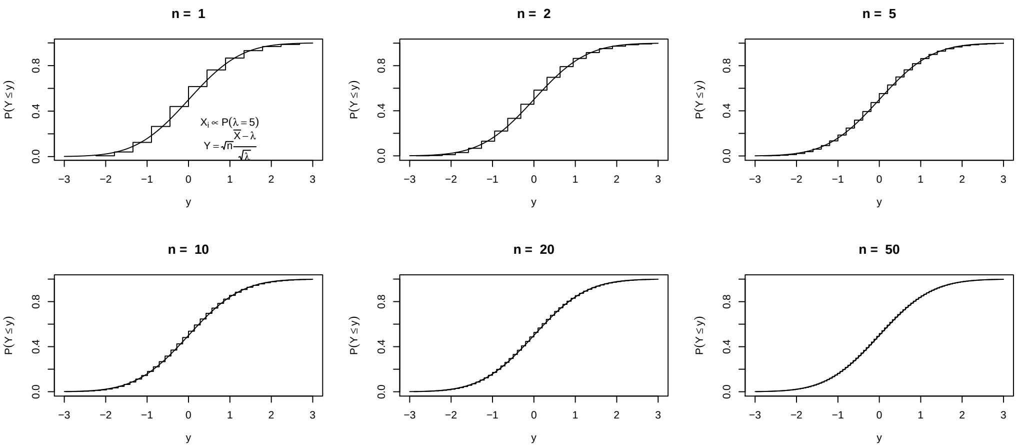


Abbildung 6: Approximation der Poissonverteilung durch die Normalverteilung.

Beispiel 18.8
$X_1, \ldots, X_n$ iid mit $\mathbb{E}(X_i) = 0$, $\mathbb{V}(X_i) = \sigma^2$ und $\mathbb{V}(X_i^2) = \tau^2 < \infty$. $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \xrightarrow{\mathbb{P}} \sigma^2$ $\sqrt{n}\,(S^2 - \sigma^2) \xrightarrow{\mathcal{D}}$?
S^2 hängt nicht von μ ab, also o.B.d.A. $\mu = 0$
ZGWS: $\quad \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \sigma^2 \right) \xrightarrow{\mathcal{D}} N(0, \tau^2)$
Gesetz der großen Zahlen: $\quad \bar{X}^2 \xrightarrow{\mathbb{P}} 0$
Damit folgt: $\quad \sqrt{n}\,(S^2 - \sigma^2) = \sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \underbrace{\bar{X}^2}_{\to 0} - \sigma^2 \right) \xrightarrow{\mathcal{D}} N(0, \tau^2)$

18.3 Konvergenzrate

Satz 18.6
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:

$$\begin{aligned} G_n(x) &= \mathbb{P}\left(\sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \leq x \right) \\ &= \Phi(x) + \underbrace{\frac{\mu_3^0}{6\,\sigma^3 \sqrt{n}}\,(1 - x^2)\,\varphi(x)}_{\textit{erste Edgeworth-Korrektur}} + o\left(\frac{1}{\sqrt{n}} \right) \end{aligned}$$

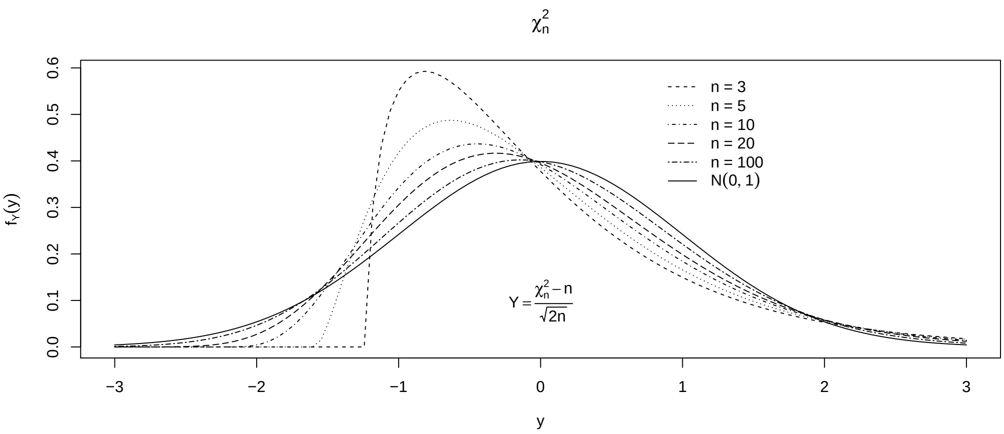


Abbildung 7: Approximation der χ^2-Verteilung durch die Normalverteilung.

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 $\mathrm{N}(0,1)$ schneller als $\frac{1}{\sqrt{n}}$. Die erste Edgeworth-Korrektur ist also für schiefe Verteilungen, etwa $\chi^2(n)$ von Interesse.

Satz 18.7 (Berry-Esseen)
Es seien $X_1, \ldots, X_n$ u.i.v., $X_i \sim F$ mit $\mathbb{E}(X_i) = \mu$, $\mathbb{V}(X_i) = \sigma^2$, $\mathbb{E}(X_i^3) < \infty \implies \exists\, c$ (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 $G_n(x) = \mathbb{P}\left(\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right) \leq x\right)$.

Beweis:

geschenkt $\square$

Bemerkung:

Satz 18.7 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 18.9

$X_1, \ldots, X_n, \quad X_i \sim Po\left(\frac{1}{n}\right) \implies \sum_{i=1}^n X_i \sim Po(1)$

Aber $\quad \sqrt{n}\,\dfrac{\left(\bar{X} - \frac{1}{n}\right)}{\sqrt{\frac{1}{n}}} \quad = \quad \dfrac{\sum X_i - n\,\frac{1}{n}}{\sqrt{n}\,\frac{1}{\sqrt{n}}} \quad = \quad \sum X_i - 1 \quad \sim \quad Po(1) - 1 \quad \neq$ $N(0,1) \quad \forall n$

18.4 Folgerungen aus dem Zentralen Grenzwertsatz

Satz 18.8 (Δ-Methode)

$$\sqrt{n}\,(X_n - \nu) \xrightarrow{\mathcal{D}} N(0, \tau^2) \implies \sqrt{n}\,(f(X_n) - f(\nu)) \xrightarrow{\mathcal{D}} N\left(0, \tau^2 \frac{\partial f(\nu)^2}{\partial \nu}\right)$$

falls $f : \mathbb{R} \to \mathbb{R}$ *in* ν *differenzierbar ist und* $\dfrac{\partial f(\nu)}{\partial \nu} \neq 0.$

Beweis:

Entwickle f in Taylorreihe

$f(x + \Delta) = f(x) + \Delta\, \left.\dfrac{\partial f(x)}{\partial x}\right|_{x=x} + \ldots + \dfrac{\Delta^r}{r!}\, \left.\dfrac{\partial^r f(x)}{\partial^r x}\right|_{x=x} + \text{Rest}$

mit $\Delta = X_n - \nu$ und $x = \nu$ folgt

$f(X_n) = f(\nu) + (X_n - \nu)\,\dfrac{\partial f(\nu)}{\partial \nu} + \text{Rest}$

$\implies \sqrt{n}\,(f(X_n) - f(\nu)) = \underbrace{\sqrt{n}\,(X_n - \nu) \cdot \dfrac{\partial f(\nu)}{\partial \nu}}_{\text{linear}} + \underbrace{\text{Rest}}_{\to 0 \text{ ohne Beweis}}$

$\overset{18.2}{\implies} \sqrt{n}\,(f(X_n) - f(\nu)) \xrightarrow{\mathcal{D}} \mathrm{N}\left(0, \tau^2 \dfrac{\partial f(\nu)^2}{\partial \nu}\right)$ $\square$

Beispiel 18.10

$X_1, \ldots, X_n$ *u.i.v.* $\quad X_i \sim N(0, \sigma^2) \implies \mathbb{E}(X_i^2) = \sigma^2,\ \mathbb{V}(X_i^2) = 2\,\sigma^4$

$\overset{ZGWS}{\implies} \quad \sqrt{n}\left(\dfrac{1}{n}\sum X_i^2 - \sigma^2\right) \xrightarrow{\mathcal{D}} N(0, 2(\sigma^2)^2)$

Gesucht: f *mit* $\dfrac{\partial f(\sigma^2)}{\partial \sigma^2} = \dfrac{c}{\sqrt{2}\,\sigma^2}$

$f = \displaystyle\int \frac{\partial f(\sigma^2)}{\partial \sigma^2}\, d\sigma^2 = \frac{c}{\sqrt{2}} \int \frac{1}{\sigma^2}\, d\sigma^2 = \frac{c}{\sqrt{2}} \log(\sigma^2)$

Setzt $c = 1 \quad \Longrightarrow$

$$\sqrt{\frac{n}{2}} \log \left(\frac{\frac{1}{n} \sum X_i^2}{\sigma^2} \right) \xrightarrow{\mathcal{D}} N(0,1)$$

Definition 18.1
Seien $X_1, \ldots, X_n$ *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* $X_1, \ldots, X_n$*.*

Satz 18.9 (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:

b)
$I_{X_i \leq x} \sim \mathrm{B}(1, F(x))$ u.i.v. $\quad \forall i = 1, \ldots, n$ und $x \in \mathbb{R}$.

$$\overset{\text{ZGWS}}{\Longrightarrow} \quad \underbrace{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^{n} I_{X_i \leq x} - F(x) \right)}_{= \sqrt{n}\,(\widehat{F}_n(x) - F(x))} \xrightarrow{\mathrm{d}} \mathrm{N}(0, F(x)(1 - F(x))) \implies \text{b)}$$

a)
Satz 17.4: $\widehat{F}_n(x) = \frac{1}{n} \sum I_{X_i \leq x} \xrightarrow{\text{f.s.}} \mathbb{E}(I_{X_i \leq x}) = \mathbb{P}(X_i \leq x) = F(x)$
c) ohne Beweis $\square$

Beispiel 18.11 (Glivenko-Cantelli: $\sim \chi^2$)

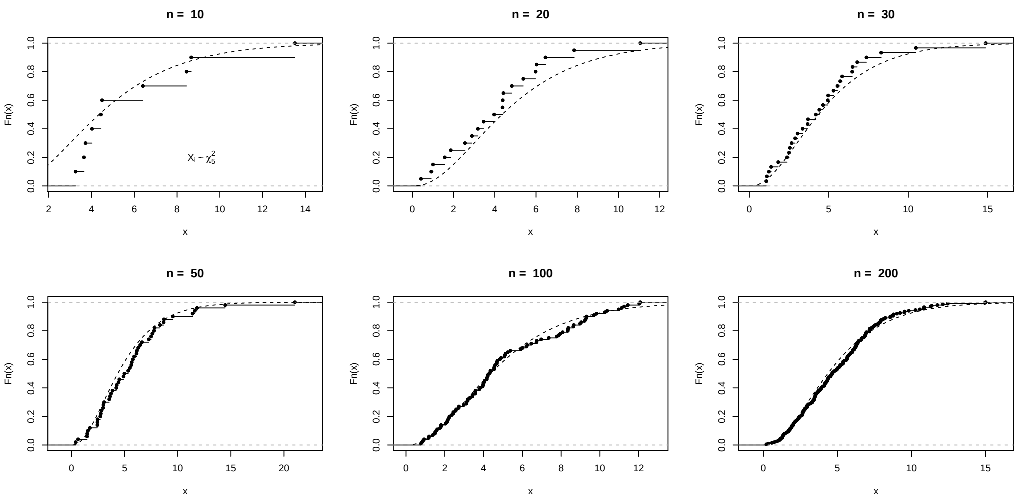


Abbildung 8: Approximation der Verteilungsfunktion der χ^2_5 Verteilung durch die empirische Verteilungsfunktion.

Beispiel 18.12 (Glivenko-Cantelli $\sim Po$)

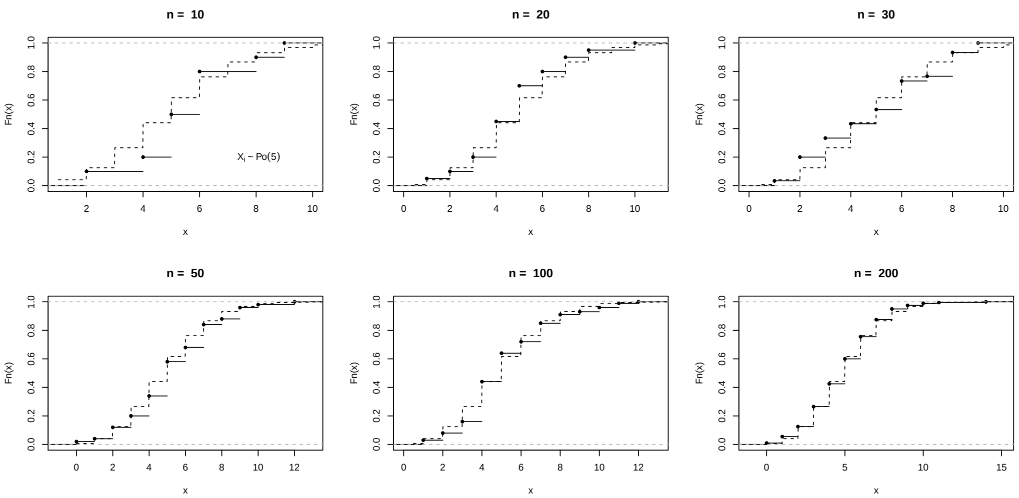


Abbildung 9: Approximation der Verteilungsfunktion der Po(5) Verteilung durch die empirische Verteilungsfunktion.

Beispiel 18.13 (Glivenko-Cantelli $\sim t(1)$)

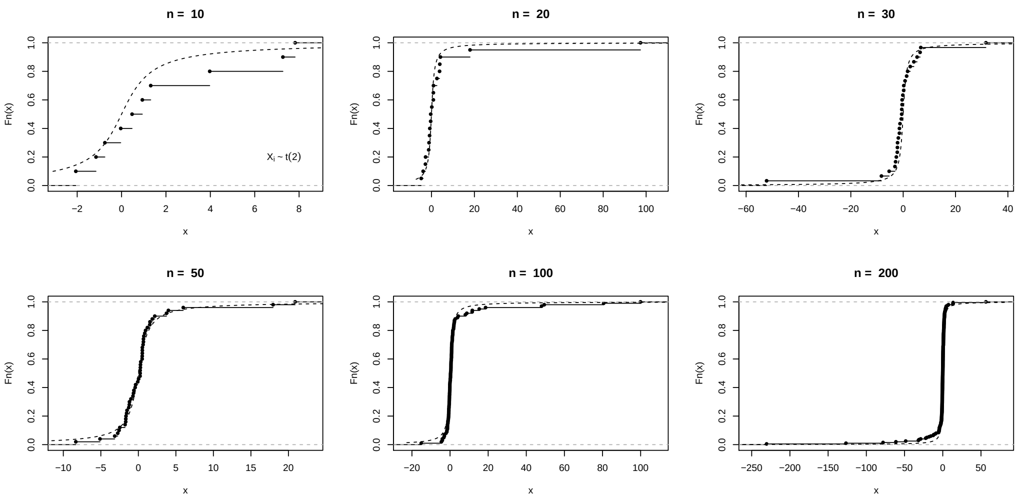


Abbildung 10: Approximation der Verteilungsfunktion der $t(1)$ (Cauchy) Verteilung durch die empirische Verteilungsfunktion.

Teil V
Anhang

A Spezielle diskrete Verteilungen

A.1 Diskrete Gleichverteilung

Zähldichte $f(\omega) := \dfrac{1}{|\Omega|} \; \forall\, \omega$, siehe Bsp. 7.1

A.2 Bernoulli- und Binomialverteilung

$X \sim B(1,p)$: Zähldichte (siehe Bsp. 7.2): $f(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$
$X \sim B(n,p))$: Zähldichte (siehe Bsp. 7.2): $f(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \forall\, x \in 0, \ldots, n, p \in [0,1], n \in \mathbb{N}$

Satz A.1 (Zusammenhang zwischen B$(1,p)$ und B(n,p))
Seien $X_1, \ldots, X_n$ stu mit $X_i \sim B(1,p)$. Dann gilt:

$$X = \sum_{i=1}^{n} X_i \sim B(n,p).$$

Beweis:

$\mathbb{P}(X_1 = 1, X_2 = 1, \ldots, X_m = 1, X_{m+1} = 0, \ldots, X_n = 0) \overset{\text{stu}}{=} \prod_{i=1}^{m} \mathbb{P}(X_i = 1) \prod_{i=m+1}^{n} \mathbb{P}(X_i = 0) = \prod_{i=1}^{m} p \prod_{i=m+1}^{n} (1-p) = p^m\,(1-p)^{n-m}$
Da es $\dbinom{n}{m}$ Möglichkeiten gibt, die 1/0 zu permutieren, folgt

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i = m\right) = \binom{n}{m} p^m\,(1-p)^{n-m} \implies \sum_{i=1}^{n} X_i \sim \mathrm{B}(n,p)$$

□

Korollar A.1
$X_1 \sim B(n_1,p)$, $X_2 \sim B(n_2,p)$ stu. $\implies$ $X_1 + X_2 \sim B(n_1 + n_2, p)$ Summenstabilität

$X \sim B(n,p) : \mathbb{E}(X) = n \cdot p$, siehe Bsp. 8.2; $\mathbb{V}(X) = n\,p\,(1-p)$, siehe 8.5.

A.3 Poisson-Verteilung

$X \sim \text{Po}(\lambda), \lambda \in \mathbb{R}^+$:

Zähldichte (siehe Bsp.7.5): $f(x) = \dfrac{\lambda^x e^{-\lambda}}{x!} \; \forall \, x \in \mathbb{N}_0$

Satz A.2 (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:

$$\begin{aligned}
b(k,n,p_n) &= \binom{n}{k} p_n^k \, (1-p_n)^{n-k} \\
&= \frac{n!}{k!\,(n-k)!} \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k} \\
&= \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \frac{n\,(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda_n}{n}\right)^{-k} \\
&= \frac{\lambda_n^k}{k!} \left(1 - \frac{\lambda_n}{n}\right)^n \frac{\frac{n}{n}\,\frac{n-1}{n}\,\frac{n-2}{n}\cdots\frac{n-k+1}{n}}{\left(1-\frac{\lambda_n}{n}\right)^k} \\
&\overset{n\to\infty}{\to} \frac{\lambda^k}{k!} e^{-\lambda} \underbrace{\frac{1\cdot\left(1-\frac{1}{n}\right)\cdot\ldots\cdot\left(1-\frac{k-1}{n}\right)}{\left(1-\frac{\lambda_n}{n}\right)^k}}_{\to 1 \text{ da } \frac{1}{n}\ldots\frac{\lambda_n}{n}\to 0} \\
&= \frac{\lambda^k}{k!} e^{-\lambda} = f(k) \quad \widehat{=} \quad \text{Dichte der Poissonverteilung mit Parameter } \lambda.
\end{aligned}$$

□

A.4 Hypergeometrische Verteilung

Definition A.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 A.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 $\mathrm{B}(n,p)$ $\Longrightarrow$ *Endlichkeitskorrektur*

Beweis:

Wir betrachten die einzelnen Ziehungen mit $Y_1, \ldots, Y_n$, $X = \sum_{i=1}^n Y_i$, 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.
$\mathbb{P}(Y_1 = i_1, \ldots, Y_n = i_n) = \frac{K \cdots (K-k+1)(N-K) \cdots (N-K-(n-k)+1)}{N(N-1) \cdots (N-n+1)}$ für $\sum_{j=1}^n i_j = k$
$\Longrightarrow$ gemeinsame Verteilung von $Y_1, \ldots, Y_n$ ist vertauschbar: $\mathbb{P}(Y_1, \ldots, Y_n) = \mathbb{P}(Y_{\pi(1)}, \ldots, Y_{\pi(n)})$ für Permutationen π.
$\Longrightarrow$ Alle Randverteilungen sind identisch $Y_i \sim \mathrm{B}(1, \frac{K}{N})$

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^n Y_i\right) &= \sum_{i=1}^n \mathbb{E}(Y_i) = n \cdot \frac{K}{N} \\
\mathbb{V}\left(\sum_{i=1}^n Y_i\right) &= \sum_{i=1}^n \mathbb{V}(Y_i) + \sum_{i \neq j, i<j} 2 \cdot \mathrm{Cov}(Y_i, Y_j)
\end{aligned}$$

Es gilt:

$$\begin{aligned}
\mathrm{Cov}(Y_i, Y_j) &= \mathrm{Cov}(Y_1, Y_2) = \mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1) \cdot \mathbb{E}(Y_2) \\
&= \frac{K}{N} \cdot \frac{K-1}{N-1} - \left(\frac{K}{N}\right)^2 = \frac{K}{N} \cdot \left(\frac{K-1}{N-1} - \frac{K}{N}\right)
\end{aligned}$$

(weil $Y_1Y_2 = 1 \iff Y_1 = 1$ und $Y_2 = 1$ und damit $\mathbb{E}(Y_1Y_2) = \mathbb{P}(Y_1 = 1, Y_2 = 1) = \frac{K}{N} \cdot \frac{K-1}{N-1}$)

$$\begin{aligned}\mathbb{V}(\sum_{i=1}^{n} Y_i) &= n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right) + \frac{n(n-1)}{2} \cdot 2\frac{K}{N}\left(\frac{K-1}{N-1} - \frac{K}{N}\right) \\ &= n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)\end{aligned}$$

□

Satz A.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 $\frac{K_m}{N_m} \to p$ 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:

$$\begin{aligned}\mathbb{P}(H_m = k) &= h(k, n, K_m, N_m) = \frac{\binom{K_m}{k}\binom{N_m-K_m}{n-k}}{\binom{N_m}{n}} \\ &= \frac{n!}{k!\,(n-k)!} \cdot \frac{\frac{K_m!}{(K_m-k)!} \cdot \frac{(N_m-K_m)!}{((N_m-K_m)-(n-k))!}}{\frac{N_m!}{(N_m-n)!}} \\ &= \binom{n}{k} \underbrace{\frac{K_m}{N_m}}_{\to p} \cdot \underbrace{\frac{K_m-1}{N_m-1}}_{\to p} \cdots \underbrace{\frac{K_m-(k-1)}{N_m-(k-1)}}_{\to p} \cdot \\ &\quad \underbrace{\frac{(N_m-K_m)}{N_m-k}}_{\to 1-p} \cdot \underbrace{\frac{N_m-K_m-1}{N_m-k-1}}_{\to 1-p} \cdots \underbrace{\frac{N_m-K_m-(n-k-1)}{N_m-(n-1)}}_{\to 1-p}\end{aligned}$$

$$\overset{m\to\infty}{\Longrightarrow} \binom{n}{k} p^k (1-p)^{n-k} = b(k,n,p) = \mathbb{P}(B = k)$$

□

A.5 Geometrische und Negativ-Binomial-Verteilung

Definition A.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 A.5 (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 A.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 in jedem Versuch 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 A.6 (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 A.2
$X \sim NegBin(n,p) \implies \mathbb{E}(X) = \frac{n}{p}$ $X = \sum_{i=1}^n X_i, \quad X_i \sim Geom(p)$ stu. $\implies \mathbb{E}(X) = n \cdot \mathbb{E}(X_i) = \frac{n}{p}$ und $\mathbb{V}(X) = \frac{n(1-p)}{p^2}$ Negative Binomialverteilung ist nicht gedächtnislos.

Beweis:

Zunächst $n = 2$.

$$
\begin{aligned}
f_{X_1,X_2}(x_1,x_2) &\overset{\text{stu}}{=} (1-p)^{x_1-1} \cdot p \cdot (1-p)^{x_2-1} \cdot p \\
\mathbb{P}(X_1 + X_2 = x) &= \mathbb{P}\left(\{\omega \in \Omega \,|\, X_1(\omega) + X_2(\omega) = x\}\right) \\
&= \mathbb{P}\left(\dot{\bigcup_{x_1 \in \mathbb{N}}} \{\omega \in \Omega \,|\, X_1(\omega) = x_1 \wedge X_2(\omega) = x - x_1\}\right) \\
&= \sum_{x_1 \in \mathbb{N}} \mathbb{P}\left(\{\omega \in \Omega \,|\, X_1(\omega) = x_1 \wedge X_2(\omega) = x - x_1\}\right) \\
&= \sum_{x_1 \in \mathbb{N}} \mathbb{P}(X_1 = x_1,\, X_2 = x - x_1) = \sum_{x_1 \in \mathbb{N}} f_{X_1,X_2}(x_1, x - x_1) \quad \text{"Faltung"} \\
&\overset{\text{stu}}{=} \sum_{x_1 \in \mathbb{N}} f_{X_1}(x_1) \cdot f_{X_2}(x - x_1)
\end{aligned}
$$

Da $f_{X_2}(x - x_1) = 0$ für $x \leq x_1$ kann man vereinfachen:

$$
\begin{aligned}
\mathbb{P}(X_1 + X_2 = x) &= \sum_{x_1=1}^{x-1} (1-p)^{x_1-1}\, p^2\, (1-p)^{x-x_1-1} \\
&= p^2 \sum_{x_1=1}^{x-1} (1-p)^{x_1-1+(x-x_1-1)} = p^2 \sum_{x_1=1}^{x-1} (1-p)^{x-2} \\
&= p^2\,(x-1) \cdot (1-p)^{x-2} \overset{n=2}{=} \binom{x-1}{n-1} p^n\, (1-p)^{x-n} \\
&\widehat{=} \text{Dichte der NegBin}(2,p)
\end{aligned}
$$

Rest: VI □

A.6 Multinomialverteilung

Definition A.4 (Multinomialverteilung)
Sei $n \in \mathbb{N}$ und $p_1, \ldots, p_k \in [0,1]$ mit $\sum_{j=1}^k p_j = 1$. Sei $Y = (Y_1, \ldots, Y_k)$ eine

k-dimensionale diskrete ZV mit Dichte

$$
\begin{aligned}
f_Y(y) &= \mathbb{P}(Y_j = y_j \quad \forall j = 1, \ldots, k) \\
&= \begin{cases} \dfrac{n!}{y_1! \cdots y_k!} \, p_1^{y_1} \cdots p_k^{y_k} & \text{falls } \sum_{j=1}^{k} y_j = n \\ 0 & \text{sonst.} \end{cases}
\end{aligned}
$$

Dann heißt die zugehörige Verteilung <u>*Multinomialverteilung*</u> *mit Parametern* n *und* $p_1, \ldots, p_k$ *:* $\quad Y \sim MN(n, p_1, \ldots, p_k)$*.*

Bemerkung:

Es genügt, $p_1, \ldots, p_{k-1}$ anzugeben, denn

$$p_k = 1 - \sum_{j=1}^{k-1} p_j.$$

Satz A.7 (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)

$$
\begin{aligned}
\mathbb{P}(Y_i = y_i) &= \mathbb{P}(Y_i = y_i, \sum_{j \neq i} Y_j = n - y_i) \\
&\overset{\mathrm{MN}(n, p_i, 1-p_i)}{=} \frac{n!}{y_i! \, (n - y_i)!} \, p_i^{y_i} \, (1 - p_i)^{n - y_i} \\
&= \binom{n}{y_i} \, p_i^{y_i} \, (1 - p_i)^{n - y_i}
\end{aligned}
$$

$\Longrightarrow \; Y_i \sim \mathrm{B}(n, p_i)$

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) $Y_i \sim \mathrm{B}(n, p_i) \implies \mathbb{V}(Y_i) = n\, p_i\, (1 - p_i)$

$$\begin{aligned}\mathbb{V}(Y_i + Y_j) &= \mathbb{V}(Y_i) + \mathbb{V}(Y_j) + 2\,\mathrm{Cov}(Y_i, Y_j) \\ &= n\, p_i\, (1 - p_i) + n\, p_j\, (1 - p_j) + 2\,\mathrm{Cov}(Y_i, Y_j)\end{aligned}$$

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):

$$\begin{aligned}\mathrm{Cov}(Y_i, Y_j) &= \frac{\mathbb{V}(Y_i + Y_j) - n\,(p_i\,(1 - p_i) + p_j\,(1 - p_j))}{2} \\ &= \frac{n\,(p_i + p_j)\,(1 - (p_i + p_j)) - n\,(p_i\,(1 - p_i) + p_j\,(1 - p_j)))}{2} \\ &= \frac{n}{2}\,(p_i\,(1 - (p_i + p_j)) + p_j\,(1 - (p_i + p_j)) - p_i\,(1 - p_i) - p_j\,(1 - p_j)) \\ &= \frac{n}{2}\,(p_i\,(1 - (p_i + p_j) - 1 + p_i) + p_j\,(1 - (p_i + p_j) - 1 + p_j)) \\ &= \frac{n}{2}(-p_i\,p_j - p_j\,p_i) = -\frac{n}{2}\,2\,p_i\,p_j = -n\,p_i\,p_j\end{aligned}$$

□

B Spezielle stetige Verteilungen

B.1 Stetige Gleichverteilung

$X \sim \mathrm{U}[a, b]$:

Dichte (siehe Bsp. 7.6): $f(x) = \dfrac{1}{b-a} I_{]a,b[}(x) \qquad \forall\, x \in [a, b]$

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

$$\mathbb{E}(X) = \frac{a+b}{2},$$
$$\mathbb{V}(X) = \frac{(b-a)^2}{12}.$$

Siehe Bsp. 8.6.

B.2 Exponential- und Gammaverteilung

Einfachste Annahme für Wartezeiten.
$X \sim \text{Exp}(\lambda), \lambda \in \mathbb{R}^+$ (siehe Bsp. 7.8):

$$f(x) = \begin{cases} 0 & x \leq 0 \\ \lambda \exp(-\lambda x) & x > 0 \end{cases}$$
$$F(x) = \begin{cases} 1 - \exp(-\lambda x) & x > 0 \\ 0 & \text{sonst} \end{cases}$$

$\mathbb{P}(X > t + s \,|\, X > t) = \mathbb{P}(X > s)$ (siehe Bsp. 15.1).
$X \sim \text{Ga}\left(k, \lambda\right), k > 0, \lambda > 0$, Dichte (siehe 7.4:

$$f_X(x; k, \lambda) = \frac{\lambda}{\Gamma(k)} \left(\lambda x\right)^{k-1} \exp(-\lambda x) \cdot I_{(0,\infty)}(x)$$

Satz B.1 (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.*

Bemerkung:

$$\text{Ga}\left(\frac{k}{2}, \frac{1}{2}\right) = \chi^2\left(k\right)$$

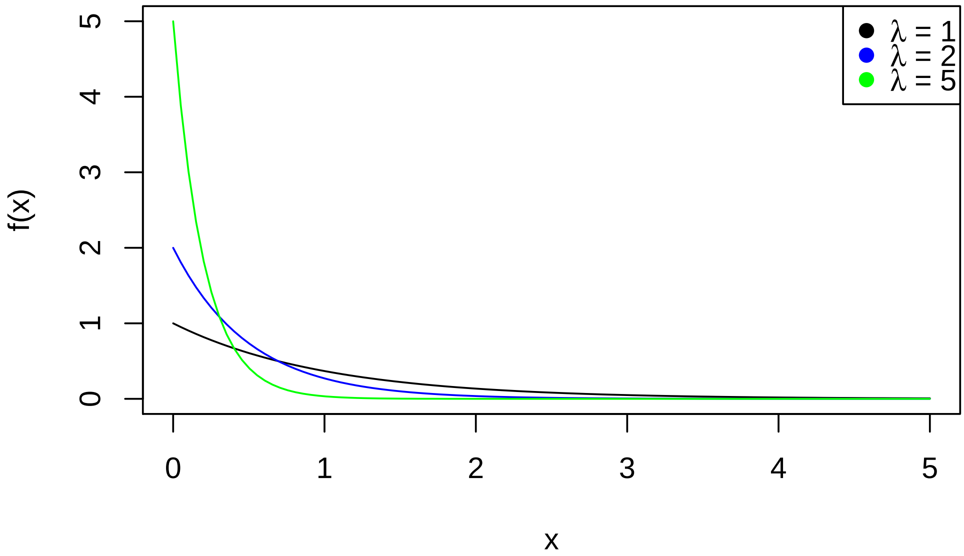


Abbildung 11: Exponential-Verteilung: Dichten

Satz B.2
$X \sim Ga(k, \lambda) \implies$

$$\begin{aligned} \mathbb{E}(X) &= \frac{k}{\lambda} \\ \mathbb{V}(X) &= \frac{k}{\lambda^2} \\ \textit{und} \quad M_X(s) &= \left(\frac{\lambda}{\lambda - s}\right)^k \quad \textit{für } s < \lambda. \end{aligned}$$

Beweis:

$$
\begin{aligned}
M_X(s) &= \mathbb{E}(\exp(sX)) \\
&= \int\limits_0^\infty \frac{\lambda}{\Gamma(k)} (\lambda x)^{k-1} \exp(-\lambda x) \exp(sx)\, dx \\
&= \int\limits_0^\infty \frac{\lambda^k}{\Gamma(k)} x^{k-1} \exp((-\lambda + s)\, x)\, dx \\
&= \lambda^k \int \frac{1}{\Gamma(k)} x^{k-1} \exp(-(\lambda - s)\, x)\, dx \\
&\overset{\text{Kreative 1}}{=} \frac{\lambda^k}{(\lambda - s)^k} \int \underbrace{\frac{(\lambda - s)^k}{\Gamma(k)} x^{k-1} \exp(-(\lambda - s)\, x)}_{\text{Dichte Ga}(k,\lambda-s)} dx \\
&= \frac{\lambda^k}{(\lambda - s)^k}
\end{aligned}
$$

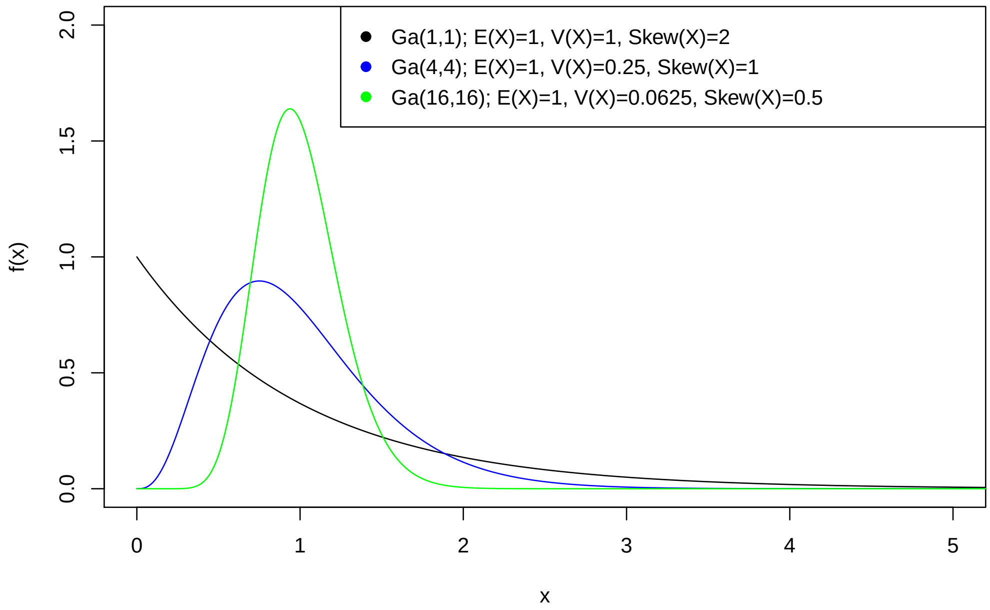
$$
\begin{aligned}
\implies \left.\frac{\partial M_X(s)}{\partial s}\right|_{s=0} &= k \cdot \lambda^k \cdot (\lambda - s)^{-k-1}\big|_{s=0} \\
&= \frac{k \cdot \lambda^k}{\lambda^{k+1}} = \frac{k}{\lambda} = \mathbb{E}(X) \\
\left.\frac{\partial^2 M_X(s)}{\partial^2 s}\right|_{s=0} &= k\,(k+1)\ \lambda^k\,(\lambda - s)^{-k-2}\big|_{s=0} \\
&= \frac{k\,(k+1)}{\lambda^2} = \mathbb{E}(X^2)
\end{aligned}
$$

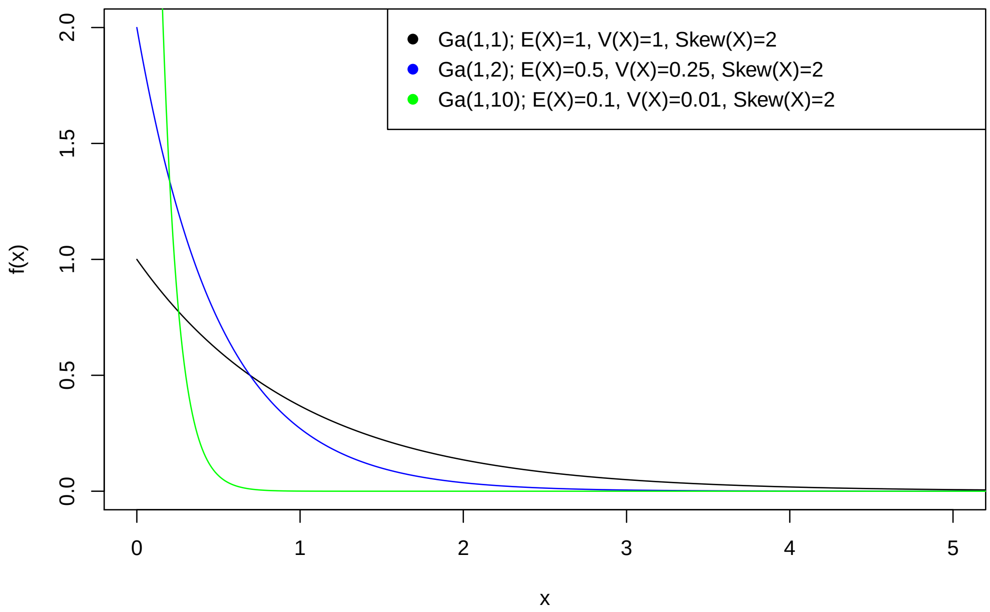
Zusammen: $\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{k\,(k+1)}{\lambda^2} - \frac{k^2}{\lambda^2} = \frac{k}{\lambda^2}$ $\square$

Bemerkung:

$X \sim \Gamma(k, \lambda) \implies$

$$
\begin{aligned}
Skew(X) &= \frac{2}{\sqrt{k}} \text{ (Schiefe)} \\
K(X) &= 3 + \frac{6}{k} \text{ (Kurtosis)}
\end{aligned}
$$

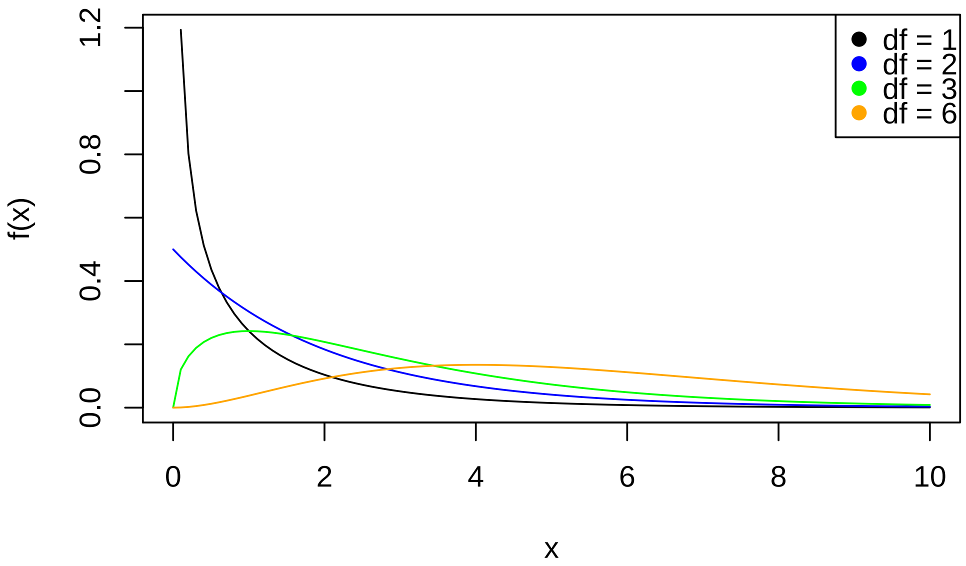

Ga(1,1); E(X)=1, V(X)=1, Skew(X)=2
Ga(4,4); E(X)=1, V(X)=0.25, Skew(X)=1
Ga(16,16); E(X)=1, V(X)=0.0625, Skew(X)=0.5
f(x)
x
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5


Ga(1,1); E(X)=1, V(X)=1, Skew(X)=2
Ga(1,2); E(X)=0.5, V(X)=0.25, Skew(X)=2
Ga(1,10); E(X)=0.1, V(X)=0.01, Skew(X)=2
f(x)
x
0.0
0.5
1.0
1.5
2.0
0
1
2
3
4
5

B.3 χ^2-Verteilung

$X \sim \chi^2(k)$ (Def. 11.1):

$$f_X(x;k) = \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}} x^{\frac{k}{2}-1} \exp\left(-\frac{1}{2}\,x\right) \cdot I_{(0,\infty)}(x)$$



Bemerkung:

$$\begin{aligned}
\int_0^\infty f_X(x)\,dx &= \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \int_0^\infty \frac{x^{\frac{k}{2}-1}}{2^{\frac{k}{2}}} \exp\left(-\frac{1}{2}\,x\right)\,dx \\
&= \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \int_0^\infty \frac{1}{2}\left(\frac{x}{2}\right)^{\frac{k}{2}-1} \exp\left(-\frac{x}{2}\right)\,dx \\
&\overset{t=\frac{x}{2},\,dx=dt\cdot 2}{=} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \cdot \underbrace{\int_0^\infty t^{\frac{k}{2}-1} \exp(-t)\,dt}_{\Gamma\left(\frac{k}{2}\right)} = 1
\end{aligned}$$

Satz B.3
$X \sim \chi^2(k) \implies \mathbb{E}(X) = k;\ \mathbb{V}(X) = 2k;\ M_X(s) = \left(\frac{1}{1-2s}\right)^{\frac{k}{2}}\ \ s < \frac{1}{2}$

Beweis:

Satz B.2 mit $k = \frac{k}{2}$ und $\lambda = \frac{1}{2}$. $\square$

Satz B.4
$X_1, \ldots, X_n$ 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:

$$\begin{aligned} M_Y(s) &= \prod_{i=1}^{n} \mathbb{E}(\exp(s\,X_i^2)) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{1-2s}} = \underbrace{\left(\frac{1}{1-2s}\right)^{\frac{n}{2}}}_{\substack{\text{Momentengenerierende Fkt.} \\ \text{von } \chi^2(n) \text{ nach Satz B.3}}} \end{aligned}$$

□

Satz B.5
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

$$f_X(x; m, n) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\,\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1+\left(\frac{m}{n}\right)\cdot x\right)^{\frac{m+n}{2}}} \cdot I_{(0,\infty)}(x).$$

Definition B.1 (F-Verteilung)
Eine ZV X mit Dichte f_X wie in Satz B.5 heißt F-verteilt mit Freiheitsgraden m und n: $\quad X \sim F(m,n)$.

Beweis:

Sei $c = \frac{m}{2}$, $d = \frac{n}{2}$, also $f_U(u) = \frac{u^{c-1}}{\Gamma(c)} \exp\left(-\frac{1}{2}u\right) I_{(0,\infty)}(u)$, genauso f_V.

$$\begin{aligned} F_{U/V}(x) &= P\left(\frac{U}{V} \leq x\right) = P\left(U \leq xV\right) \\ &= \int_0^\infty \int_0^{xv} f_{U,V}(u,v)\,du\,dv \overset{\text{stu}}{=} \int_0^\infty \left(\int_0^{xv} f_U(u) f_V(v)\,du\right) dv \\ &= \frac{1}{\Gamma(c)\Gamma(d)2^{c+d}} \int_0^\infty \left(\int_0^{xv} u^{c-1} e^{-u/2}\,du\right) v^{d-1} e^{-v/2}\,dv \end{aligned}$$

mit Stammfunktion G von g gilt: $\frac{d}{dx}\left(\int_0^{xy} g(t)\,dt\right) = \frac{d}{dx}\left(G(xy) - G(0)\right) = yg(xy)$, also:

$$\begin{aligned} f_{U/V}(x) = \frac{d}{dx} F_{U/V}(x) &= \frac{1}{\Gamma(c)\Gamma(d)2^{c+d}} \int_0^\infty \left(v(vx)^{c-1} e^{-(vx)/2}\right) v^{d-1} e^{-v/2}\,dv \\ &= \frac{x^{c-1}}{\Gamma(c)\Gamma(d)2^{c+d}} \frac{\Gamma(c+d)}{\left(\frac{x+1}{2}\right)^{c+d}} \underbrace{\int_0^\infty \frac{\left(\frac{x+1}{2}\right)^{c+d}}{\Gamma(c+d)} v^{c+d-1} e^{-((x+1)/2)v}\,dv}_{\equiv 1 \text{ da Dichte der } \Gamma\left(c+d, \frac{x+1}{2}\right)}. \end{aligned}$$

Also:

$$f_{U/V}(x) = \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} \frac{x^{c-1}}{(x+1)^{c+d}}$$

und mit

$$\begin{aligned} f_X(x) = f_{\frac{U/m}{V/n}}(x) &= f_{\frac{n}{m}U/V}(x) = \tfrac{m}{n} f_{U/V}(\tfrac{m}{n}x) \\ &= \frac{m}{n} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{(\frac{m}{n}x)^{\frac{m}{2}-1}}{(\frac{m}{n}x+1)^{\frac{m+n}{2}}} \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \frac{x^{\frac{m-2}{2}}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \end{aligned}$$

□

Satz B.6

$$\begin{aligned} X \sim F(m,n) \implies \mathbb{E}(X) &= \frac{n}{n-2} \text{ \textit{für} } n > 2 \text{ \textit{und}} \\ \mathbb{V}(X) &= \frac{2\,n^2\,(m+n-2)}{m\,(n-2)^2\,(n-4)} \text{ \textit{für} } n > 4. \end{aligned}$$

Beweis:

$$\mathbb{E}(X) = \mathbb{E}\left(\tfrac{U/m}{V/n}\right) \stackrel{\text{stu}}{=} \left(\frac{m}{n}\right)^{-1} \mathbb{E}(U) \cdot \mathbb{E}\left(\frac{1}{V}\right) = \left(\frac{m}{n}\right)^{-1} \cdot m \cdot \underbrace{\mathbb{E}\left(\frac{1}{V}\right)}_{=\frac{1}{n-2} \text{ o.B.}} = \frac{n}{n-2}$$

□

Bemerkung:

$$X \sim \mathrm{F}\,(m,n) \implies \frac{1}{X} \sim \mathrm{F}\,(n,m)$$

B.4 Beta-Verteilung

$X \sim \mathrm{Beta}(\alpha,\beta), \alpha > 0, \beta > 0$:
Dichte (siehe Def. 12.5):

$$f_X(x;\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \cdot I_{(0,1)}(x)$$

Bemerkung:

Für unabhängige $X \sim \Gamma(\alpha, \lambda)$ und $Y \sim \Gamma(\beta, \lambda)$ gilt: $\frac{X}{X+Y} \sim \text{Beta}(\alpha, \beta)$

Bemerkung:

$X \sim \text{Beta}(\alpha, \beta) \implies$

$$\begin{aligned}
E(X) &= \frac{\alpha}{\alpha+\beta} \\
V(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
Skew(X) &= \frac{2(\beta-\alpha)\sqrt{\alpha+\beta+1}}{(\alpha+\beta+2)\sqrt{\alpha\beta}} \text{ (Schiefe)} \\
K(X) &= \frac{6\left((\alpha-\beta)^2(\alpha+\beta+1)-\alpha\beta(\alpha+\beta+2)\right)}{\alpha\beta(\alpha+\beta+2)(\alpha+\beta+3)} \text{ (Kurtosis)}
\end{aligned}$$

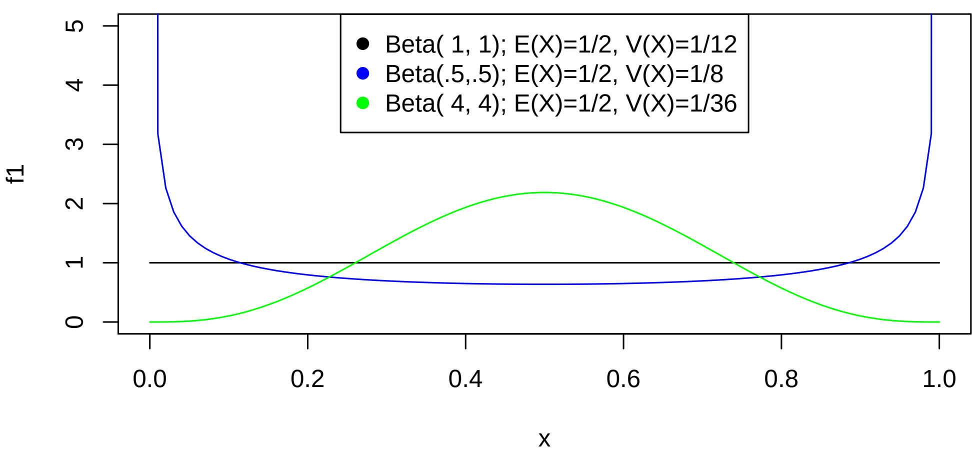


Abbildung 12: Beta-Verteilung: Dichten

B.5 Student-t-Verteilung

Definition B.2 (t-Verteilung)
Eine ZV X mit Dichte

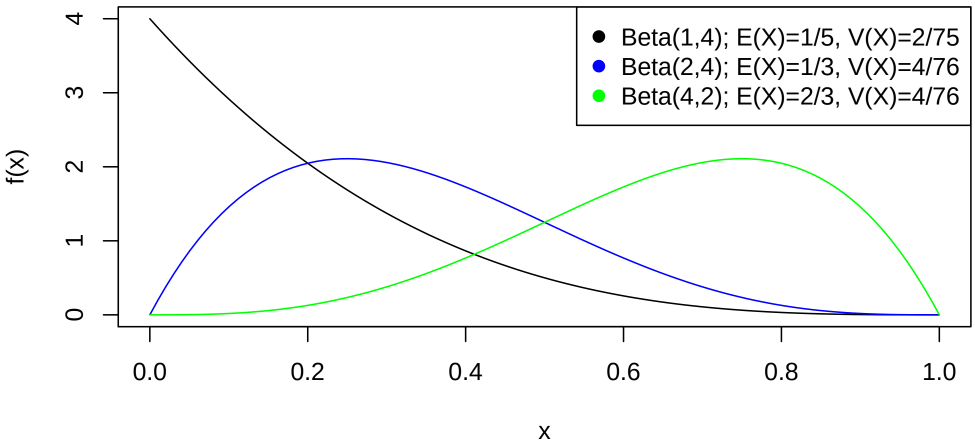


Abbildung 13: Beta-Verteilung: Dichten

$$f_X(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \frac{1}{\sqrt{k\pi}} \frac{1}{\left(1+\frac{x^2}{k}\right)^{\frac{k+1}{2}}}.$$

heißt t-verteilt mit k Freiheitsgraden: $\quad X \sim t(k)$.

Bemerkung:

Eine ZV $X \sim \mathrm{t}(1)$ entspricht Cauchy-Verteilung, siehe Def. 11.2.

Satz B.7
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:

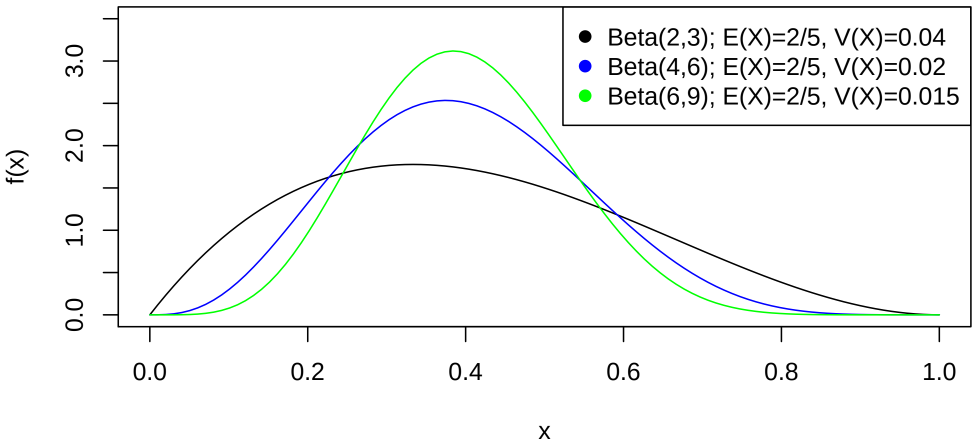


Abbildung 14: Beta-Verteilung: Dichten

$$f_{Z,U}(z,u) = \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma\left(\frac{k}{2}\right)} \left(\frac{1}{2}\right)^{\frac{k}{2}} u^{\frac{k}{2}-1} \exp\left(-\frac{1}{2}\,u\right) \exp\left(-\frac{1}{2}\,z^2\right) I_{(0,\infty)}(u)$$

mit $g(z,u) = (x,y) = \left(\frac{z}{\sqrt{u/k}}, u\right)$ und $h(x,y) = g^{-1}(x,y) = \left(\sqrt{\frac{y}{k}}x, y\right)$ dann

also $J = \begin{pmatrix} \sqrt{\frac{y}{k}} & \sqrt{\frac{1}{k}}y^{-1/2} \\ 0 & 1 \end{pmatrix} \implies |J| = \sqrt{\frac{y}{k}}$

$$\begin{aligned} f_X(x) &= \int_0^\infty f_{Z,U}\left(x\sqrt{\tfrac{y}{k}}, y\right) \sqrt{\tfrac{y}{k}}\, dy \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{k}{2})2^{k/2}} \int_0^\infty \exp\left(-\frac{\left(x\frac{y^{1/2}}{k^{1/2}}\right)^2}{2}\right) y^{(k/2)-1} \exp\left(-(\frac{y}{2})\right) \frac{y^{1/2}}{k^{1/2}}\, dy \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{k}{2})2^{k/2}k^{1/2}} \int_0^\infty \underbrace{y^{\frac{k+1}{2}-1} \exp\left(-\tfrac{1}{2}\left(1+\frac{x^2}{k}\right)y\right)}_{\propto \text{ Dichte } \Gamma(\tilde{k},\lambda) \text{ mit } \tilde{k}=\frac{k+1}{2};\, \lambda=\frac{1}{2}\left(1+\frac{x^2}{k}\right)}\, dy \end{aligned}$$

Also "kreative 1" und Multiplikation mit $\frac{\Gamma(\tilde{k})\lambda^{\tilde{k}}}{\Gamma(\tilde{k})\lambda^{\tilde{k}}}$, damit Integral $= 1$ und

$$\begin{aligned}
f_X(x) &= \frac{\Gamma\left(\frac{k+1}{2}\right)}{\left(\frac{1}{2}\left(1+\frac{x^2}{k}\right)\right)^{(k+1)/2}} \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(\frac{k}{2}) 2^{k/2} k^{1/2}} \\
&\quad \underbrace{\int_0^\infty \frac{\left(\frac{1}{2}\left(1+\frac{x^2}{k}\right)\right)^{(k+1)/2}}{\Gamma\left(\frac{k+1}{2}\right)} y^{\frac{k+1}{2}-1} \exp\left(-\frac{1}{2}\left(1+\frac{x^2}{k}\right) y\right) dy}_{\equiv 1} \\
&= \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{(2\pi)^{1/2} 2^{k/2} k^{1/2}} \frac{2^{(k+1)/2}}{\left(1+\frac{x^2}{k}\right)^{(k+1)/2}} \\
&= \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \frac{1}{\sqrt{\pi k}} \left(1+\frac{x^2}{k}\right)^{-(k+1)/2}
\end{aligned}$$

□

Satz B.8
$X \sim t(k) \implies$

$$\begin{aligned}
m_n &= 0, \ \textit{falls } k > n \textit{ und } n \textit{ ungerade} \\
m_n &= k^{n/2} \cdot \frac{1 \cdot 3 \cdot 5 \cdot 7 \ldots m_{n-1}}{(k-2)\cdot(k-4)\cdot(k-6)\ldots m_{k-n}}, \\
&\quad \textit{falls } k > n \textit{ und } n \textit{ gerade}
\end{aligned}$$

Beweis:

a) Symmetrie in x

b) geschenkt.

□

Bemerkung:

Beweis für ungerade (zentrale) Momente über Symmetrie der Verteilung. Speziell:
$\mathbb{E}(X) = 0$ für $k > 1$, existiert nicht (!) für $k \leq 1$. $\mathbb{V}(X) = \dfrac{k}{k-2}$ für $k > 2$, existiert nicht (!) für $k \leq 2$.

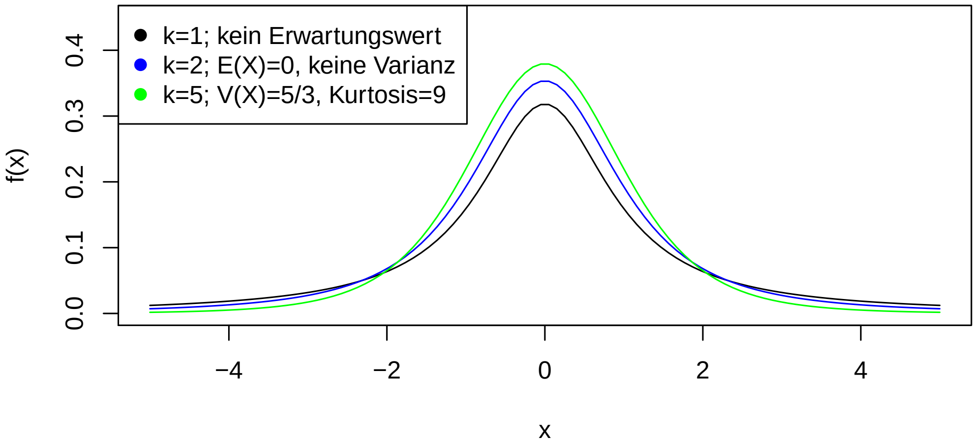


Abbildung 15: t-Verteilung: Dichten

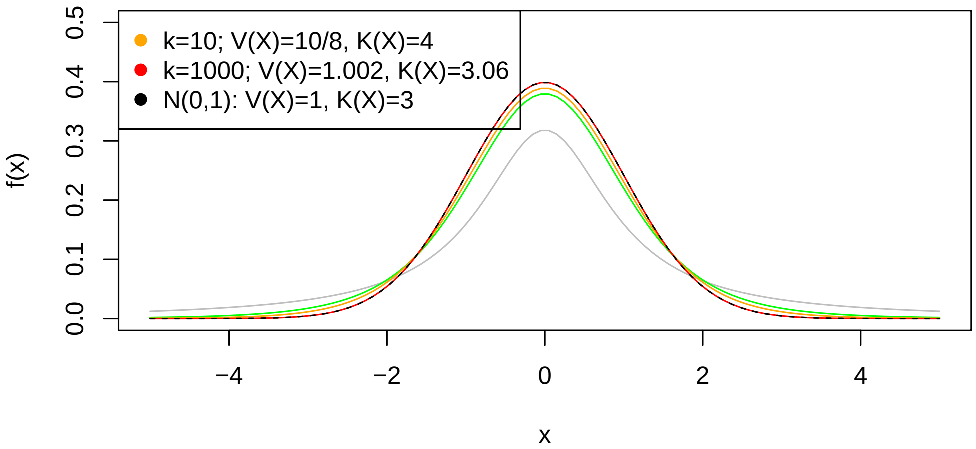


Abbildung 16: t-Verteilung: Dichten

Literatur

R. Hable. *Einführung in die Stochastik: Ein Begleitbuch zur Vorlesung*. Springer, 2015.

E. Lehmann. *Elements of Large-Sample Theory*. Springer, 2001.

D. Meintrup and S. Schäffler. *Stochastik: Theorie und Anwendungen (Statistik und ihre Anwendungen)*. Springer, 2005.

K. Schmidt. *Maß und Wahrscheinlichkeit*. Springer, 2011.

A. Sklar. Random variables, distribution functions, and copulas – a personal look backward and forward. In L. Rüschendorf, B. Schweizer, and M. Taylor, editors, *Distributions With Fixed Marginals & Related Topics*. 1997.

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 und Fabian Scheipl. Das ursprüngliche Skript basiert auf Meintrup and Schäffler (2005).