Kapitel 1

Maßräume und Wahrscheinlichkeitsräume

1.1 Diskrete Wahrscheinlichkeitsräume

Wahrscheinlichkeitsräume dienen der Modellierung von Zufallsexperimenten, also von Experimenten, deren Ausgänge zufälligen Einflüssen unterliegen. Wir bezeichnen mit Ω die nicht-leere Menge aller (möglichen) Ausgänge eines Zufallsexperiments. Ω wird *Stichprobenraum*, *Ergebnismenge* oder *Grundraum* genannt. Die Elemente $\omega \in \Omega$ heißen Ergebnisse.
Wir setzen in diesem Abschnitt voraus, dass Ω höchstens abzählbar ist.

1.1 Beispiel. (i) Beim Würfeln mit einem echten Würfel ist $\Omega = \{1, 2, 3, 4, 5, 6\}$ ein geeigneter Stichprobenraum.

(ii) Beim wiederholten Werfen mit einem Würfel kann man das Experiment „Würfeln bis zum ersten Mal eine 6 fällt'"betrachten. Hierbei bietet sich $\Omega = \mathbb{N} = \{1, 2, 3, ...\}$ an, da beliebig häufig andere Zahlen auftreten können, bis zum ersten Mal eine 6 fällt.

Unter der Annahme, dass Ω höchstens abzählbar ist, wollen wir nun jedem Ergebnis eine Wahrscheinlichkeit für sein Auftreten zuordnen:

1.2 Definition. Sei $\Omega \neq \varnothing$und höchstens abzählbar und $\rho : \Omega \to \mathbb{R}$ mit:

(i) $\rho(\omega) \geq 0$ für alle $\omega \in \Omega$

(ii) $\sum\limits_{\omega\in\Omega} \rho(\omega) = 1$

Dann wird ρ eine *Zähldichte* genannt und (Ω, ρ) ein *diskretes Zufallsexperiment.* Wir interpretieren $\rho(\omega)$ als die Wahrscheinlichkeit für das Eintreten des Ergebnisses ω.

Oft ist man an Wahrscheinlichkeiten von Mengen von Ergebnissen interessiert, etwa dass beim Würfeln eine gerade Zahl fällt (hier: $\{2, 4, 6\}$). Allgemein bezeichnen wir Teilmengen von Ω als *Ereignisse.* Wir sagen das Ereignis A *tritt ein*, falls das Ergebnis ω des Zufallsexperiments Element von A ist. Wichtige Ereignisse sind:

- $\varnothing$: Unmögliches Ereignis
- Ω: Sicheres Ereignis
- $\{\omega\}$: Elementarereignisse (für $\omega \in \Omega$)

Sind A und B Ereignisse so interpretiert man:

Eintreten von	Interpretation
$A \cup B$	Mindestens eines der Ereignisse A oder B tritt ein.
$A \cap B$	A und B treten gemeinsam ein.
$B \setminus A$	B tritt ein, A aber nicht.
A^c	A tritt nicht ein.

A und B werden *unvereinbar* genannt, falls $A \cap B = \varnothing$.

Gegeben ein diskretes Zufallsexperiment (Ω, p) betrachten wir die Abbildung

$$\mathcal{P} : 2^\Omega \to \mathbb{R}, A \mapsto \sum_{\omega\in A} \rho(\omega) \tag{1.1}$$

Hier ist 2^Ω die Potenzmenge von Ω und wir betrachten $\mathcal{P}(A)$ als Wahrscheinlichkeit für das Eintreten von A.

1.3 Motivation. Man beachte, dass für endliches Ω von der Häufigkeitsinterpretation nahegelegt wird: Würfelt man etwa n-mal mit einem echten Würfel, so ergibt sich die relative Häufigkeit für das Auftreten gerader Zahlen als Summe der relativen Häufigkeiten für das Auftreten der Zahlen 2,4 bzw. 6, d.h.

$$\frac{\#\text{„gerader Zahl“}}{n} = \frac{\#\text{„2“}}{n} + \frac{\#\text{„4“}}{n} + \frac{\#\text{„6“}}{n}$$

Die Definition in (1.1) für Wahrscheinlichkeiten macht für überabzählbares Ω i.a. aufgrund der Summenbildung keinen Sinn. Zur Verallgemeinerung ist die folgende äquivalente Formulierung hilfreich:

1.4 Satz. Sei $\Omega \neq \varnothing$ höchstens abzählbar und $\mathcal{P} : 2^\Omega \to \mathbb{R}$.Dann sind äquivalent:

(i) Es gibt eine Zähldichte ρ auf Ω so, dass (1.1) gilt.

(ii) $\mathcal{P}$ erfüllt folgende Eigenschaften:

a) $\mathcal{P}(A) \geq 0$ für alle $A \in 2^\Omega$

b) $\mathcal{P}(\Omega) = 1$

c) $\mathcal{P}$ ist *σ-additiv*, d.h. für alle Folgen $(A_n)_{n\in\mathbb{N}}$ in 2^Ω gilt:

$$A_n \cap A_m = \varnothing \text{ für alle } n, m \in \mathbb{N} \text{ mit } n \neq m \Rightarrow \mathcal{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mathcal{P}(A_n)$$

Beweis. $(i) \Rightarrow (ii)$: a) und b) folgen direkt aus und der Definition der Zähldichte. Zu c): Seien $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt. Dann:

$$\sum_{n=1}^{\infty} \mathcal{P}(A_n) = \sum_{n=1}^{\infty} \sum_{\omega\in A_n} \rho(\omega) = \sum_{\omega\in \bigcup_{n=1}^{\infty} A_n} \rho(\omega) = \mathcal{P}\Big(\bigcup_{n=1}^{\infty} A_n\Big),$$

wobei die Vertauschung der Summationsreihenfolge aufgrund der Nicht-Negativität der Zähldichte durch den großen Umordnungssatz gerechtfertigt ist.

$(ii) \Rightarrow (i)$: Sei $\rho(\omega) := \mathcal{P}(\{\omega\})$. Aufgrund der σ-Additivität gilt für jedes $A \in 2^\Omega$:

$$\mathcal{P}(A) = \sum_{\omega\in A} \mathcal{P}(\{\omega\}) = \sum_{\omega\in A} \rho(\omega),$$

also gilt (1.1). ρ ist eine Zähldichte, da $\rho(\omega) = \mathcal{P}(\{\omega\}) \geq 0$ nach a) und $\sum_{\omega\in\Omega} \rho(\omega) = \mathcal{P}(\Omega) = 1$ nach b).

□

Erinnerung. Der große Umordnungssatz besagt (siehe Königsberger, Analysis 1):
Sei $\mathbb{N} = \bigcup_{j\in\mathbb{N}} T_j$ für disjunkte Mengen $(T_j)_{j\in\mathbb{N}}$ und $(a_n)_{n\in\mathbb{N}}$ eine Folge von reellen Zahlen mit $\sum_{n=1}^{\infty} |a_n| < \infty$.
Sei

$$b_j = \begin{cases} 0, & T_j = \varnothing, \\ \sum_{n\in T_j} a_n, & T_j \neq \varnothing. \end{cases}$$

Dann: $\sum_{n=1}^{\infty} a_n = \sum_{j=1}^{\infty} b_j$.

1.5 Definition. Sei $\Omega \neq \varnothing$ höchstens abzählbar.

(i) Eine Abbildung $\mathcal{P} : 2^\Omega \to \mathbb{R}$, die die Eigenschaften aus Satz 1.4 (ii) erfüllt, wird *diskretes Wahrscheinlichkeitsmaß* auf $(\Omega, 2^\Omega)$ genannt.

(ii) Ein Tripel $(\Omega, 2^\Omega, \mathcal{P})$, wobei $\mathcal{P}$ ein diskretes W-Maß auf $(\Omega, 2^\Omega)$ ist, heißt *diskretes Wahrscheinlichkeitsmaß*.

1.2 Allgemeine Wahrscheinlichkeitsräume

Häufig treten Zufallsexperimente mit überabzählbarem Stichprobenraum auf.

1.6 Beispiel. (i) Wird ein Münzwurf beliebig oft ausgeführt, so bildet sich als Stichprobenraum

$$\Omega = \{0,1\}^{\mathbb{N}} = \{(\omega_n)_{n\in\mathbb{N}} | \omega_n \in \{0,1\} \textit{ für alle } n \in \mathbb{N}\}$$

an.

(ii) Wird ein Glücksrad gedreht, so interessiert man sich für den Winkel $\phi \in (0, 2\pi]$. Hier ist das Intervall $(0, 2\pi]$ ein geeigneter Stichprobenraum.

Bei überabzählbaren Stichprobenräumen ist die Potenzmenge häufig zu groß, um dort W-Maße zu definieren, die ein Zufallsexperiment geeignet beschreiben. Für den wiederholten Münzwurf gilt etwa:

1.7 Satz. Sei $\Omega = \{0,1\}^{\mathbb{N}}$. Dann existiert **keine** σ-additive Abbildung $\mathcal{P} : 2^{\Omega} \to \mathbb{R}$ mit $\mathcal{P}(A) \geq 0$ für alle $A \in 2^{\Omega}, \mathcal{P}(\Omega) = 1$, die die folgende Invarianzeigenschaft erfüllt: Für alle $n \geq 1$ und $A \in 2^{\Omega}$ gilt:

$$\mathcal{P}(T_n A) = \mathcal{P}(A),$$

wobei

$$T_n : \Omega \to \Omega, (\omega_1, \omega_2, ...) \mapsto (\omega_1, ..., \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, ...).$$

Beweis. Wir nehmen an, es gibt eine solche Abbildung. Für $\omega, \omega' \in \Omega$ sei

$$\omega \sim \omega' :\Leftrightarrow \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \omega_n = \omega'_n.$$

Man prüft leicht nach, dass $\sim$ eine Äquivalenzrelation definiert. Nach dem Auswahlaxiom existiert eine Menge $A^* \subset \Omega$, die aus jeder Äquivalenzklasse genau ein Element enthält. Sei

$$\varphi = \{S \subset \mathbb{N} \,|\, \text{S hat höchstens endlich viele Elemente}\}.$$

φ ist abzählbar unendlich als disjunkte Vereinigung der endlichen Mengen $\{S \in \varphi | max\varphi = m\}, m \in \mathbb{N}$ und $\{\varnothing\}$. Für $S = \{n_1, ..., n_k\} \in \varphi$ sei $T_S := T_{n_1} \circ T_{n_k}$ mit der Konvention $T_\varnothing = id$. Dann gilt:

a) $\Omega = \bigcup\limits_{S\in\varphi} T_S A^*$, denn zu jedem $\omega \in \Omega$ existiert $\omega' \in A^*$ mit $\omega \sim \omega'$ und also $\omega = T_S \omega' \in T_S A^*$ für ein $S \in \varphi$.

b) Die Mengen $(T_S A^*)_{S\in\varphi}$ sind paarweise disjunkt. Denn: Ist $T_S A^* \cap T_{S'} A^* \neq \varnothing$ für $S, S' \in \varphi$, so existieren $\omega, \omega' \in A^*$ mit $T_S \omega = T_{S'} \omega'$. Es folgt: $\omega = \omega'$ und also $S = S'$.

c) $\mathcal{P}(T_S A^*) = \mathcal{P}(A^*)$.

Die σ-Additivität impliziert zusammen mit a), b) und c):

$$\mathcal{P}(\Omega) = \mathcal{P}(\bigcup_{S\in\varphi} T_S A^*) = \sum_{S\in\varphi} \mathcal{P}(T_S A^*) = \sum_{S\in\varphi} \mathcal{P}(A^*) \in \{o, \infty\}$$

im Widerspruch zu $\mathcal{P}(\Omega) = 1$. □

Der Ausweg besteht darin, nicht mehr allen Teilmengen von Ω Wahrscheinlichkeiten zuordnen zu wollen, sondern nur noch „geeigneten"' Teilmengen.

1.8 Definition. Sei $\Omega \neq \varnothing$. Ein Mengensystem $\mathcal{A} \subset 2^{\Omega}$ wird σ-Algebra genannt, falls gilt:

- $\Omega \in \mathcal{A}$
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A} \Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}$

1.9 Beispiel. (i) 2^{Ω} ist eine σ-Algebra.

(ii) $\{\varnothing, \Omega\}$ ist eine σ-Algebra, die sog. *triviale σ-Algebra.*

1.10 Lemma. Seien $(\mathcal{A}_i)_{i\in I}$ σ-Algebren auf Ω, wobei I eine nicht-leere Menge ist. Dann ist $\bigcap\limits_{i\in I} \mathcal{A}_i$ eine σ-Algebra auf Ω.

Beweis.
- $\Omega \in \mathcal{A}_i$ für alle $i \in I$. Also $\Omega \in \bigcap\limits_{i\in I} \mathcal{A}_i$.
- $A \in \mathcal{A}_i$ für alle $i \in I$. $\Rightarrow A^c \in \mathcal{A}_i$ für alle $i \in I$.

- $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}_i$ für alle $i \in I$. $\Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{A}_i$ für alle $i \in I$.

$\square$

1.11 Definition. Sei $\Omega \neq \varnothing$ und $\mathcal{M} \subset 2^\Omega$ ein System von Teilmengen von Ω. Dann wird

$$\sigma(\mathcal{M}) = \cap\{\mathcal{A} \mid \mathcal{A}\ \sigma - Algebra\ auf\ \Omega\ mit\ \mathcal{M} \in \mathcal{A}\}$$

die von $\mathcal{M}$ *erzeugte σ-Algebra* genannt.

Lemma 1.10. stellt sicher, dass $\sigma(\mathcal{M})$ eine σ-Algebra auf Ω definiert.

1.12 Definition. Sei $\Omega = \mathbb{R}^D$. Wir notieren mit

$$[a,b) := [a_1,b_1) \times ... \times [a_D,b_D),$$

$a = (a_1,...,a_D) \in \mathbb{R}^D, b = (b_1,...,b_D) \in \mathbb{R}^D$ die nach rechts halboffenen Intervalle in $\mathbb{R}^D$. Entsprechend sind die Notationen $(a,b], [a,b], (a,b)$ zu lesen. Dann erzeugen die Mengensysteme

$$J_1 := \{(a,b] \mid a,b \in \mathbb{R})^D\},$$

$$J_2 := \{[a,b) \mid a,b \in \mathbb{R})^D\},$$
$$J_3 := \{(a,b) \mid a,b \in \mathbb{R})^D\},$$
$$J_4 := \{[a,b] \mid a,b \in \mathbb{R})^D\}$$

allesamt die gleiche σ-Algebra, die *Borelsche σ-Algebra* auf $\mathbb{R}^D$ genannt wird und die wir mit $\mathcal{B}_D$ notieren. $\mathcal{B}_D$ wird auch von den offenen Mengen im $\mathbb{R}^D$ erzeugt (Übung).

1.13 Definition. (i) Sei $\Omega \neq \varnothing$ und $\mathcal{A}$ eine σ-Algebra auf Ω. Dann wird $(\Omega, \mathcal{A})$ ein *Messraum* genannt.

(ii) Sei $(\Omega, \mathcal{A})$ ein Messraum. $\mathcal{P} : \mathcal{A} \to \mathbb{R}$ wird *Wahrscheinlichkeitsmaß* auf $(\Omega, \mathcal{A})$, falls:

a) $\mathcal{P}(\Omega) = 1$

b) $\mathcal{P}(A) \geq 0$ für alle $A \in \mathcal{A}$

c) $\mathcal{P}$ ist σ-additiv, d.h. für alle Folgen $(A_n)_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen in $\mathcal{A}$ gilt:

$$\mathcal{P}(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{\infty} \mathcal{P}(A_n).$$

(iii) Ist $(\Omega, \mathcal{A})$ ein Messraum und $\mathcal{P}$ ein W-Maß auf $(\Omega, \mathcal{A})$, so wird $(\Omega, \mathcal{A}, \mathcal{P})$ ein *Wahrscheinlichkeitsraum* genannt.

1.14 Bemerkung. Sei $\mathcal{P}$ ein W-Maß auf $(\Omega, \mathcal{A})$ und $A \in \mathcal{A}$. Dann folgt aus der σ-Additivität:

$$1 = \mathcal{P}(\Omega) = \mathcal{P}(A \cup A^c \cup \varnothing \cup \varnothing \cup ...) = \mathcal{P}(A) + \mathcal{P}(A^c) + \sum_{i=1}^{\infty} \mathcal{P}(\varnothing).$$

Also folgt $\mathcal{P}(\varnothing) = 0$ und $\mathcal{P}(A) \leq 1$ für alle $A \in \mathcal{A}$.

1.15 Definition. Sei $(\Omega, \mathcal{A})$ ein Messraum. Eine Abbildung $\mu : \mathcal{A} \to \mathbb{R} \cup \{+\infty\}$ wir ein *Maß* auf $(\Omega, \mathcal{A})$ genannt, falls die Eigenschaften b) und c) aus Der Definition eines W-Maßes erfüllt sind. Und statt a) gilt

$$a')\quad \mu(\varnothing) = 0.$$

Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt dann ein *Maßraum*.
Ein Maß μ heißt *endlich*, falls $\mu(\Omega) < \infty$ und *σ-endlich*, falls eine Folge $(B_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ exisitiert mit $\mu(B_n) < \infty$ für alle $n \in \mathbb{N}$ und $\Omega = \bigcup\limits_{n\in\mathbb{N}} B_n$.

1.3 Konstruktion von Maßen

Motivation. Sei $\Omega \neq \varnothing$. Häufig ist uns klar, wie wir ein Maß auf gewissen Teilmengen von Ω definieren wollen. So kann man z.B. den Intervallen in $\mathbb{R}$ die Intervalllänge als „Maß“ zuordnen.

Frage. Unter welchen Bedingungen ist es möglich, ein auf gewissen Teilmengen festgelegtes Maß auf eine σ-Algebra auszudehnen?

Wir betrachten zuerst die Situation, dass $\Omega = \mathbb{R}$ ist. Dann wird jede nicht fallende, rechtsstetige Funktion $F : \mathbb{R} \to \mathbb{R}$ eine *maßerzeugende Funktion* genannt. Gilt ferner $\lim\limits_{x\to\infty} F(x) = 1$ und $\lim\limits_{x\to-\infty} F(x) = 0$, dann heißt F eine *Verteilungsfunktion.*

Sei $J_1 := \{(a, b] \mid a \leq b \in \mathbb{R}\}$ die Menge aller linksoffenen Intervalle in $\mathbb{R}$. Ziel ist es eine σ-Algebra $\mathcal{A}$ mit $J_1 \subset \mathcal{A}$ zu konstruieren und eine Maß μ auf $\mathcal{A}$ zu definieren, so dass $\mu((a, b]) = F(b) - F(a)$. Sei $\mathcal{F}$ das Mengensystem, das aus allen endlich disjunkten Vereinigungen von Mengen aus J_1 besteht. Man überlegt sich leicht (Übung), dass $\mathcal{F}$ einen Ring im Sinne folgender Definition bildet:

1.16 Definition. Sei $\Omega \neq \varnothing$ und $\mathcal{R} \subset 2^\Omega$. $\mathcal{R}$ wir ein *Ring* auf Ω geannt, falls

- $\varnothing \in \mathcal{R}$
- $A, B \in \mathcal{R} \Rightarrow A \setminus B \in \mathcal{R}$ und $A \cup B \in \mathcal{R}$

Fordert man zudem $\Omega \in \mathcal{R}$, so heißt $\mathcal{R}$ eine *Algebra.*

1.17 Definition. Sei $\mathcal{R}$ ein Ring auf Ω. $\mu_0 : \mathcal{R} \to \mathcal{R} \cup \{+\infty\}$ wir ein *Inhalt* genannt, falls

- $\mu_0(\varnothing) = 0$
- $\mu_0(A) \geq 0$ für alle $A \in \mathcal{R}$
- $A, B \in \mathcal{R}$ disjunkt $\Rightarrow \mu_0(A \cup B) = \mu_0(A) + \mu_0(B)$ (endliche Additivität)

Ist μ_0 sogar σ-additiv, d.h. für alle Folgen $(A_N)_{n\in\mathbb{N}}$ von paarweise disjunkten Mengen in $\mathcal{R}$ gilt:

$$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{R} \Rightarrow \mu_0(\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n=1}^{\infty} \mu_0(A_n),$$

so wird μ_0 ein *Prämaß* genannt.

1.18 Satz. Sei $F : \mathbb{R} \to \mathbb{R}$ eine maßerzeugende Funktion. Dann existiert genau ein Inhalt $\mu_0 : \mathcal{F} \to \mathbb{R}$ mit $\mu_0((a, b]) = F(a) - F(b)$ für alle $a \leq b$.

Beweis.

- **Eindeutigkeit:** Sei $\mu_0 : \mathcal{F} \to \mathbb{R}$ ein Inhalt. Sei $A \in \mathcal{F}$ mit $A = I_1 \cup ... \cup I_k$, wobei $I_1, ..., I_k$ disjunkt aus J_1. Dann folgt aufgrund der Additivität:

$$\mu_0(A) = \mu_0(\bigcup_{k=1}^{K} I_k) = \sum_{k=1}^{K} \mu_0(I_k).$$

Jeder Inhalt auf $\mathcal{F}$ ist also durch die Werte auf den linksoffenen Intervallen eindeutig bestimmt.

- **Konstruktion von μ_0:** Für $I = (a, b] \in J_1$ definieren wir $\mu_0((a, b]) = F(b) - F(a)$. Sei $(a, b]$ dargestellt als endliche Vereinigung von Mengen aus J_1, also $(a, b] = \bigcup\limits_{n=1}^{N-1} (a_n, a_{n+1}]$ mit $a = a_1 \leq a_2 \leq ... \leq a_n = b$. Dann:

$$(*) \quad \mu_0((a, b]) = F(a) - F(b) = \sum_{n=1}^{N-1} (F(a_{n+1})) - F(a_n)) = \sum_{n=1}^{N-1} \mu_0((a_n, a_{n+1}]).$$

Sei nun $A \in \mathcal{F}$ und seien zwei Zerlegungen von A als Vereinigung disjunkter Mengen in J_1 gegeben: $A = \bigcup\limits_{k=1}^{k} I_k = \bigcup\limits_{m=1}^{M} \tilde{I}_m$. Dann folgt aus $(*)$:

$$\mu_0(I_k) = \mu_0(I_k \cap A) = \mu_0(\bigcup_{m=1}^{M} (I_k \cap \tilde{I}_m)) = \sum_{m=1}^{M} \mu_0(I_k \cap \tilde{I}_m) \text{ und } \mu_0(\tilde{I}_m) = \sum_{k=1}^{K} \mu_0(I_k \cap \tilde{I}_m).$$

Somit

$$\sum_{k=1}^{K} \mu_0(I_k) = \sum_{m=1}^{M} \mu_0(\tilde{I}_m).$$

Also ist die Abbildung

$$\mu_0 : \mathcal{F} \to \mathbb{R}_k, A \mapsto \sum_{k=1}^{K} \mu_0(I_k),\ I_1, ..., I_k \in J_1 \text{ disjunkt mit } \bigcup_{k=1}^{K} I_k = A$$

wohldefiniert.

- μ_0 **ist ein Inhalt:**
  - $\mu_0(\varnothing) = \mu_0((0,0]) = F(0) - F(0) = 0$
  - Aufgrund der Konstruktion reicht es, $A \in J_1$ zu betrachten. Für $A \in J_1$ folgt die Nichtnegativität aus der Monotonie von F.
  - Additivität: Seien $A, B \in \mathcal{F}$ disjunkt mit Darstellungen $A = \bigcup_{k=1}^{K} I_k, B = \bigcup_{m=1}^{M} \tilde{I}_m$ als disjunkte Vereinigungen von Mengen aus J_1. Dann ist $A \cup B = (\bigcup_{k=1}^{K} I_k) \cup (\bigcup_{m=1}^{M} \tilde{I}_m)$ eine disjunkte Vereinigung. Also:

$$\mu_0(A \cup B) = \sum_{k=1}^{K} \mu_0(I_k) + \sum_{m=1}^{M} \mu_0(\tilde{I}_m) = \mu_0(A) + \mu_0(B).$$

□

Um zu zeigen, dass das eben konstruierte μ_0 ein Prämaß ist, verwenden wir das folgende Lemma:

1.19 Lemma. Für einen Inhalt μ_0 auf einem Ring $\mathcal{R}$ gilt:

(a) μ_0 ist ein Prämaß.

$\Leftrightarrow (b)$ μ_0 ist stetig nach unten, d.h. für jede Folge $(A_n)_{n\in\mathbb{N}} \in \mathcal{R}$ mit $A_n \subset A_{n+1}$ für alle $n \in \mathbb{N}$ und $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$ gilt:

$$\mu_0(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu_0(A_n).$$

$\Rightarrow (c)$ μ_0 ist stetig nach oben, d.h. für jeden Folge (A_n) in $\mathcal{R}$ mit $A_n \supset A_{n+1}$ für alle $n \in \mathbb{N}$, $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{R}$ und $\mu_0(A_1) < \infty$ gilt:

$$\mu_0(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mu_0(A_n).$$

$\Leftrightarrow (d)$ μ_0 ist stetig von oben in $\varnothing$, d.h. für jede Folge $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{R}$ mit $A_n \supset A_{n+1}$ für alle $n \in \mathbb{N}$, $\bigcap_{n\in\mathbb{N}} A_n = \varnothing$ und $\mu_0(A_1) < \infty$ gilt:

$$\lim_{n\to\infty} \mu_0(A_n) = 0.$$

Ist μ_0 endlich, d.h. $\mu_0(A) < \infty$ für alle $A \in \mathcal{R}$, so gilt auch $(c) \Rightarrow (b)$.

Beweis. Übung. □

1.20 Satz. Der in Satz 1.18 konstruierte Inhalt ist ein Prämaß.

Beweis. Der in Satz 1.18 konstruierte Inhalt ist endlich. Nach Lemma 1.19 reicht es also, die Stetigkeit von oben in $\varnothing$ zu zeigen.
Sei $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{F}$ mit $A_n \supset A_{n+1}$ mit $\lim_{n\to\infty} \mu_0(A_n) = \delta > 0$. Dann ist zu zeigen, dass $\bigcap_{n\in\mathbb{N}} A_n \neq \varnothing$.

Schritt 1: Zu jedem $\epsilon > 0$ und jedem $A \in \mathcal{F}$ existiert ein $B \in \mathcal{F}$ mit $\overline{B} \subset A$ und $\mu_0(B) \geq \mu_0(A) - \epsilon$. Nach Definition von $\mathcal{F}$ und μ_0 muss der Beweis nur für $A = (a, b]$ mit $a < b$ geführt werden. Sei $B = (c, b]$ für ein $c \in (a, b)$. Dann gilt $\overline{B} \subset A$ und

$$\mu_0(B) = F(b) - F(c) \to F(b) - F(a) = 13$$

für $c \searrow a$ wegen der Rechtsstetigkeit von $\mathcal{F}$. Insbesondere ist $\mu_0(B) \geq \mu_0(A) - \epsilon$ falls c hinreichend nahe an a ist.

Schritt 2: Zu jedem A_n sei $B_n \in \mathcal{F}$ so gewählt, dass $\overline{B}_n \subset A_n$ und $\mu_0(B_n) \geq \mu_0(A_n) - 2^{-n}\delta$. Derartige B_n existieren nach Schritt 1. Sei $C_n = B_1 \cap ... \cap B_n$. Wir zeigen mittels Induktion nach n, dass

$$\mu_0(C_n) \geq \mu_0(A_n) - \delta(1 - 2^{-n}).$$

$n = 1$**:** Klar nach Konstruktion.

$n-1 \to n$**:** Da μ_0 endlich ist, gilt:

$$\mu_0(C_n) = \mu_0(C_{n-1} \cap B_n) = \mu_0(C_{n-1}) + \mu_0(B_n) - \mu_0(C_{n-1} \cup B_n).$$

Da $B_n \cup C_{n-1} \subset A_n \cup A_{n-1} = A_{n-1}$, folgt aus der Induktionsannahme

$$\begin{aligned}\mu_0(C_n) &\geq \mu_0(A_{n-1}) - \delta(1 - 2^{-(n-1)}) + \mu_0(A_n) - 2^{-n}\delta - \mu_0(A_{n-1}) \\ &= \mu_0(A_n) - \delta(1 - 2^{-(n-1)} + 2^{-n}) \\ &= \mu_0(A_n) - \delta(1 - 2^{-n}).\end{aligned}$$

Schritt 3: Wir zeigen, dass $\bigcap\limits_{n\in\mathbb{N}} \overline{C}_n \neq \varnothing$.

Da nach Annahme $\mu_0(A_n) \geq \delta$, folgt aus Schritt 2:

$$\mu_0(C_n) \geq 2^{-n}\delta > 0.$$

Also ist jede der Mengen C_n nicht-leer und also $\overline{C}_n \neq \varnothing$ für alle $n \in \mathbb{N}$. Da die Mengen A_n beschränkt sind und $\overline{C}_n \subset A_n$, ist $(\overline{C}_n)_{n\in\mathbb{N}}$ eine Folge von kompakten Mengen in $\mathbb{R}$ mit $\overline{C_n} \supset \overline{C_{n+1}}$ für alle $n \in \mathbb{N}$. Wähle $x_n \in \overline{C}_n$ für jedes $n \in \mathbb{N}$. Dann ist $(x_n)_{n\in\mathbb{N}}$ eine beschränkte Folge in $\mathbb{R}$ und hat also eine konvergente Teilfolge $(x_{n_k})_{k\in\mathbb{N}}$ mit Grenzwert x. Für jedes $m \in \mathbb{N}$ und $n_k \geq m$ gilt: $x_{n_k} \in \overline{C}_m$. Die Kompaktheit von $\overline{C}_m$ impliziert, dass $x \in \overline{C}_m$ für alle $m \in \mathbb{N}$. Insbesondere $\bigcap\limits_{m\in\mathbb{N}} \overline{C}_m \neq \varnothing$.

Schritt 3 schließt den Beweis ab, da

$$\varnothing \neq \bigcap_{m\in\mathbb{N}} \overline{C}_m \subset \bigcap_{m\in\mathbb{N}} A_m.$$

□

Gegeben eine maßerzeugende Funktion F auf $\mathbb{R}$, so haben wir ein Prämaß μ_0 auf dem Ring $\mathcal{F}$ konstruiert mit $\mu_0((a,b]) = F(b) - F(a)$. Dieses Prämaß wollen wir zu einem Maß auf der Borelschen σ-Algebra ausweiten. Dazu:

1.21 Satz. (Ausweitungssatz von Carathéodory)
Sei μ_0 ein Prämaß auf einem Ring $\mathcal{R}$ auf Ω. Dann gilt:

(i) Es existiert ein Maß μ auf $\sigma(\mathcal{R})$ mit $\mu(A) = \mu_0(A) \, \forall A \in \mathcal{R}$.

(ii) Ist das Prämaß μ_0 σ-endlich, so ist das Maß in (i) eindeutig bestimmt.

Beweisskizze folgt später.

Als Folgerung ergibt sich:

1.22 Satz. Sei $F : \mathbb{R} \to \mathbb{R}$ eine maßerzeugende Funktion. Dann existiert genau ein Maß μ_F auf der borelschen σ-Algebra $\mathcal{B}$ mit $\mu_F((a,b]) = F(b) - F(a) \, \forall a \leq b \in \mathbb{R}$.
Ferner gilt: Für eine weitere maßerzeugende Funktion G gilt $\mu_F = \mu_G$ genau dann, wenn eine Konstante $c \in \mathbb{R}$ existiert mit $F(x) = G(x) + c$ für alle $x \in \mathbb{R}$.

Beweis. Da $J_1 = \{(a,b] \mid a \leq b \in \mathbb{R}\} \subset \mathcal{F} \subset \mathcal{B}$ folgt $\mathcal{B} = \sigma(J_1) \subset \sigma(\mathcal{F}) \subset \sigma(\mathcal{B}) = \mathcal{B}$. Die von $\mathcal{F}$ erzeugte σ-Algebra auf $\mathbb{R}$ ist also die Borelsche σ-Algebra. Existenz und Eindeutigkeit des Maßes μ_F folgen dann aus Satz 1.20 und dem Satz von Carathéodory.
Ist $G = F+c$, so folgt $\mu_F = \mu_G$ aus der Eindeutigkeitsaussage des Satzes von Carathéodory. Ist andererseits $\mu_F = \mu_G$, so folgt $c := G(0)-F(0)$ für alle $x \geq 0$. $G(x) = \mu_G((0,x])+G(0) = \mu_F((0,x])+F(0)+c = F(x)+c$. Analog folgt der Fall $x < 0$ mit Hilfe von $(x,0]$. □

1.23 Bemerkung. Die maßerzeugende Funktion $\mathcal{F}(x) = x$ ($x \in \mathbb{R}$) führt zum Lebesgue-Maß auf $\mathcal{B}$, das wir mit λ bezeichnen.

1.24 Folgerung. Die Abbildung {Verteilungsfunktion auf $\mathbb{R}$} $\to$ {ω-Maße auf $(\mathbb{R},\mathcal{B})$}, $F \mapsto \mu_F$ (definiert in Satz 1.22) ist bijektiv.

Beweis.

- μ_F ist ein ω-Maß da:

$$\mu_F(\mathbb{R}) = \mu_F(\bigcup_{n\in\mathbb{N}}(-n,n]) \overset{1.13}{=} \lim_{n\to\infty} \mu_F((-n,n]) = \lim_{n\to\infty}(F(n) - F(-n)) = 1 - 0 = 1.$$

- Injektivität: Sei $\mu_F = \mu_G$ für Verteilungsfunktionen F und G. Nach Satz 1.22 existiert ein $c \in \mathbb{R}$, so dass für alle $x \in \mathbb{R}$ gilt: $G(x) = F(x) = c$. Aus $x \to -\infty$ folgt $c = 0$.

- Surjektivität: Sei μ ein ω-Maß aif $(\mathbb{R},\mathcal{B})$. Definiere $F(x) := \mu((-\infty,x])$. Dann gilt für alle $a \le b \in \mathbb{R}$:

$$\mu((a,b]) = \mu((-\infty,b]) \setminus (-\infty,a]) = \mu((-\infty,b]) - \mu((-\infty,a]) = F(b) - F(a.)$$

- Es bleibt zu zeigen, dass F eine Verteilungsfunktion ist:
  - (i) Monotonie: Klar.
  - (ii) Rechtsstetigkeit: Sei $x \in \mathbb{R}$ und $x_n \searrow x$. Dann:

$$F(x) = \mu((-\infty,x]) = \mu(\bigcap_{n\in\mathbb{N}}(-\infty,x_n]) \overset{1.19}{=} \lim_{n\to\infty} \mu((-\infty,x_n]) = \lim_{n\to\infty} F(x_n).$$

  - (iii) Normierung: $F(-\infty) = 0, F(+\infty) = 1$ folgt wie in (ii) mit

$$\mathbb{R} = \bigcup_{n\in\mathbb{N}}(-\infty,x_n] \text{ für } x_n \nearrow +\infty \text{ und } \varnothing = \bigcap_{n\in\mathbb{N}}(-\infty,-x_n] \text{ für } x_n \searrow -\infty.$$

□

1.25 Bemerkung. Eine Abbildung $F : \mathbb{R}^D \to \mathbb{R}$ *maßerzeugend*, falls gilt:

- F ist *rechtsstetig*, d.h. für alle Folgen $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^D$ mit $x_n \ge x_{n+1}$ für alle $n \in \mathbb{N}$ („ $\ge$ “ komponentenweise) und mit $\lim\limits_{n\to\infty} x_n = x \in \mathbb{R}$ gilt

$$F(x) = \lim_{n\to\infty} F(x_n).$$

- F ist $\triangle$*-monoton*, d.h. $\forall a \le b \in \mathbb{R}^D$ gilt

$$\triangle_a^b F := \sum_{x\in\{0,1\}^D} (-1)^{\sum\limits_{d=1}^{D} x_d} F(a_1^{x_1} b_1^{1-x_1}, ..., a_D^{x_D} b_D^{1-x_D}) \ge 0.$$

Ist F ferner normiert in dem Sinn, dass

- $\lim\limits_{n\to\infty} F(x_n) = 0$ falls $x_{n+1} \le x_n$ und $x_{n,d} \to -\infty$ für $d = 1, ..., D$
- $\lim\limits_{n\to\infty} F(x_n) = 0$ falls $x_{n+1} \ge x_n$ und $x_{n,d} \to +\infty$ für $d = 1, ..., D$

so wird F eine *D-dimensionale Verteilungsfunktion* genannt. Analog zum Fall $D = 1$, kann man zeigen:

(1) Zu jeder maßerzeugenden Funktion $F : \mathbb{R}^D \to \mathbb{R}$ existiert genau ein Maß μ_F auf $(\mathbb{R}^D, \mathcal{R}_D)$ mit $F((a,b]) = \triangle_a^b F \,\forall a \le b \in \mathbb{R}^D$.

(2) Die Abbildung $\{D\text{-dimensionale Verteilungsfunktion}\} \to \{\omega\text{-Maße auf } (\mathbb{R}^D, \mathcal{B}_D)\}$, $F \mapsto \mu_F$ ist bijektiv.

Beweis. (Beweisskizze zu Satz 1.21)

(i) Existenz: Sei μ_0 ein Prämaß auf $\mathbb{R}$. Zu $A \subset \Omega$ sei

$$\mu^*(A) = \inf\{\sum_{n=1}^{\infty} \mu_0(B_n) \,|\, (B_n)_{n\in\mathbb{N}} \text{ in } \mathbb{R} \text{ mit } A \subset \bigcup_{n\in\mathbb{N}} B_n\} (*)$$

mit der Konventioninf $\varnothing = +\infty$. Dann gelten:

- $\mu^*(\varnothing) = 0$
- $\mu^*(A_1) \le \mu * (A_2)$ für $A_1 \subset A_2 \subset \Omega$
- $\mu^*(\bigcup\limits_{n\in\mathbb{N}} A_n) \le \sum\limits_{n=1}^{\infty} \mu^*(A_n)$ $A_n \subset \Omega$ für alle $n \in \mathbb{N}$

Eine Abbildung $\mu^* : 2^\Omega \to \mathbb{R}$ mit diesen drei Eigenschaften nennt man ein *äußeres Maß*. Eine Menge $A \subset \Omega$ heißt μ^**-messbar*, falls

$$\mu^*(Q) \ge \mu^*(Q \cap A) + \mu^*(Q \cap A^c) \,\forall Q \subset \Omega.$$

Sei $\mathcal{A}^* = \{A \subset \Omega \,|\, A \text{ ist } \mu^*\text{-messbar}\}$. Es gilt dann für das in $(*)$ definierte äußere Maß:

(1) $\mathcal{R} \subset \mathcal{A}^*$

(2) $\mu^*(A) = \mu_0(A) \,\forall A \in \mathcal{R}$

Es reicht dann zu zeigen, dass:

(3) $\mathcal{A}^*$ ist eine σ-Algebra.

(4) $\mu : \mathcal{A}^* \to \mathbb{R} \cup \{+\infty\}, A \mapsto \mu^*(A)$ ist ein Maß.

Dies zu zeigen ist nicht schwierig, aber technisch.

(ii) Eindeutigkeit: Der Beweis der Eindeutigkeit beruht auf dem Konzept des Dynkin-Systems: Ein System $\mathcal{D}$ von Teilmengen von Ω wird *Dynkin-System* genannt, falls

- $\Omega \in \mathcal{D}$
- $A \in \mathcal{D} \Rightarrow A^c \in \mathcal{D}$
- $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ mit paarweise disjunkten $A_i \Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{D}$ □

Den Zusammenhang zwischen Dynkin-System und σ-Algebra erläutert der folgende Satz:

1.26 Satz. (i) Ein Dynkin-System $\mathcal{D}$ auf Ω ist genau dann eine σ-Algebra, falls mit je zwei Mengen $A, B \in \mathcal{D}$ auch $A \cap B \in \mathcal{D}$.

(ii) Ist ε ein Mengensystem auf Ω, das abgeschlossen gegen endliche Durchschnitssabbildungen ist, d.h. $A, B \in \varepsilon \Rightarrow A \cap B \in \varepsilon$, so gilt für das von ϵ erzeugte Dynkin-System $\delta(\varepsilon)$: $\delta(\varepsilon) = \sigma(\varepsilon)$.

Beweis. (i) Übung

(ii) Da jede σ-Algebra ein Dynkin-System ist, folgt $\delta(\varepsilon) \subset \sigma(\varepsilon)$. Es reicht also zu zeigen, dass $\delta(\varepsilon)$ eine σ-Algebra ist, also nach (i), dass $\delta(\varepsilon)$ $\cap$-stabil ist.
Sei $\varepsilon_1 := \{B \in \delta(\varepsilon) \,|\, A \cap B \in \delta(\varepsilon) \; \forall\, A \in \varepsilon\}$. Nach Voraussetzung gilt $\varepsilon \subset \varepsilon_1$. Man prüft leicht nach, dass ε_1 ein Dynkin-System ist. Also folgt $\delta(\varepsilon) \subset \varepsilon_1 \subset \delta(\varepsilon)$, d.h. $\varepsilon_1 = \delta(\varepsilon)(*)$.
Sei nun $\varepsilon_2 := \{B \in \delta(\varepsilon) \,|\, A \cap B \in \delta(\varepsilon) \; \forall\, A \in \delta(\varepsilon)\}$. Dann ist auch ε_2 ein Dynkin-System und nach $(*)$ gilt $\varepsilon \subset \varepsilon_2$ und $\delta(\varepsilon) \subset \varepsilon_2 \subset \delta(\varepsilon)$. Somit $\delta(\varepsilon) = \varepsilon_2$, d.h. $A \cap B \in \delta(\varepsilon) \; \forall\, B, A \in \delta(\varepsilon)$. □

Zurück zum Satz von Carathéodory: Sei μ_0 ein σ-endliches Prämaß auf $\mathcal{R}$ und $(B_n)_{n\in\mathbb{N}}$ eine aufsteigende Folge von Mengen in $\mathcal{R}$ mit $\bigcup\limits_{n\in\mathbb{N}} B_n = \Omega$ und $\mu_0(B_n) < \infty$. Seien $\mu_1, \mu_2 : \sigma(\mathcal{R}) \to \mathbb{R} \cup \{+\infty\}$ Fortsetzungen von μ_0 zu Maßen auf $\sigma(\mathcal{R})$.
Es gilt also $\mu_1(A) = \mu_2(A) = \mu_0(A) \; \forall\; A \in \mathcal{R}$. Sei $\mathcal{D}_n = \{A \in \sigma(\mathcal{R}) \,|\, \mu_1(A \cap B_n) = \mu_2(A \cap B_n)\}$. Dann gilt $\mathcal{R} \subset \mathcal{D}_n$ und $\mathcal{D}_n$ ist ein Dynkin-System $\forall\, n \in \mathbb{N}$. Da $\mathcal{R}$ als Ring $\cap$-stabil ist, folgt aus Satz 1.26(ii): $\delta(\mathcal{R}) = \sigma(\mathcal{R}) = \mathcal{D}_n$.
Also gilt $\forall\, n \in \mathbb{N}$ und $A \in \sigma(\mathcal{R})$: $\mu_1(A \cap B_n) = \mu_2(A \cap B_n)$. Indem man n gegen ∞ laufen lässt, folgt aus der Stetigkeit von unten: $\mu_1(A) = \mu_2(A) \, \forall\, A \in \sigma(\mathcal{R})$.

1.4 Einige wichtige Verteilungen

a) Diskrete Gleichverteilung und hypergeometrische Verteilung

Sei $\Omega \neq \varnothing$ endlich und für $A \subset \Omega$ sei $|A|$ die Anzahl der Elemte in A. Die Abbildung

$$\mathcal{P} : 2^{\Omega} \to \mathbb{R}, A \mapsto \frac{|A|}{|\Omega|}$$

definiert ein ω-Maß auf $(\Omega, 2^{\Omega})$ mit Zähldichte

$$\rho : \Omega \to \mathbb{R}, \omega \mapsto \frac{1}{|\Omega|}.$$

$\mathcal{P}$ wir *diskrete Gleichverteilung* auf Ω genannt. Ist $\mathcal{P}$ eine diskrete Gleichverteilung auf Ω, so wird (Ω, ρ) auch *Laplace-Experiment* auf Ω genannt, wobei ρ die Zähldichte von $\mathcal{P}$ ist.

1.27 Beispiele.

- Fairer Münzwurf: $\Omega = \{K, Z\}$
- Echter Würfel: $\Omega = \{1, 2, ..., 6\}$

1.28 Beispiel. (Uhrenmodelle)
In einer Urne seien N von 1 bis N durchnummerierte Kugeln. Es werden K Kugeln sukzessiv gezogen. Man unterscheidet:

- Ziehen mit Zurücklegen: ja/ nein
- Beachtung der Reihenfolge: ja/ nein

(a) Ziehen mit Zurücklegen unter Beachtung der Reihenfolge:

$$\Omega_I = \{(\omega_1, ..., \omega_K) \,|\, \omega_i \in \{1, ..., N\} \;\forall\, i = 1, ..., K\}, |\Omega_I| = N^K.$$

(b) Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge:

$$\Omega_{II} = \{(\omega_1, ..., \omega_K) \,|\, \omega_i \in \{1, ..., N\} \textit{ und } \omega_i \neq \omega_j \forall\, i \neq j\}, |\Omega_{II}| = \frac{N!}{(N-K)!}.$$

(c) Ziehen ohne Zurücklegen ohne Beachtung der Reihenfolge:

$$\Omega_{III} = \{(\omega_1, ..., \omega_K) \,|\, \omega_i \in \{1, ..., N\} \textit{ und } \omega_1 < \omega_2 < ... < \omega_K\}, |\Omega_{III}| = \binom{N}{K}.$$

Idee: Zu jedem $\omega \in \Omega_{III}$ sei M_ω die Menge der Ergebnisse aus Ω_{II}, die nach Sortieren der Größe zu ω werden. Nach (b) gilt $|M_\omega| = K!$. Somit $\Omega_{II} = \sum\limits_{\omega \in \Omega_{III}} |M_\omega| = K!|\Omega_{III}|$.

(d) Ziehen mit Zurücklegen ohne Beachtung der Reihenfolge:

$$\Omega_{IV} = \{(\omega_1, ..., \omega_K) \,|\, \omega_i \in \{1, ..., N\} \textit{ und } \omega_1 \leq \omega_2 \leq ... \leq \omega_K\}, |\Omega_{IV}| = \binom{N+K-1}{K}.$$

Idee: Die Abbildung $\Omega_{IV} \to \{(\omega_1', ..., \omega_K') \,|\, \omega_i' \in \{1, ..., N+K-1 \text{ und } \omega_1' \leq \omega_2' \leq ... \leq \omega_K'\}, (\omega_1, ..., \omega_K) \mapsto (\omega_1, \omega_2 + 1, \omega_3 + 2, ..., \omega_K + K - 1)$ ist bijektiv. Mit (c) folgt $|\Omega_{IV}| = \binom{N+K-1}{K}$.

1.29 Beispiel. Aus einer Urne mit r roten und s schwarzen Kugeln werden ohne Zurücklegen nacheinander rein zufällig $n \leq r + s$ Kugeln gezogen.

Frage: Wie groß ist die Wahrscheinlichkeit, dass dabei genau k rote Kugeln gezogen werden?

Modell: Urne mit $r + s$ von 1 bis $r + s$ durchnummerierten Kugeln, wobei die Nummern $1, ..., r$ rot gefärbt sind. Wir Ziehen ohne Zurücklegen unter Beachtung der Reihenfolge.

Laplace-Experiment auf

$$\Omega_{II} = \{(\omega_1, ..., \omega_n) \,|\, \omega_i \in \{1, ..., r+s\}, \omega_i \neq \omega_j \text{ für } i \neq j\}$$

. Für $\max\{0, n-s\} \leq k \leq \min\{r, n\}$ sei

$$A_k := \{\omega \in \Omega_{II} \,|\, |\{\omega_i \,|\, \omega_i \leq r\}| = k\}, |A_k| = \binom{n}{k} \frac{r!}{(r-k)!} \frac{s!}{(s-n+k)!}.$$

Denn:

- Ziehen der k Positionen, an denen die roten Kugeln positioniert werden: $\binom{n}{k}$ Möglichkeiten
- Füllen der gezogenen k Positionen mit roten Kugel (Beachtung der Reihenfolge): $\frac{r!}{(r-k)!}$ Möglichkeiten
- Füllen der übrigen $n - k$ Positionen mit schwarzen Kugeln: $\frac{s!}{(s-n+k)!}$ Möglichkeiten

Also: $\mathcal{P}(A_k) = \frac{|A_k|}{|\Omega_{II}|} = \frac{\binom{r}{k}\binom{s}{n-k}}{\binom{s+r}{n}}$

Neuer Stichprobenraum

$$\Omega = \{k \in \mathbb{N}_0 \,|\, \max\{0, n-s\} \leq k \leq \min\{r, n\}\}$$

. Da $\bigcup\limits_{k \in \Omega} A_k = \Omega_{II}$ und die A_k paarweise disjunkt sind folgt:

$$H_{r,s,n} : \Omega \to \mathbb{R}, k \mapsto \binom{n}{k} \frac{r!}{(r-k)!} \frac{s!}{(s-n+k)!}$$

ist eine Zähldichte auf Ω. Das von $H_{r,s,n}$ auf $(\Omega, 2^\Omega)$ induzierte Wahrscheinlichkeitsmaß wird *hypergeometrische Verteilung* genannt.

b) Binomialverteilung und Poisson-Verteilung

1.30 Definition. Sei

$$\Omega = \{0, ..., n\}, b_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}, k \in \Omega, p \in [0,1].$$

Dann ist $b_{n,p}$ eine Zähldichte auf Ω. Das diskrete Zufallsexperiment $(\Omega, b_{n,p})$ heißt *Binomial(n,p)-Experiment*, das zugehörige Wahrscheinlichkeitsmaß auf $(\Omega, 2^\Omega)$ heißt *Binomial(n,p)-Verteilt* ($n \in \mathbb{N}, p \in [0,1]$). Im Fall $n = 1$ spricht man von einem *Bernoulli(p)-Experiment* bzw. von der *Bernoulli (p)-Verteilung.*

1.31 Bemerkung. (i) Bei einem Bernoulli-Experiment sind nur zwei Versuchsausgänge möglich. Üblicherweise interpretiert man „1" als Erfolg und „0" als Fehlversuch. p gibt dann die Erfolgswahrscheinlichkeit an.

(ii) $b_{n,p}(k)$ gibt die Wahrscheinlichkeit für k Erfolge bei n-facher „unabhängiger" Durchführung von Bernoulli(p)-Experimenten an.

(iii) Weitere verwandte Zufallsexperimente:

- Warte auf den ersten Erfolg bei unabhängiger Durchführung von Bernoulli(p)-Experimenten
$$\Omega = \mathbb{N}, G_p(k) = p(1-p)^{k-1}.$$
Das von G_p induzierte Wahrscheinlichkeitsmaß heißt *geometrische Verteilung* mit Parameter $p \in [0,1]$.
- Anzahl der Fehlversuche bis zum n-ten Erfolg bei „unabhängiger" Durchführung von Bernoulli(p)-Experimenten
$$\Omega = \mathbb{N}_0, B^-_{n,p}(k) = \binom{n+k-1}{k} p^n (1-p)^k.$$
Dann definiert $B^-_{n,p}$ eine Zähldichte auf Ω. Das zugehörige Wahrscheinlichkeitsmaß wird *negative Binomialverteilung* genannt.

Analog zum Beispiel 1.29 zeigt man:

1.32 Satz. Werden aus einer Urne mit r roten und s schwarzen Kugeln nacheinander mit Zurücklegen „rein zufällig" (d.h. hier verwenden wir Ω_I) n Kugeln gezogen, so ist die Wahrscheinlichkeit dafür, dass dabei genau k rote Kugeln gezogen werden, gleich $b_{n,\frac{r}{r+s}}(k)$ für $k = 0, ..., n$.

Sind in der Urne sehr viele Kugeln, so sollte die Unterscheidung zwischen „mit Zurücklegen" und „ohne Zurücklegen" im Kontext von Satz 1.32 kaum eine Rolle spielen.

1.33 Satz. (Binomialapproximation der hypergeometrischen Verteilung)
Hängt im Kontext von Beispiel 1.29, Satz 1.32 die Anzahl der roten und schwarzen Kugeln r_N und s_N von $N \in \mathbb{N}$ ab, dass $s_N + r_N \to \infty (N \to \infty)$ und $\frac{r_N}{r_N + s_N} \to p \in (0,1)(N \to \infty)$ so gilt für alle $n \in \mathbb{N}$ und $k = 0, ..., n$:

$$\lim_{n \to \infty} H_{r_N + s_N, n}(k) = b_{n,p}(k)$$

.

Beweis. Da auch $s_N \to \infty, r_N \to \infty$ folgt etwa $\binom{r_N}{s_N} = \frac{r_N^k}{k!} \cdot \frac{r_N(r_N-1)\cdots(r_N+k-1)}{\underbrace{r_N r_N \cdots r_N}_{k-mal}} \sim \frac{r_N^k}{k!}(N \to \infty)$. Somit

$H_{r_N+s_N,n}(k) = \frac{\binom{r_N}{k}\binom{s_N}{n-k}}{\binom{r_N+s_N}{n}} \sim \frac{\frac{r_N^k}{k!} \frac{s_N^{n-k}}{(n-k)!}}{\frac{(r_N+s_N)^n}{n!}} \sim \binom{n}{k} p^k (1-p)^{n-k} (n \to \infty).$ □

1.34 Beispiel. In einer Gegend erkranken im langjährigen Mittel 3 Personen an Krebs. Im letzten Jahr waren es 5. Ist dies ein „außergewöhnliches Ereignis"?
Wir modellieren die Anzahl der Erkrankten pro Jahr durch ein Binomial(n,p)-Experiment mit $p = \frac{3}{n}$ wobei n die Anzahl der Bewohner der Gegend angibt. Wir betrachten die Fälle $n = 100$ und $n = 100000$. Als Wahrscheinlichkeit für die Erkrankung von mindestens 5 Personen ergibt sich dann:

$$\sum_{k=5}^{n} b_{n,\frac{3}{n}}(k) = 1 - \sum_{k=0}^{4} \binom{n}{k} (\frac{3}{n})^k (1 - \frac{3}{n})^{n-k} \approx \begin{cases} 0,181, & n = 100 \\ 0,185, & n = 100000 \end{cases}.$$

Die Wahrscheinlichkeit hängt hier kaum von der Populationsgröße ab und ist mit über 18% nicht so klein, dass man von einem „außergewöhnlichen" Ereignis sprechen kann.

Die erste Beobachtung ist Konsequenz des folgenden Satzes:

1.35 Satz. (Poisson-Approximation der Binomialverteilung)
Hängt in einem Binomialexperiment die Erfolgswahrscheinlichkeit p_n so von der Anzahl der Versuche $n \in \mathbb{N}$ ab, dass $\lim\limits_{n\to\infty} np_n = \lambda \in (0,\infty)$, dann gilt für alle $k \in \mathbb{N}_0$:

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

Das durch die Zähldichte $Poiss_\lambda(k) = e^{-\lambda}$ definierte Wahrscheinlichkeitsmaß heißt *Poisson* (λ)*-Verteilung.*

Beweis. $\binom{n}{k}p_n^k(1-p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(1-\frac{np_n}{n})^n}{(1-p_n)^k}(np_n)^k \cdot \frac{1}{k!} \to e^{-\lambda}\frac{\lambda^k}{k!}$ für $n\to\infty$, da $p_n \to 0$ und $\lambda = np$. □

Die Poisson-Verteilung wird zur Modellierung seltener Phänomene verwendet.

c) (Wahrscheinlichkeits-) Maße auf $(\mathbb{R}^D, \mathcal{B}_D)$

1.36 Beispiel. Sei $f : \mathbb{R} \to \mathbb{R}$ (Riemann-) integrierbar über kompakten Intervallen $[a,b]$ mit

- $f(x) \geq 0$
- $\int_\mathbb{R} f(x)dx = 1$.

Die Funktion $F : \mathbb{R} \to \mathbb{R}, x \mapsto \int_{-\infty}^x f(y)dy$ ist dann eine Verteilungsfunktion. Sei $\mathcal{P}_f$ das von F induzierte Wahrscheinlichkeitsmaß auf $(\mathbb{R},\mathcal{B})$ für das also gilt: $\mathcal{P}_f((a,b]) = \int_a^b f(x)dx$ $(a \leq b)$. In diesem Fall nennt man f die *(Riemann-) Dichte* von $\mathcal{P}_f$.

F	Name von $\mathcal{P}_F$	Parameter
$\frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$	Gleichverteilung auf $[a,b]$	$a < b \in \mathbb{R}$
$\lambda e^{-\lambda x}\mathbb{1}_{[0,\infty)}(x)$	Exponentialverteilung	$\lambda > 0$
$\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Normalverteilung	$\mu \in \mathbb{R}, \sigma > 0$
	(Standartverteilung)	$(\mu = 0, \sigma = 1)$

1.37 Beispiel. (Produktmaße)
Seien $F_1, ..., F_D : \mathbb{R} \to \mathbb{R}$ maßerzeugend. Seien $\mu_1, ..., \mu_D$ die zugehörigen Maße auf $(\mathbb{R},\mathcal{B})$ (Satz 1.22). Dann ist

$$F : \mathbb{R}^D \to \mathbb{R}, (x_1, ..., x_D) \mapsto \prod_{d=1}^D F_d(x_d)$$

maßerzeugend auf $\mathbb{R}^D$. Denn F ist:

- Rechtsstetig: Klar
- $\triangle$-monoton:

$$\begin{aligned}
\triangle_a^b F &= \sum_{x\in\{0,1\}^D} (-1)^{\sum\limits_{d=1}^D x_d} F_1(a_1^{x_1}b_1^{1-x_1})\cdots F_D(a_D^{x_D}b_D^{1-x_D}) \\
&= (F_D(b_D) - F_D(a_D)) \sum_{x\in\{0,1\}^D} (-1)^{\sum\limits_{d=1}^{D-1} x_d} F_1(a_1^{x_1}b_1^{1-x_1})\cdots F_D(a_{D-1}^{x_{D-1}}b_{D-1}^{1-x_{D-1}}) \\
&= ... = \prod_{d=1}^D (F_d(b_d) - F_d(a_d)) \geq 0
\end{aligned}$$

Ferner ist F eine Verteilungsfunktion, falls alle $F_d, d = 1, ..., D$, eindimensionale Verteilungsfunktionen sind. Nach Bemerkung 1.25 existiert ein eindeutiges Maß $\mu_1 \otimes ... \otimes \mu_D$ auf $(\mathbb{R}^D, \mathcal{B}_D)$ mit

$$\mu_1 \otimes ... \otimes \mu_D((a,b]) = \prod_{d=1}^D \mu_d((a_d,b_d]) \text{ für alle } a \leq b \in \mathbb{R}^D,$$

$\mu_1 \otimes ... \otimes \mu_D$ wird *Produktmaß* genannt.
Ist speziell $F_1(x) = ... = F_D(x) = x$, so erhalten wir das Lebesgue-Maß auf $(\mathbb{R}^D, \mathcal{B}_D)$.

Kapitel 2

Zufallsvariablen und Unabhängigkeit

2.1 Zufallsvariablen

2.1 Definition. Seien $(\Omega, \mathcal{A})$ und $(\Omega', \mathcal{A}')$ zwei Messräume. Eine Abbildung $X : \Omega \to \Omega'$ wird $\mathcal{A}/\mathcal{A}'$-*messbar* (kurz: *messbar*) genannt, falls für alle $B \in \mathcal{A}'$ gilt:

$$X^{-1}(B) \in \mathcal{A}.$$

Messbare Funktionen werden auch *Zufallsvariablen* genannt. Ist $(\Omega', \mathcal{A}') = (\mathbb{R}^D, \mathcal{B}_D)$, sow ird X auch $\mathbb{R}^D$-*wertig* genannt. Ist $(\Omega', \mathcal{A}') = (\mathbb{R} \cup \{+\infty, -\infty\}, \overline{\mathcal{B}})$, wobei $B \in \overline{\mathcal{B}}$ genau dann, wenn $B \cap \mathbb{R} \in \mathcal{B}$, so spricht man von einer *numerischen Zufallsvariable.*

2.2 Beispiele. (i) Sei $(\Omega, \mathcal{A})$ ein Messraum und $A \subset \Omega$. Dann ist $\mathbb{1}_A$ genau dann eine numerische Zufallsvariable, wenn $A \in \mathcal{A}$. Denn für alle $B \in \overline{\mathcal{B}}$ gilt:

$$X^{-1}(B) \in \{\varnothing, \Omega, A, A^c\}.$$

(ii) Sei $(\Omega, \mathcal{A})$ ein Messraum und $(A_i)_{i=1,\dots,n}$ eine Zerlegung von Ω, d.h. die A_i sind paarweise disjunkt und $\bigcup\limits_{i=1}^{n} A_i = \Omega$. Sind dann $A_i \in \mathcal{A}$ für alle $i = 1, ..., n$ und sind $\alpha_1, ..., \alpha_n$ reelle Zahlen, so ist

$$X = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$$

eine $\mathbb{R}^D$-wertige Zufallsvariable. Denn sei Œ $\alpha_i \neq \alpha_j$ für $i \neq j$. Dann folgt für $B \in \mathcal{B}_D$:

$$X^{-1}(B) = \bigcup_{i, \alpha_i \in B} \{\omega \,|\, X(\omega) = \alpha_i\} = \bigcup_{i, \alpha_i \in B} A_i \in \mathcal{A}.$$

Zum Nachweis der Messbarkeit ist hilfreich:

2.3 Satz. Seien $(\Omega, \mathcal{A}), (\Omega', \mathcal{A}')$ zwei Messräume und es sei ε ein Erzeugendensystem von $\mathcal{A}'$ (d.h. $\varepsilon \subset \mathcal{A}'$ und $\sigma(\varepsilon) = \mathcal{A}'$). Dann ist $X : \Omega \to \Omega'$ genau dann $\mathcal{A}/\mathcal{A}'$-messbar, wenn $X^{-1}(B) \in \mathcal{A}$ für alle $B \in \varepsilon$.

Der Beweis beruht auf dem *Prinzip der geeigneten Mengen*:
Sei $\mathcal{A}'$ eine σ-Algebra. Zum Nachweis einer Eigenschaft a auf $\mathcal{A}'$ betrachtet man

$$\mathcal{M} = \{A' \in \mathcal{A}' \,|\, A' \text{ hat Eigenschaft } a\}.$$

Enthält $\mathcal{M}$ ein Erzeugendensystem ε von $\mathcal{A}'$ und ist $\mathcal{M}$ eine σ-Algebra, so folgt

$$\mathcal{A}' = \sigma(\varepsilon) \subset \sigma(\mathcal{M}) = \mathcal{M} \subset \mathcal{A}',$$

d.h. $\mathcal{M} = \mathcal{A}'$ (also hat jedes $A' \in \mathcal{A}'$ Eigenschaft a).

Beweis. Nach dem Prinzip der geeigneten Mengen reicht es zu zeigen, dass

$$\mathcal{M} = \{B \in \mathcal{A}' \,|\, X^{-1}(B) \in \mathcal{A}\}$$

eine σ-Algebra ist:

- $\Omega' \in \mathcal{M}$, da $X^{-1}(\Omega') = \Omega$

- $A \in \mathcal{M} \Rightarrow X^{-1}(A) \in \mathcal{A} \Rightarrow (X^{-1}(A))^c \in \mathcal{A} \Rightarrow X^{-1}(A^c) \in \mathcal{A} \Rightarrow A^c \in \mathcal{M}$
- $(A_n)_{n\in\mathbb{N}} \subset \mathcal{M} \Rightarrow \forall n \in \mathbb{N} : X^{-1}(A_n) \in \mathcal{A} \Rightarrow X^{-1}(\bigcup\limits_{n\in\mathbb{N}} A_n) = \bigcup\limits_{n\in\mathbb{N}} X^{-1}(A_n) \in \mathcal{A} \Rightarrow \bigcup\limits_{n\in\mathbb{N}} A_n \in \mathcal{M}$ □

2.4 Folgerung. (i) Sei $(\Omega, \mathcal{A})$ ein Messraum. $X : \Omega \to \mathbb{R}\cup\{+\infty, -\infty\}$ ist genau dann numerisch, wenn für alle $a \in \mathbb{R}$ gilt

$$\{\omega \mid X(\omega) \leq a\} \in \mathcal{A} \quad [\text{bzw.}\{\omega \mid X(\omega) \geq a\}].$$

(ii) Jede stetige Funktion $X : \mathbb{R}^M \to \mathbb{R}^D$ ist $\mathcal{B}_M/\mathcal{B}_D$-messbar.

Beweis. (i) Übung

(ii) Folgt aus Satz 2.3, da die offenen Mengen $\mathcal{B}_D$ erzeugen und bei stetigen Funktionen die Urbilder offener Mengen offen sind. □

2.5 Folgerung. Sei $(\Omega, \mathcal{A})$ ein Messraum und seien $X_1, ..., X_D : \Omega \to \mathbb{R}$ $\mathcal{A}/\mathcal{B}$-messbar. Dann ist $X = \begin{pmatrix} x_1 \\ : \\ x_D \end{pmatrix}$ eine $\mathbb{R}^D$-wertige Zufallsvariable.

Beweis. Für $a \leq v \in \mathbb{R}^D$ gilt: $\underbrace{X^{-1}((a, b]) = \bigcap\limits_{d=1}^{D} X_d^{-1}((a_d, b_d])}_{\in\mathcal{A}} \in \mathcal{A}$. Da die linkshalboffenen Intervalle $\mathcal{B}_D$ erzeugen, ist Satz 2.3 anwendbar. □

2.6 Satz. Seien $(\Omega_i, \mathcal{A}_i)_{i=1,2,3}$ Messräume und $X_1 : \Omega_1 \to \Omega_2, X_2 : \Omega_2 \to \Omega_3$ messbar. Dann ist $X_2 \circ X_1 : \Omega_1 \to \Omega_3$ messbar.

Beweis. Für alle $B \in \mathcal{A}_3$ gilt:

$$(X_2 \circ X_1)^{-1}(B) = \{\omega \in \Omega_1 \mid X_2 \circ X_1(\omega) \in B\} = \{\omega \in \Omega_1 \mid X_1(\omega) \in X_2^{-1}(B)\} \in \mathcal{A}_1.$$

□

2.7 Beispiele. Seien $X_1, ..., X_D : \Omega \to \mathbb{R}$ reellwertige Zufallsvariablen, wobei der Messraum $(\Omega, \mathcal{A})$ gegeben ist. Dann sind

$$\sum_{d=1}^{D} X_d \text{ und } \prod_{d=1}^{D} X_d$$

reellwertige Zufallsvariablen. Nach Folgerung 2.5 ist $X = \begin{pmatrix} x_1 \\ : \\ x_D \end{pmatrix}$ eine $\mathbb{R}^D$-wertige Zufallsvariable und nach Folgerung 2.4(ii) sind $+ : \mathbb{R}^D \to \mathbb{R}, (x_1, ..., x_D) \mapsto x_1 + ... + x_D$ und $\cdot : \mathbb{R}^D \to \mathbb{R}, (x_1, ..., x_D) \mapsto x_1 \cdots x_D$ als stetige Funktionen messbar. Also sind die Kompositionen $\sum\limits_{d=1}^{D} X_d$ und $\prod\limits_{d=1}^{D} X_d$ messbar.

2.8 Satz. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von numerischen Zufallsvariablen auf einem Messraum $(\Omega, \mathcal{A})$. Dann gilt:

(i) $\sup\limits_{n\in\mathbb{N}} X_n$, $\inf\limits_{n\in\mathbb{N}} X_n$, $\limsup\limits_{n\to\infty} X_n$, $\liminf\limits_{n\to\infty} X_n$ sind numerische Zufallsvariablen.

(ii) $A := \{\omega \mid \lim\limits_{n\to\infty} X_n(\omega) \in [-\infty, +\infty] \text{ existiert}\} \in \mathcal{A}$ und $X := \mathbb{1} \cdot \lim\limits_{n\to\infty} X_n$ ist eine numerische Zufallsvariable.

Beweis. (i) Für alle $A \in \mathbb{R}$ gilt

$$\{\omega \mid \sup_{n\to\infty} X_n(\omega) \geq a\} = \bigcup_{n\in\mathbb{N}} \{\omega \mid X_n(\omega) > a\} \in \mathcal{A},$$

$$\{\omega \mid \inf_{n\to\infty} X_n(\omega) \leq a\} = \bigcup_{n\in\mathbb{N}} \{\omega \mid X_n(\omega) \leq a\} \in \mathcal{A}.$$

Also sind $\sup\limits_{n\in\mathbb{N}} X_n$ und $\inf\limits_{n\in\mathbb{N}}$ numerische Zufallsvariablen nach Folgerung 2.4(i). Nach Definition sind dann auch $\limsup\limits_{n\to\infty} X_n = \inf\limits_{n\in\mathbb{N}} \sup\limits_{n\geq m} X_m$ und $\liminf\limits_{n\to\infty} X_n = \sup\limits_{n\in\mathbb{N}} \inf\limits_{n\geq m} X_m$ numerische Zufallsvariablen.

(ii) Da $\limsup_{n\in\mathbb{N}} X_n$ und $\liminf_{n\in\mathbb{N}} X_n$ numerische Zufallsvariablen sind, folgt: $A = \{\omega \in \Omega \mid \limsup_{n\to\infty} X_n(\omega) = \liminf_{n\to\infty} X_n(\omega)\} \in \mathcal{A}$ (Vgl. Übung).
Also $\{\omega \mid X(\omega) \leq a\} = (\underbrace{\{\omega \mid \limsup_{n\to\infty} X_n(\omega) \leq a\}}_{\in\mathcal{A}} \cap \underbrace{A}_{\in\mathcal{A}}) \cup (\underbrace{A^c}_{\in\mathcal{A}} \cap \underbrace{\{\omega \mid 0 \leq a\}}_{\in\{\varnothing,\Omega\}\in\mathcal{A}}) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
Dann: Folgerung 2.4. $\square$

2.9 Definition. Seien $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{A}')$ ein Messraum und $X : \Omega \to \Omega' \mathcal{A}|\mathcal{A}'$-messbar. Dann heißt

$$\mathcal{P}_X(B) := P(X^{-1}(B)), B \in \mathcal{A}'$$

die *Verteilung von X*. $\mathcal{P}_X$ definiert ein Wahrscheinlichkeitsmaß auf $(\Omega', \mathcal{A}')$.
Ist X $\mathbb{R}^D$-wertig und $x \in \mathbb{R}^D$, so wird

$$F_X := \mathcal{P}_X((-\infty, x])$$

die *Verteilungsfunktion* von X genannt. Sie ist eine D-dimensionale Verteilungsfunktion im Sinn von Bemerkung 1.25

2.10 Satz. Sei F eine D-dimensionale Verteilungsfunktion. Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ und eine $\mathbb{R}^D$-wertige Zufallsvariable X auf $(\Omega, \mathcal{A}, \mathcal{P})$ so, dass F die Verteilungsfunktion von X ist.

Beweis. Sei $(\Omega, \mathcal{A}) = (\mathbb{R}^D, \mathcal{B}_D)$ und sei $\mathcal{P}$ das durch $\mathcal{P}((a, b]) := \triangle_a^b F$ für alle $a, b \in \mathbb{R}^D$ eindeutig bestimmte Wahrscheinlichkeitsmaß (Bemerkung 1.25). Setze $X : \mathbb{R}^D \to \mathbb{R}^D, \omega \mapsto \omega$. Dann ist X eine Zufallsvariable und es gilt für alle $X \in \mathbb{R}^D$

$$F_X(x) = \mathcal{P}(\{\omega \mid X(\omega) \leq x\}) = \mathcal{P}((-\infty, x]) = F(x).$$

$\square$

Konvention.

- Schreibweise: $\{X \in B\} := \{\omega \mid X(\omega) \in B\}$
- Wir sagen eine Eigenschaft gilt *μ-fast sicher*, falls die Menge der $\omega \in \Omega$, für die diese Eigenschaft nicht gilt, unter μ das Maß 0 hat.
  Z.B. $X \leq 0$ μ-fast sicher bedeutet $\mu(\{X > 0\}) = 0$.
- Zwei Zufallsvariablen X und Y sind gleich bis auf *Modifikation*, falls $X = Y$ μ-fast sicher.
- Rechnen mit $\pm\infty$: Für $a \in \mathbb{R}$ vereinbaren wir

$$a \pm \infty = \pm\infty \qquad \frac{a}{\pm\infty} = 0$$

$$a \cdot (+\infty) = \begin{cases} +\infty & , a > 0 \text{ oder } a = +\infty \\ -\infty & , a < 0 \text{ oder } a = -\infty \\ 0 & , a = 0 \end{cases} \qquad a \cdot (-\infty) = -(a \cdot (+\infty))$$

$$(+\infty) + (+\infty) = +\infty \qquad (-\infty) - (+\infty) = -\infty$$

  Undefiniert bleiben: $\infty - \infty$ und $\frac{\infty}{\infty}$
- $X(\Omega, \mathcal{A}) \to (\Omega' \mathcal{A}')$ Zufallsvariable bedeutet: $X : \Omega \to \Omega'$ ist $\mathcal{A}/\mathcal{A}'$-messbar.

2.2 Unabhängigkeit

2.11 Definition. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum.

(i) Zwei Ereignisse A und B werden *stochastisch unabhängig* genannt, falls

$$\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B).$$

(ii) Eine Familie von Ereignissen $(A_i)_{i\in I} \in \mathcal{A}, I \neq \varnothing$ beliebig, heißt *stochastisch unabhängig*, falls für alle endlichen $J \subset I$ gilt:

$$\mathcal{P}(\bigcap_{j\in J} A_j) = \prod_{j\in J} \mathcal{P}(A_j).$$

Die Familie wird *paarweise unabhängig* genannt, falls für alle $i \neq j \in I$ gilt:

$$\mathcal{P}(A_i \cap A_j) = \mathcal{P}(A_i)\mathcal{P}(A_j).$$

2.12 Bemerkung. (i) Wir betrachten ein Laplace-Experiment auf einer endlichen Menge Ω. Dann sind $A, B \in 2^{\Omega}, B \neq \varnothing$ stochastisch unabhängig genau dann, wenn

$$\frac{|A|}{|\Omega|} = \frac{|A \cap B|}{|B|}.$$

Das bedeutet: Das Merkmal, das durch A beschrieben wird, ist relativ zur Gesamtpopulation des Stichprobenraums gleich häufig wie relativ zur Subpopulation B.

(ii) Sei (A, B, C) eine Familie von drei Ereignissen. Es reicht nicht aus für die Unabhängigkeit zu überprüfen, dass $\mathcal{P}(A \cap B \cap C) = \mathcal{P}(A)\mathcal{P}(B)\mathcal{P}(C)$.
Beispiel: $A = B, C \neq \varnothing, \mathcal{P}(A) \in (0,1)$. Dann: $\mathcal{P}(A \cap B \cap C) = 0 = \mathcal{P}(A)\mathcal{P}(B)\mathcal{P}(C)$, aber $\mathcal{P}(A \cap B) = \mathcal{P}(A) \neq \mathcal{P}(A)^2 = \mathcal{P}(A)\mathcal{P}(B)$.

(iii) Unabhängigkeit impliziert paarweise Unabhängigkeit. Die Umkehrung gilt i.A. nicht.
Beispiel: Sei ein Laplace-Experiment auf $\{K, Z\}^2$ („zweifacher Münzwurf") gegeben. Wir betrachten:

$$\begin{aligned} A &= \{(K,Z),(K,K)\} \triangleq \text{„erster Wurf Kopf"} \\ B &= \{(K,Z),(Z,K)\} \triangleq \text{„zweiter Wurf Kopf"} \\ C &= \{(K,K),(Z,Z)\} \triangleq \text{„erster Wurf = zweiter Wurf"} \end{aligned}$$

Dann ist die Familie (A, B, C) nicht stochastisch unabhängig, aber paarweise unabhängig.

2.13 Beispiel. Sei $(\mathbb{R}^2, \mathcal{B}_2, \mathcal{P}_1 \otimes \mathcal{P}_2)$ ein Wahrscheinlichkeitsraum, wobei $\mathcal{P}_1, \mathcal{P}_2$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$ sind (siehe Beispiel 1.38). Dann sind Ereignisse von der Form $A_\times \mathbb{R}$ und $\mathbb{R} \times A_2$ für $A_1, A_2 \in \mathcal{B}$ unabhängig. Denn:

$$\begin{aligned} \mathcal{P}_1 \otimes \mathcal{P}_2((A_1 \times \mathbb{R}) \cap (\mathbb{R} \times A_2)) &= \mathcal{P}_1 \otimes \mathcal{P}_2(A_1 \times A_2) \\ &= \mathcal{P}_1(A_1)\mathcal{P}_2(A_2) \\ &= (\mathcal{P}(A_1)\mathcal{P}_2(\mathbb{R}))(\mathcal{P}(\mathbb{R})\mathcal{P}(A_2)) \\ &= \mathcal{P}_1 \otimes \mathcal{P}_2(A_1 \times \mathbb{R})\mathcal{P}_1 \otimes \mathcal{P}_2(\mathbb{R} \times A_2) \end{aligned}$$

Dies gilt nach Beispiel 1.37, falls A_1 und A_2 linkshalboffene Intervalle sind und folgt aus Satz 2.15 für den allgemeinen Fall. In diesem Sinn beschreiben Produktmaße die unabhängige Kopplung von Zufallsexperimenten.

2.14 Definition. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{A}_i)_{i \in I}, I \neq \varnothing$, eine Familie von Unter-σ-Algebren von $\mathcal{A}$ (d.h. jedes $\mathcal{A}_i$ ist σ-Algebra mit $\mathcal{A}_i \subset \mathcal{A}$). Eine solche Familie von σ-Algebren wird *stochastisch unabhängig* genannt, falls die Familie $(A_i)_{i \in I}$ stochastisch unabhängig ist, wenn immer $A_i \in \mathcal{A}_i$ für alle $i \in I$.

2.15 Satz. Sei $(\mathcal{A}_i)_{i \in I}$ eine Familie von Unter-σ-Algebren von $\mathcal{A}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$. $(\varepsilon_i)_{i \in I}$ seien $\cap$-stabile Erzeuger von $\mathcal{A}_i$, d.h.

- $\sigma(\varepsilon_i) = \mathcal{A}_i$ für alle $i \in I$,
- für alle $i \in I$: $A, B \in \varepsilon \Rightarrow A \cap B \in \varepsilon_i$.

Dann ist die Familie $(A_i)_{i \in I}$ unabhängig, wenn immer $A_i \in \varepsilon_i$ für alle $i \in I$, so ist die Familie $(\mathcal{A}_i)_{i \in I}$ unabhängig.

Beweis. Œ sei I endlich ($(\mathcal{A}_i)_{i \in I}$ ist unabhängig genau dann, wenn jede endliche Teilfamilie unabhängig ist). Sei $i_0 \in I$ fixiert. Sei $\mathcal{D}_{i_0}$ die Menge aller $E \subset \mathcal{A}_{i_0}$ so, dasss

$$\mathcal{P}(A_{i_1} \cap ... \cap A_{i_n} \cap E) = \mathcal{P}(A_{i_0}) \cdots \mathcal{P}(A_n)\mathcal{P}(E)$$

für alle $\{i_1, ..., i_n\} \subset I \setminus \{i_0\}$ und alle $A_{i_j} \subset \varepsilon_{i_j}, j = 1, ..., n$. Wir zeigen, dass $\mathcal{D}_{i_0}$ ein Dynkin-System ist:

- $\Omega \in \mathcal{D}_{i_0}$: Klar
- $E \subset \mathcal{D}_{i_0} \Rightarrow \mathcal{P}(A_{i_1} \cap ... \cap A_{i_n} \cap E^c) = \mathcal{P}(A_{i_1} \cap ... \cap A_{i_n} \cap \Omega) - \mathcal{P}(A_{i_1} \cap ... \cap A_{i_n} \cap E) = \mathcal{P}(A_{i_1}) \cdots \mathcal{P}(A_{i_n}) \underbrace{(1 - \mathcal{P}(E))}_{\mathcal{P}(E^c)} \Rightarrow E^c \in \mathcal{D}_{i_0}$
- Seien $(E_k)_{k \in \mathbb{N}}$ in $\mathcal{D}_{i_0}$ paarweise disjunkt. Dann:

$$\begin{aligned} \mathcal{P}(A_{i_1} \cap ... \cap A_{i_n}(\bigcup_{k \in \mathbb{N}} E_k)) &= \mathcal{P}(\bigcup_{k \in \mathbb{N}} (A_{i_1} \cap ... \cap A_{i_n} \cap E_k)) \stackrel{\sigma-Add.}{=} \sum_{k \in \mathbb{N}} \mathcal{P}(A_{i_1} \cap ... \cap A_{i_n} \cap E_k) \\ &= \sum_{k \in \mathbb{N}} \mathcal{P}(A_{i_1}) \cdots \mathcal{P}(A_{i_n})\mathcal{P}(E_k) \stackrel{\sigma-Add.}{=} \mathcal{P}(A_{i_1}) \cdots \mathcal{P}(A_{i_n})\mathcal{P}(\bigcup_{k \in \mathbb{N}} E_k) \end{aligned}$$

Also: $\bigcup_{k \in \mathbb{N}} E_k \in \mathcal{D}_{i_0}$.

Da $\varepsilon_{i_0} \subset \mathcal{D}_{i_0}$, folgt aus dem Dynkin-Lemma 1.26, dass

$$\mathcal{A}_{i_0} = \sigma(\varepsilon_{i_0}) = \delta(\varepsilon_{i_0}) \subset \mathcal{D}_{i_0} \subset \mathcal{A}_{i_0}, \text{ d.h } \mathcal{A}_{i_0} = \mathcal{D}_{i_0}$$

Wir haben gezeigt: $(\mathcal{A}_i)_{i\in I}$ ist unabhängig, wenn immer $A_i \in \varepsilon_i$ für $i \in I \setminus \{i_0\}$ und $A_{i_0} \in \mathcal{A}_{i_0}$. Da I endlich ist, folgt durch Iteration dieses Vorgehens: $(A_i)_{i\in I}$ ist unabhängig, wenn immer $A_i \in \mathcal{A}_i$ für alle $i \in I$. □

2.16 Satz. Sei $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ eine Zufallsvariable. Dann ist

$$\sigma(X) := \{X^{-1}(A) \mid A \in \mathcal{A}'\}$$

eine Unter-σ-Algebra von $\mathcal{A}$. Sie wird die von X *erzeugte σ-Algebra* genannt.

Beweis. Übung □

2.17 Definition. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum und $(X_i)_{i\in I}$ $(I \neq \varnothing)$ eine Familie von Zufallsvariablen mit Werten in $(\Omega_i, \mathcal{A}_i), i \in I$. Dann wird $(X_i)_{i\in I}$ *stochastisch unabhängig* genannt, falls die Familie $(\sigma(X_i))_{i\in I}$ unabhängig ist.

Bemerkung. Eine Familie $(X_i)_{i\in I}$ von Zufallsvariablen ist unabhängig genau dann, wenn jede endliche Teilfamilie unabhängig ist.

2.18 Satz. Seien $X_1, ..., X_D$ reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$. Sei $\mathcal{F}_X$ die Verteilungsfunktion der $\mathbb{R}^D$-wertigen Zufallsvariable $X = (X_1, ..., X_D)^T$. Dann ist die Familie $(X_1, ..., X_D)$ genau dann unabhängig, wenn für alle $x = (x_1, ..., x_D)^T \in \mathbb{R}^D$:

$$(*) \qquad \mathcal{F}_X(x_1, ..., x_D) = \prod_{d=1}^{D} \mathcal{F}_{X_d}(x_d).$$

Beweis. (a) Seien $X_1, ..., X_D$ unabhängig und $x_1, ..., x_D \in \mathbb{R}$. Dann:

- $\forall d = 1, ..., D : \{X_d \leq x_d\} \in \sigma(X_d)$

Da nach Voraussetzung die σ-Algebren $(\sigma(X_d))_{d=1,...,D}$ unabhängig sind, folgt

$$F_X(x_1, ..., x_D) = \mathcal{P}(\bigcap_{d=1}^{D}\{X_d \leq x_d\}) = \prod_{d=1}^{D} \mathcal{P}(\{X_d \leq x_d\}) = \prod_{d=1}^{D} F_{X_D}(x_d).$$

(b) Es gelte nun $(*)$. Sei $K \subset D$ und $\{d_1, ..., d_K\} \subset \{1, ..., D\}$. Seien $x_1, ..., x_K \in \mathbb{R}$ und $Y^{(n)} = \begin{pmatrix} Y_1^{(n)} \\ \vdots \\ Y_D^{(n)} \end{pmatrix}$

mit $Y_s^{(n)} = \begin{cases} x_i, & d = d_i \text{ für ein } i = 1, ..., K \\ n, & \text{sonst} \end{cases}$. Dann:

$$\begin{aligned}
\mathcal{P}(\{X_{d_1} \leq x_1\} \cap ... \cap \{X_{d_K} \leq x_K\}) &\overset{stetig}{\underset{v.u.}{=}} \lim_{n\to\infty} \mathcal{P}(\bigcap_{d=1}^{D}\{X_d \leq Y_d^{(n)}\}) = \lim_{n\to\infty} F_X(Y_1^{(n)}, ..., Y_D^{(n)}) \\
&\overset{(*)}{=} \prod_{d=1}^{D} \lim_{n\to\infty} F_{X_d}(Y_d^{(n)}) = F_{X_{d_1}}^{(x_1)} \cdots F_{X_{d_K}}^{(x_K)} \\
&= \mathcal{P}(\{X_{d_1} \leq x_1\}) \cdots \mathcal{P}(\{X_{d_K} \leq x_K\})
\end{aligned}$$

Somit ist Satz 2.15 mit den Erzeugern $\varepsilon_d = \{\{X_d \leq x\} \mid x \in \mathbb{N}\}$ anwendbar. Man beachte dabei, dass nach Definition $\sigma(X_d) \supset \sigma(\varepsilon_d)$. Ferner ist X_d $\sigma(\varepsilon_d)/\mathcal{B}$-messbar nach Satz 2.3. Also gilt auch $\sigma(\varepsilon_d) \supset \sigma(X_d)$. □

2.19 Beispiel. Sei X eine standardnormalverteilte Zufallsvariable auf $(\Omega, \mathcal{A}, \mathcal{P})$. Seien $Y := |X|$ und $Z := sgn(X)$. Dann:

$$F_Y(y) = \mathcal{P}(\{|X| \leq y\}) = \begin{cases} 0 & , y < 0 \\ 2\int\limits_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx & , y \geq 0 \end{cases}$$

$$F_Z(z) = \mathcal{P}(\{sgn(X) \leq z\}) = \begin{cases} 0 & , z < -1 \\ \frac{1}{2} & , -1 \leq z < 1 \\ 1 & , z \geq 1 \end{cases}$$

$$F_{(Y,Z)}(y,z) = \mathcal{P}(\{|X| \leq y\} \cap \{sgn(X) \leq z\}) = \begin{cases} 0 & , y < 0 \text{ oder } z < -1 \\ \int\limits_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx & , y \geq 0, -1 \leq z < 1 \\ 2\int\limits_0^y \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx & , y \geq 0, z \geq 1 \end{cases}$$

Also sind $|X|$ und $sgn(X)$ unabhängig.

2.20 Satz. Seien $X_1, ..., X_D$ Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathcal{P})$ mit Werten in $(\Omega_d, \mathcal{A}_d), d = 1, ..., D$. Die Zufallsvariablen X_d seine diskret in dem Sinn, dass eine höchstens abzählbare Menge $Y_d \subset \Omega_d$ existiert mit $X_d(\Omega) \subset Y_d$ und $\{y\} \in \mathcal{A}_d$ für alle $y \in Y_d$. Dann sind $X_1, ..., X_D$ unabhängig genau dann, wenn für alle $Y_d \in Y_d, d = 1, ..., D$:

$$\mathcal{P}(\bigcap_{d=1}^{D} \{X_d = Y_d\}) = \prod_{d=1}^{D} \mathcal{P}(\{X_d = Y_d\}).$$

Beweis. Es gilt $\sigma(X_d) = \{X_d^{-1}(B) \,|\, B \subset Y_d\}$. Seien $B_d \subset Y_d$ für $d = 1, ..., D$. Dann:

$$\begin{aligned} \mathcal{P}(\bigcap_{d=1}^{D} \{X_d \in B_d\}) &= \sum_{Y_1 \in B_1} \cdots \sum_{Y_D \in B_D} \mathcal{P}(\bigcap_{d=1}^{D} \{X_d = Y_d\}) \\ &= (\sum_{Y_1 \in B_1} \mathcal{P}(\{X_1 = Y_1\})) \cdots (\sum_{Y_D \in B_D} \mathcal{P}(\{X_D = Y_D\}) \\ &= \prod_{d=1}^{D} \mathcal{P}(\{X_d \in B_d\}) \end{aligned}$$

□

2.3 Bedingte Wahrscheinlichkeiten

2.21 Definition. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum, $A, B \in \mathcal{A}$ mit $\mathcal{P}(B) > 0$. Dann heißt

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

die *bedingte Wahrscheinlichkeit* von A gegeben B. Es handelt sich um die Wahrscheinlichkeit für das Eintreten von A unter der Vorinformation, dass B eintritt.

2.22 Satz. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{A}$ mit $\mathcal{P}(B) > 0$. Dann definiert

$$\mathcal{P}_B : \mathcal{A} \to \mathbb{R}, A \mapsto \mathcal{P}(A|B)$$

ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ mit $\mathcal{P}_B(B) = 1$.

Beweis. Einfache Übung. □

2.23 Bemerkung. (i) Sei $(\Omega, \mathcal{A})$ ein Messraum und $B \subset \Omega$. Dann ist

$$\mathcal{A}_B := \{A \cap B \,|\, A \in \mathcal{A}\}$$

eine σ-Algebra auf B (Übung A.5, B.4), die sogenannte *Spur-σ-Algebra*.

(ii) Ist $\mathcal{P}$ ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$ und $B \in \mathcal{A}$ mit $\mathcal{P}(B) > 0$, so ist $\mathcal{P}_B$ ein Wahrscheinlichkeitsmaß auf $(B, \mathcal{A}_B)$.

Wichtige Rechenregeln:

2.24 Satz. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum , $(B_i)_{i \in I}$ eine höchstens abzählbare Familie von paarweise disjunkten Ereignissen $B_i \in \mathcal{A}$ mit $\bigcup\limits_{i \in I} B_i = \Omega$ und $\mathcal{P}(B_i) > 0$. Sei $A \in \mathcal{A}$. Dann gelten:

(i) Die *Formel von der totalen Wahrscheinlichkeit*:

$$\mathcal{P}(A) = \sum_{i \in I} \mathcal{P}(A|B_i)\mathcal{P}(B_i).$$

(ii) *Formel von Bayes*: Falls $\mathcal{P}(A) > 0$, so gilt für alle $j \in I$

$$\mathcal{P}(B_j|A) = \frac{\mathcal{P}(A|B_j)\mathcal{P}(B_j)}{\sum\limits_{i \in I} \mathcal{P}(A|B_j)\mathcal{P}(B_j)}.$$

Beweis. (i) $\sum_{i\in I} \mathcal{P}(A|B_i)\mathcal{P}(B_i) = \sum_{i\in I} \mathcal{P}(A\cap B_i) \overset{\sigma-Add.}{=} \mathcal{P}((\bigcup_{i\in I} B_i)\cap A) = \mathcal{P}(\Omega\cap A) = \mathcal{P}(A)$

(ii) $\mathcal{P}(B_j|A) = \frac{\mathcal{P}(B_j\cap A)\mathcal{P}(B_j)}{\mathcal{P}(A)} = \frac{\mathcal{P}(A|B_j)\mathcal{P}(B_j)}{\mathcal{P}(A)}$, dann ersetze $\mathcal{P}(A)$ durch die Formel in (i). $\square$

2.25 Beispiel. Bei einer Diagnosemethode gibt die Sensitivität die Wahrscheinlichkeit an, dass eine kranke Person als krank erkannt wird und die Spezifität gibt die Wahrscheinlichkeit an, dass eine gesunde Person als gesund erkannt wird. Seien Sensitivität und Spezifität jeweils $99,8\%$.
Wie hoch ist die Wahrscheinlichkeit, dass eine positiv getestete Person tatsächlich erkrankt ist? Wir nehmen an, dass die Person zu einer Risikogruppe gehört, in der die Erkrankung mit einer Wahrscheinlichkeit von $\frac{1}{10000}$ auftritt.
Sei

$$A = \text{„positiver Befund“}, \\ B = \text{„tatsächlich erkrankt“}.$$

Dann liefert die Formel von Bayes:

$$\begin{aligned} \mathcal{P}(B|A) &= \frac{\mathcal{P}(A|B)\mathcal{P}(B)}{\mathcal{P}(A|B)\mathcal{P}(B) + \mathcal{P}(A|B^c)\mathcal{P}(B^c)} \\ &= \frac{0,998\cdot 0,0001}{0,998\cdot 0,0001 + 0,002\cdot 0,9999} \\ &\approx 0,0475 \end{aligned}$$

Kapitel 3

Das Maß-Integral

Motivation. Eine Zufallsvariable X modelliere einen Würfelwurf mit einem echten Würfel. Unter dem Erwartungswert von X versteht man das gewichtete Mittel:

$$\sum_{i=1}^{6} i\mathcal{P}(\{X=i\}) = \sum_{i=1}^{6} \frac{1}{6} i = 3,5.$$

Um dies zu verallgemeinern wollen wir Integrale der Form

$$\int_{\mathbb{R}} x\mathcal{P}_x(dx) \text{ bzw. } \int_{\Omega} f(\omega)\mu(\omega)$$

einen Sinn geben.

3.1 Konstruktion der Maßintegrale

Schritt 1: Für nicht negative Treppenfunktionen:

3.1 Definition. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und X eine nicht negative Treppenfunktion in Normalform, d.h.:

$$X = \sum_{i=1}^{m} \alpha_i \mathbb{1}_{A_i}, A_i \in \mathcal{A} \text{ paarweise disjunkt mit } \bigcup_{i=1}^{m} A_i = \Omega, \alpha_i \geq 0$$

Dann wird

$$\int_{\Omega} X(\omega)\mu(d\omega) := \sum_{i=1}^{m} \alpha_i \mu(A_i)$$

das *Maßintegral von X bezüglich μ* genannt (Alternativ: *μ-Integral von X*, Kurznotation: $\int X d\mu$).

Bemerkung.
- Analog zum Beweis von Satz 1.18 sieht man, dass das μ-Integral wohldefiniert ist (d.h. eine Definition hängt nicht von der speziellen Wahl von α_i und A_i ab).
- Hier ist $\int X d\mu = +\infty$ möglich, falls μ nicht endlich ist.
- Wir bezeichnen die Menge aller Treppenfunktionen in Normalform mit ε.

3.2 Lemma. Seien $X, Y \in \varepsilon, \alpha \geq 0, A \in \mathcal{A}$. Dann:

(i) $\int \mathbb{1}_A d\mu = \mu(A)$

(ii) $\int \alpha X d\mu = \alpha \int X d\mu$

(iii) $\int (X+Y) d\mu = \int X d\mu + \int Y d\mu$

(iv) $X \leq Y$ μ-fast sicher $\Rightarrow \int X d\mu \leq \int Y d\mu$

Beweis. (i), (ii): Klar

(iii), (iv): Seien $X = \sum\limits_{i=1}^{m} \alpha_i \mathbb{1}_{A_i}, Y = \sum\limits_{j=1}^{n} \beta_i \mathbb{1}_{B_j}$ Darstellungen in Normalform. Dann:

$$X = \sum_{i=1}^{m} \sum_{j=1}^{n} \alpha_i \mathbb{1}_{A_i \cap B_j}, Y = \sum_{j=1}^{n} \sum_{j=1}^{m} \beta_i \mathbb{1}_{A_i \cap B_j}$$

Also:

$$X + Y = \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha_i + \beta_j)\mathbb{1}_{A_i \cap B_j}$$

Somit:

$$\int X + Y d\mu = \sum_{i=1}^{m}\sum_{j=1}^{n}(\alpha_i + \beta_j)\mu(A_i \cap B_j)$$

$$\int X d\mu = \sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\mu(A_i \cap B_j) \qquad (*)$$

$$\int Y d\mu = \sum_{i=1}^{m}\sum_{j=1}^{n}\beta_j\mu(A_i \cap B_j) \qquad (*)$$

Insbesondere folgt also (iii).
Zu (iv): Ist $X \leq Y$ μ-fast sicher, so folgt $\alpha_i \leq \beta_j$ für alle i, j mit $\mu(A_i \cap B_j) > 0$. Somit folgt aus $(*)$:

$$\int Y d\mu \leq \sup_n \int X_n d\mu.$$

$\square$

3.3 Lemma. Sei $(X_n)_{n\in\mathbb{N}} \subset \varepsilon$ eine aufsteigende Folge (mit $X_{n+1} \geq X_n$) und $Y \in \varepsilon$ mit $Y \leq \sup_{n\in\mathbb{N}} X_n$. Dann gilt:

$$\int Y d\mu \leq \sup_n \int X_n d\mu.$$

Beweis. Fixiere ein $\beta \in (0,1)$ und setze $B_n = \{X_n \geq \beta Y\}$. Dann gilt $B_n \subset B_{n+1}$ und $\bigcup_{n\in\mathbb{N}} B_n = \Omega$. Also gilt für alle $A \in \mathcal{A}$ aufgrund der Stetigkeit von μ von unten: $\lim_{n\to\infty} \mu(A \cap B_n) = \mu(A)$.
Sei $Y = \sum_{j=1}^{m} \alpha_j \mathbb{1}_{A_j}$ eine Darstellung von Y in Normalform . Dann:

$$\int Y d\mu = \sum_{j=1}^{m}\alpha_j\mu(A_j) = \lim_{n\to\infty}\sum_{j=1}^{m}\alpha_j\mu(A_j \cap B_n) = \lim_{n\to\infty}\int Y\mathbb{1}_{B_n} d\mu.$$

Da $X_n \geq \beta Y \mathbb{1}_{B_n}$ folgt aus Lemma 3.4 (ii) und (iv): $\int X_n d\mu \geq \beta \int Y\mathbb{1}_{B_n} d\mu$. Somit:

$$\sup_{n\in\mathbb{N}} \int X_n d\mu \geq \lim_{n\to\infty} \beta \int Y\mathbb{1}_{B_n} d\mu = \beta \int Y d\mu.$$

Mit $\beta \nearrow 1$ folgt die Behauptung. $\square$

Bemerkung. Auf der rechten Seite in Lemma 3.3 kann „sup" durch „lim" ersetzt werden (nach Lemma 3.2 (iv)).

Schritt 2: Für nicht negative numerische Zufallsvariablen:

Es sei ε^* die Menge aller nicht negativen numerischen Zufallsvariablen.

3.4 Satz. Zu jedem $X \in \varepsilon^*$ existiert eine Folge $(X_n)_{n\in\mathbb{N}}$ in ε mit $X_{n+1} \geq X_n$ und $X = \sup_{n\in\mathbb{N}} X_n$.

Beweis. Seien

$$A_{i,n} = \begin{cases} \{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\} & , i = 0, ..., n2^n - 1 \\ \{X \geq n\} & , i = n2^n \end{cases}$$

Dann sind $A_{i,n} \in \mathcal{A}$ für alle i, n und für ein festes n sind alle Mengen $(A_{i,n})_i$ paarweise disjunkt mit $\bigcup_i A_{i,n} = \Omega$. Also sind $(X_n)_{n\in\mathbb{N}}$ definiert durch:

$$X_n = \sum_{i=1}^{n2^n} \frac{i}{2^n}\mathbb{1}_{A_{i,n}}.$$

Nun gilt für alle $\omega \in \Omega$: $\sup_n X_n(\omega) = X(\omega)$. Denn ist $X(\omega) = +\infty$, dann folgt $X_n(\omega) = 1$ und also: $\sup_n X_n(\omega) = +\infty$. Ist $X(\omega) < +\infty$, so gilt für alle $n > X(\omega)$, dass $X_n(\omega) \leq X(\omega) \leq X_n(\omega) + 2^{-n}$ und also: $\sup_{n\in\mathbb{N}}(X_n(\omega)) = X(w)$. $\square$

3.5 Satz. Sowohl Satz 3.5 als auch sein Beweis sind Bestandteil eines der kommenden Übungsblätter.

Schritt 3: Sei $X \in \varepsilon^*$ und $(X_n)_{n\in\mathbb{N}}$ eine nichtfallende Folge in ε mit $\sup_{n\in\mathbb{N}} X_n = X$. Dann wird

$$\int_\Omega X(\omega)\mu(d\omega) := \sup_{n\in\mathbb{N}} X_n(\omega)\mu(d\omega)$$

das μ-Integral der Zufallsvariable X genannt.

Bemerkung. (i) Das Maßintegral μ ist wohldefiniert. Sind $(X_n)_{n\in\mathbb{N}}$ und $(Y_n)_{n\in\mathbb{N}}$ approximierend Folgen für X in ε (d.h. nichtfallend mit $\sup_{n\in\mathbb{N}} X_n = \sup_{n\in\mathbb{N}} Y_n = X$), dann gilt nach Lemma 3.3 für jedes $n \in \mathbb{N}$: $\int Y_m \leq \sup_n \int X_n d\mu$. Also:

$$\sup_{m\in\mathbb{N}} \int Y_m d\mu \leq \sup_{n\in\mathbb{N}} \int X_n d\mu.$$

Durch Vertauschen der Rollen von (Y_n) und (X_n) folgt

$$\sup_{n\in\mathbb{N}} \int Y_m d\mu = \sup_{n\in\mathbb{N}} \int X_n d\mu.$$

(ii) Sei $X \in \varepsilon$. Dann ist $(X_n)_{n\in\mathbb{N}}$ mit $X_n = X$ für alle $n \in \mathbb{N}$ eine approximierende Folge für X. Also stimmen Definition 3.5 für Treppenfunktionen mit der Definition 3.1 überein.

3.6 Lemma. Die Eigenschaften (ii)-(vi) aus Lemma 3.2 gelten auch für Integranden aus ε^*.

Beweis. Übung □

3.7 Satz. (Beppo-Levi)
Sei $(X_n)_{n\in\mathbb{N}}$ eine nichtfallende Folge in ε^*. Dann gilt: $\sup_{n\in\mathbb{N}} X_n \in \varepsilon^*$ und

$$\int \sup_{n\in\mathbb{N}} X_n d\mu = \sup_{n\in\mathbb{N}} \int X_n d\mu.$$

Beweis. Zu jedem X_n sei $(X_{n,m})_{m\in\mathbb{N}}$ eine nichtfallende Folge in ε mit $\sup_{m\in\mathbb{N}} X_{n,m} = X_n$ (s. Satz 3.4). Sei

$$Y_m := \sup(X_{n,m}, ..., X_{m,m}).$$

Dann ist $(Y_m)_{m\in\mathbb{N}}$ eine nichtfallende Folge in ε. Da $(X_n)_{n\in\mathbb{N}}$ nichtfallend ist, folgt:

$$X_m = \sup(X_1, ..., X_m) \geq \sup(X_{1,m}, ..., X_{m,m}) = Y_m.$$

Also:

$$\sup_{m\in\mathbb{N}} Y_m \leq \sup_{m\in\mathbb{N}} X_m.$$

Andererseits gilt für $m \geq n$ $Y_m \geq X_{n,m}$, d.h.

$$X_n = \sup_{m\in\mathbb{N}} X_{n,m} \leq \sup_{m\in\mathbb{N}} Y_m.$$

Zusammen erhalten wir: $\sup_{m\in\mathbb{N}} Y_m = \sup_{n\in\mathbb{N}} X_n$. Nach Definition des μ-Integrals folgt:

$$\int (\sup_{n\in\mathbb{N}} X_n) d\mu = \sup_{m\in\mathbb{N}} \int Y_m d\mu \overset{2.6\ (iv)}{\leq} \sup_{m\in\mathbb{N}} \int X_m d\mu.$$

Die andere Ungleichung folgt aus Lemma 3.6 (iv), da $X_m \leq \sup_{n\in\mathbb{N}} X_n$ und also:

$$\int X_m d\mu \leq \int (\sup_{n\in\mathbb{N}} X_n) d\mu \Rightarrow \int X_m d\mu \leq \int (\sup_{n\in\mathbb{N}} X_n) d\mu.$$

□

Schritt 4: Für numerische Zufallsvariablen:

3.8 Definition. Sei X eine numerische Zufallsvariable und seien

$$X_+ := \max\{0, X\} \text{ und } X_- := \max\{0, -X\}.$$

Ist $\int X_+ d\mu < \infty$ oder $\int X_- d\mu < \infty$, so wird X *μ-quasi-integrierbar* genannt und definiert das μ-Integral von X durch

$$\int X d\mu = \int X_+ d\mu - \int X_- d\mu.$$

Ist $\mu = \mathcal{P}$ ein Wahrscheinlichkeitsmaß, so notiert man auch

$$E[X] := \int X d\mathcal{P}$$

und nennt $E[X]$ den *Erwartungswert* von X.

Bemerkung. Ist X μ-integrierbar, so gilt $\int X d\mu \in \mathbb{R}$.

3.9 Satz. Die folgenden Aussagen sind äquivalent für eine numerische Zufallsvariable:

(i) X ist μ-integrierbar.

(ii) Es gibt μ-integrierbare $Y_1, Y_2 \in \varepsilon^*$ mit $X = Y_1 - Y_2$.

(iii) Es gibt ein μ-integrierbares $Z \in \varepsilon^*$ mit $|X| \leq Z$.

(iv) $|X|$ ist μ-integrierbar.

Gilt (ii) so gilt insbesondere: $\int X d\mu = \int Y_1 d\mu - \int Y_2 d\mu$.

Beweis. **(i)⇒(ii)** Setze $Y_1 = X_+$ und $Y_2 = X_-$.

(ii)⇒(iii) Sei $Z := Y_1 + Y_2$. Dann gilt $Z \in \varepsilon^*$ und Z ist μ-integrierbar (wegen der Linearität des μ-Integrals auf ε^* (Lemma 3.6)).

$$\left.\begin{aligned} +X &= Y_1 - Y_2 \leq Y_1 + Y_2 = Z \\ -X &= Y_2 - Y_1 \leq Y_2 + Y_1 = Z \end{aligned}\right\} \Rightarrow |X| = Z$$

(iii)⇒(iv) Nach Lemma 3.6 (iv) folgt $\int |X| d\mu \leq \int Z d\mu < \infty$.

(iv)⇒(i) Wie im vorherigen Schritt folgt aus

$$\begin{cases} X_+ \leq |X| \\ X_- \leq |X| \end{cases},$$

dass X_+ und X_- μ-integrierbar sind und also:

$$\int X_+ d\mu < \infty \text{ und } \int X_- d\mu < \infty.$$

Es gelte nun (ii) und damit auch (i). Dann: $X = Y_1 - Y_2 = X_+ - X_-$. Also: $Y_1 + X_- = Y_2 + X_+$. Aus der Linearität des μ-Integrals auf ε^* (Lemma 3.6) folgt:

$$\int Y_1 d\mu + \int X_- d\mu = \int Y_2 d\mu + \int X_+ d\mu$$

$$\Rightarrow \int X d\mu = \int X_+ d\mu - \int X_- d\mu = \int Y_1 d\mu - \int Y_2 d\mu.$$

□

3.10 Satz. Seine X, Y μ-integrierbare numerische Zufallsvariablen, $\alpha \in \mathbb{R}$. Dann sind $X + Y$ (sofern überall definiert), αX, $\max\{X, Y\}$, $\min\{X, Y\}$ μ-integrierbar und es gilt:

$$\int \alpha X d\mu = \alpha \int X d\mu \text{ und } \int X + Y d\mu = \int X d\mu + \int Y d\mu.$$

Beweis.

- Da
$$(\alpha X)_+ = \begin{cases} \alpha X_+ & ,\alpha \geq 0 \\ \alpha X_- & ,\alpha < 0 \end{cases}$$
$$(\alpha X)_- = \begin{cases} \alpha X_- & ,\alpha \geq 0 \\ \alpha X_+ & ,\alpha < 0 \end{cases}$$
folgt aus Lemma 3.6 etwa für $\alpha < 0$:
$$\int (\alpha X)_+ d\mu = -\alpha \int X_- d\mu < \infty$$
$$\int (\alpha X)_- d\mu = -\alpha \int X_+ d\mu < \infty$$
Also ist αX μ-integrierbar und
$$\int \alpha X d\mu = -\alpha \int X_- d\mu - (-\alpha \int X_+ d\mu) = \alpha \int X d\mu.$$
- $X + Y = (X_+ + Y_+) - (X_- - Y_-)$. Nach Lemma 3.6 sind positiver Teil $X_+ + Y_+$ und ngativer Teil $X_- - Y_-$. Somit folgt aus Satz 3.10 (ii), dass $X + Y$ μ-integrierbar ist und :
$$\begin{aligned} \int (X+Y) d\mu &= \int X_+ + Y_+ d\mu - \int X_- + Y_- d\mu \\ &\overset{3.6}{=} \int X_+ d\mu + \int Y_+ d\mu - \int X_- d\mu - \int Y_- d\mu \\ &= \int X d\mu + \int Y d\mu \end{aligned}$$
- $$|\max\{X, Y\}| \leq |X| + |Y|$$
$$|\min\{X, Y\}| \leq |X| + |Y|$$
Da mit X und Y auch $|X|$ und $|Y|$ (Satz 3.10) und somit auch $|X| + |Y|$ μ-integrierbar ist, ist $|X| + |Y|$ jeweils eine μ-integrierbare Majorante und die μ-Integrierbarkeit folgt wiederum aus Satz 3.10. □

3.11 Satz. Sind X und Y μ-integrierbare numerische Funktionen mit $X \leq Y$ so folgt:
$$\int X d\mu \leq \int Y d\mu,$$
insbesondere also:
$$\left| \int X d\mu \right| \leq \int |X| d\mu.$$

Beweis. Aus $X \leq Y$ folgt $X_+ \leq Y_+$ und $X_- \geq Y_-$. Also ist (Satz 3.6)
$$\int X d\mu = \int X_+ d\mu - \int X_- d\mu \leq \int Y_+ d\mu - \int Y_- d\mu = \int Y d\mu.$$
□

3.2 Maße mit Dichten und der Transformationssatz

3.12 Satz. (i) Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $X \in \varepsilon^*$. Dann sind durch
$$\nu(A) := \int_A X d\mu := \int_\Omega \mathbb{1}_A \cdot X \, d\mu, A \in \mathcal{A},$$
ein Maß auf $(\Omega, \mathcal{A})$ definiert. X wird *(Radon-Nikodym-) Dichte von ν bezüglich μ* genannt.
ν ist ein Wahrscheinlichkeitsmaß genau dann, wenn:
$$\int_\Omega X d\mu = 1.$$

(ii) Eine numerische Zufallsvariable Y ist ν-integrierbar genau dann, wenn $Y \cdot X$ μ-integrierbar ist. In diesem Fall gilt:
$$\int Y d\nu = \int X \cdot Y \, d\mu.$$

Bemerkung. Für die Dichte X verwendet man häufig die suggestive Schreibweise $X = \frac{d\nu}{d\mu}$. Dann wird (ii) zu:

$$\int Y d\nu = \int Y \frac{d\nu}{d\mu} d\mu.$$

Beweis. (i) Es ist $\nu(\varnothing) = \int 0 d\mu = 0$ und $\nu(A) = \int \underbrace{\mathbb{1}_A X}_{\geq 0} d\mu \geq 0$.

Ferner gilt: $\nu(\Omega) = \int X d\mu$ und also $\nu(\Omega) = 1 \Leftrightarrow \int X d\mu = 1$.
Zur σ-Additivität: Seien $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ paarweise disjunkt. Dann ist die Folge $(X\mathbb{1}_{\bigcup_{k=1}^n A_k})_{n\in\mathbb{N}}$ in ε^* nichtfallend. Also folgt aus dem Satz von Beppo-Levi (Satz 3.7):

$$\begin{aligned}\nu(\bigcup_{k=1}^\infty A_k) &= \int X \cdot \mathbb{1}_{\bigcup_{k=1}^\infty A_k} d\mu = \int \sup_{n\in\mathbb{N}} (X \cdot \mathbb{1}_{\bigcup_{k=1}^n A_k}) d\mu = \sup_{n\in\mathbb{N}} \int X \cdot \mathbb{1}_{\bigcup_{k=1}^n A_k} d\mu \\ &= \sup_{n\in\mathbb{N}} \int X \cdot (\sum_{k=1}^n \mathbb{1}_{A_k}) d\mu = \sup_{n\in\mathbb{N}} \sum_{k=1}^n \int X \cdot \mathbb{1}_{A_k} d\mu = \sum_{k=1}^\infty \nu(A_k).\end{aligned}$$

(ii) Sei $Y = \sum_{i=1}^m a_i \mathbb{1}_{A_i} \in \varepsilon$. Dann:

(1) $\int Y d\nu = \sum_{i=1}^m a_i \nu(A_i) = \sum_{i=1}^m a_i \int \mathbb{1}_{A_i} X d\mu = \int (\sum_{i=1}^m a_i \mathbb{1}_{A_i}) X d\mu = \int Y X d\mu$

(2) Sei $Y \in \varepsilon^*$ und $(Y_n)_{n\in\mathbb{N}}$ eine nichtfallende approximierende Folge in ε für Y. Dann ist $(Y_n X)_{n\in\mathbb{N}}$ eine nichtfallende Folge in ein ε^* mit $\sup_{n\in\mathbb{N}} (Y_n X) = YX$. Also folgt aus (1) und dem Satz von Beppo-Levi:

$$\int XY d\mu = \sup_{n\in\mathbb{N}} \int XY_n d\mu \underset{(1)}{=} \sup_{n\in\mathbb{N}} \int Y_n d\nu = \int Y d\nu.$$

(3) Sei Y eine numerische Zufallsvariable. Dann folgt aus (2):

$$\int Y_+ d\nu = \int Y_+ X d\mu = \int (YX)_+ d\mu \text{ und } \int Y_- d\nu = \int (YX)_- d\mu.$$

Damit liefert die Definition des Maßintegrals die Behauptung. □

3.13 Bemerkung. (a) Sei $(\Omega, \mathcal{A}) = (\mathbb{R}^D, \mathcal{B}_D)$ und $\lambda^{\otimes D}$ das D-dimensionale Lebesgue-Maß auf $(\mathbb{R}^D, \mathcal{B}_D)$. Dann sprechen wir vom *Lebesgue-Integral* von X, falls X $\mathcal{B}_D/\mathcal{B}$-messbar ist und $\int_{\mathbb{R}^D} X d\lambda^{\otimes D}$ existiert. Ist f über einer Jordan-messbaren (d.h. $\mathbb{1}_A$ ist Riemann-integrierbar)Teilmenge von $\mathbb{R}^D$ eigentlich Riemann-integrierbar und ist $\mathbb{1}_A f$ $\mathcal{B}_D/\mathcal{B}$-messbar, dann ist $\mathbb{1}_A f$ Lebesgue-integrierbar und $\int \mathbb{1}_A f d\lambda^{\otimes D} = \int_A f(x) dx$ ($\rightarrow$Riemann-Integral).

(b) Hat also eine Maß ν auf $(\mathbb{R}^D, \mathcal{B}_D)$ eine Dichte f bezüglich des Lebesguemaßes und ist $\mathbb{1}_A f$ für ein $A \in \mathcal{B}_D$ eigentlich Riemann-integrierbar, so folgt

$$\nu(A) = \int_A f(x) dx.$$

Hat andererseit ein Wahrscheinlichkeitsmaß $\mathcal{P}_f$ auf $(\mathbb{R}, \mathcal{B})$ eine $\mathcal{B}/\mathcal{B}$-messbare Riemann-Dichte f im Sinn von Beispiel 1.36, so gilt für alle $a < b \in \mathbb{R}$

$$\mathcal{P}_f((a, b]) = \int_a^b f(x) dx = \int \mathbb{1}_{(a,b]} f d\lambda.$$

Nach dem Eindeutigkeitsresultat aus Satz 1.21 folgt

$$\mathcal{P}_f(A) = \int \mathbb{1}_A f d\lambda, \text{ für alle } A \in \mathcal{B}.$$

Also ist f dann auch Radon-Nikodym-Dichte von $\mathcal{P}_F$ bezüglich λ.

Eine weitere Möglichkeit zur Berechnung von Integralen liefert folgender Transformationssatz:
Dazu sein $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, $(\Omega', \mathcal{A}')$ ein Messraum, $T : \Omega \rightarrow \Omega'$ $\mathcal{A}/\mathcal{A}'$-messbar. Das *Bildmaß* von μ unter T auf $(\Omega', \mathcal{A}')$ ist definiert durch

$$\mu_T(A') = \mu(T^{-1}(A')), A' \in \mathcal{A}'.$$

Ist μ ein Wahrscheinlichkeitsmaß, so ist μ_T gerade die Verteilung von T. In dieser Situation gilt:

3.14 Satz. Ist $X : (\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ eine numerische Zufallsvariable, so ist X μ-messbar genau dann, wenn $X \circ T$ μ-integrierbar ist, und in diesem Fall gilt:

$$\int_{\Omega'} X d\mu = \int_{\Omega} X \circ T d\mu.$$

Beweis. (1) $X = \sum_{i=1}^{m} a_i \mathbb{1}_{A_i'} \in \varepsilon$ $(A_i' \in \mathcal{A}')$. Dann: $X \circ T = \sum_{i=1}^{m} a_i \mathbb{1}_{T^{-1}(A_i')}$. Somit:

$$\int_{\Omega'} X d\mu_T = \sum_{i=1}^{m} a_i \mu_T(A_i') = \sum_{i=1}^{m} a_i \mu(T^{-1}(A_i')) = \int_{\Omega} X \circ T d\mu.$$

(2) $X \in \varepsilon^*$ und $(X_n)_{n\in\mathbb{N}}$ eine nichtfallende Folge in ε mit $\sup_{n\in\mathbb{N}} X_n = X$. Dann ist $(X_n \circ T)$ eine nichtfallende Folge für $X \circ T$.

(3) Sei X eine numerische Zufallsvariable. Dann folgt die Aussage aus (2), da $(X \circ T)_\pm = X_\pm \circ T$.

□

3.15 Beispiel. Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum, X eine $\mathbb{R}^D$ -wertige Zufallsvariable, $g : \mathbb{R}^D \to \mathbb{R}$ messbar. Die Verteilung von X habe eine Dichte f bezüglich des Lebesguemaßes $\lambda^{\otimes D}$. Dann gilt nach den Sätzen 3.13 und 3.15:

$$g(X) \text{ ist } \mathcal{P}\text{-integrierbar } \Leftrightarrow g\ \mathcal{P}_X\text{-integrierbar } \Leftrightarrow g \cdot f\ \lambda^{\otimes D}\text{-integrierbar.}$$

In diesem Fall gilt:

$$E[g(X)] = \int_{\mathbb{R}^D} g\, d\mathcal{P}_X = \int_{\mathbb{R}^D} g \cdot f\, d\lambda^{\otimes D}.$$

3.16 Beispiel. Sei X eine reellwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$. Es gebe eine höchsten abzählbare Menge $Y \subset \mathbb{R}$ mit $\mathcal{P}(\{X \in Y\}) = 1$. Dann gilt für jede messbare Funktion $g : \mathbb{R} \to \mathbb{R}$:

$$g \circ X\ \mathcal{P}\text{-integrierbar } \Leftrightarrow \sum_{y\in Y} |g(y)|\, \mathcal{P}(\{X = y\}) < \infty.$$

In diesem Fall gilt:

$$E[g(X)] = \sum_{y\in Y} g(y) \mathcal{P}(\{X = y\}).$$

Denn aus Satz 3.15 folgt:

$$g \circ X\ \mathcal{P}\text{-integrierbar } \Leftrightarrow g \text{ ist } \mathcal{P}_{\mathcal{X}}\text{-integrierbar.}$$

Es reicht also zu zeugen, dass für nicht-integrierbares g gilt:

$$\int g\, d\mathcal{P}_{\mathcal{X}} = \sum_{y\in Y} g(y)\, \mathcal{P}(\{X = y\})$$

__Schritt 1:__ $g = \sum_{i=1}^{m} a_i \mathbb{1}_{A_i} \in \varepsilon$

$$\int g\, d\mathcal{P}_{\mathcal{X}} = \sum_{i=1}^{m} a_i \mathcal{P}(A_i) = \sum_{i=1}^{m} a_i \sum_{y\in A_i\cap Y} \mathcal{P}_{\mathcal{X}}(\{y\}) = \sum_{i=1}^{m} \sum_{y\in A_i\cap Y} g(y)\, \mathcal{P}_{\mathcal{X}}(\{y\}) = \sum_{y\in Y} g(y)\, \mathcal{P}(\{X = y\}).$$

__Schritt 2:__ $g \in \varepsilon^*$

Sei $(g_n)_{n\in\mathbb{N}}$ eine nichtfallende approximierende Folge für g in ε. Dann:

$$\int g\, d\mathcal{P}_{\mathcal{X}} = \sup_{n\in\mathbb{N}} \int g_n\, d\mathcal{P}_{\mathcal{X}} \underset{\text{Schritt 1}}{=} \sup_{n\in\mathbb{N}} \sum_{y\in Y} g_n(y)\, \mathcal{P}(\{X = y\}) = \sum_{y\in Y} g(y)\, \mathcal{P}(\{X = y\}).$$

Hier ist „$\leq$“ klar. Ist $(g_k)_{k\in\mathbb{N}}$ eine Abzählung von Y, so folgt „$\geq$“ aus:

$$\sup_{n\in\mathbb{N}} \sum_{y\in Y} g_n(y)\, \mathcal{P}(\{X = y\}) \geq \sup_{n\in\mathbb{N}} \sum_{k=1}^{K} g_n(y_k)\, \mathcal{P}(\{X = y_k\}) = \sum_{k=1}^{K} g(y_k)\, \mathcal{P}(\{X = y\})$$

und dem Grenzwert $k \to \infty$.

3.17 Beispiel. Seien X und Y unabhängige standardnormalverteilte Zufallsvariablen und U, V seien $[0,1)$-wertige Zufallsvariablen, s.d.

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \sqrt{-2\ln(1-V)} \begin{pmatrix} \cos 2\pi U \\ \sin 2\pi U \end{pmatrix}.$$

Dann:

$$\begin{aligned} \mathcal{P}_{(X,Y)}((a_1,b_1]\times(a_2,b_2]) &= \mathcal{P}(\{X\in(a_1,b_1]\}\cap\{Y\in(a_2,b_2]\}) \\ &\overset{\text{unabh.}}{=} \mathcal{P}(\{X\in(a_1,b_1]\})\mathcal{P}(\{Y\in(a_2,b_2]\}) \\ &= \int_{a_1}^{b_1} \frac{1}{2\pi} e^{-\frac{x^2}{2}}\,dx \int_{a_2}^{b_2} \frac{1}{2\pi} e^{-\frac{y^2}{2}}\,dy \\ &= \int_{(a_1,b_1]\times(a_2,b_2]} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\,d(x,y) \end{aligned}$$

Wie in Bemerkung 3.14 folgt daher, dass $\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$ Dichte von (X,Y) bezüglich des Lebesguemaßes $\lambda^{\otimes 2}$ ist.
Sei nun $u, v \in [0,1)$ und sei $g(r,\varphi) = (r\cos\varphi, r\sin\varphi)$ die Polarkoordinaten Abbildung. Dann gilt:

$$\begin{aligned} &\mathcal{P}(\{U\le u\}\cap\{V\le v\}) = \mathcal{P}(\{0\le U\le u\}\cap\{0\le V\le v\}) \\ &\mathcal{P}\left(\left\{\begin{pmatrix} X \\ Y \end{pmatrix} \in g([0,\sqrt{-2\ln(1-v)}]\times[0,2\pi u])\right\}\right) \\ &= E\left[\mathbb{1}_{\left(\begin{smallmatrix} X \\ Y \end{smallmatrix}\right)\in g([0,\sqrt{-2\ln(1-v)}]\times[0,2\pi u])}\right] \\ &\overset{\substack{\text{Bsp. 3.16} \\ \text{Bem. 3.14}}}{=} \int_{\{\left(\begin{smallmatrix} X \\ Y \end{smallmatrix}\right)\in g(\dots)\}} \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}\,d(x,y) \\ &\overset{\substack{\text{Transformations-} \\ \text{satz (Analysis)}}}{=} \int_{[0,\sqrt{-2\ln(1-v)}]\times[0,2\pi u]} \frac{1}{2\pi} e^{-\frac{r^2}{2}}\,r\,d(r,\varphi) \\ &\overset{\text{Fubini}}{=} \int_0^{\sqrt{-2\ln(1-v)}} r e^{-\frac{r^2}{2}}\,dr \int_0^{2\pi u} \frac{1}{2\pi}\,d\varphi \\ &= -((1-v)-1)u = vu \end{aligned}$$

Mit $u \to 1$ (bzw. $v \to 1$) folgt, dass:

$$F_v(x) = F_u(x) = \begin{cases} 0 & , x < 0 \\ x & , x \in [0,1) \\ 1 & , x \ge 1 \end{cases}$$

Also sind U und V auf $[0,1)$ gleichverteilt. Ferner sind dann U und V unabhängig (nach Satz 2.18), da die gemeinsame Verteilungsfunktion

$$F_{(U,V)}(u,v) = F_U(u)F_V(v)$$

zunächst für $(u,v) \in [0,1]^2$ und dann auch für $(u,v) \in \mathbb{R}^2$.
Insbesondere folgt, dass Winkel und Längen von $\left(\begin{smallmatrix} X \\ Y \end{smallmatrix}\right)$ unabhängig sind. Dreht man die Argumentation um, so sieht man, dass für unabhängige auf $[0,1]$ gleichverteilte Zufallsvariablen auch

$$X = \sqrt{-2\ln(1-V)}\cos(2\pi U) \qquad Y = \sqrt{-2\ln(1-V)}\sin(2\imath U)$$

zwei unabhängige standardnormalverteilte Zufallsvariablen sind. Dies ist das *Box-Miller-Verfahren* zur Erzeugung normalverteilter Zufallsvariablen.

3.3 L^p-Räume und Konvergenzsätze

3.18 Definition. Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $p \in [1,\infty)$. Ein messbare, reellwertige Funktion $X : \Omega \to \mathbb{R}$ heißt *p-fach integrierbar*, falls $|X|^p$ μ-integrierbar ist. Wir notieren de Menge der p-fach integrierbaren Funktionen mit $\mathcal{L}^p(\mu)$ bzw. $\mathcal{L}^p(\Omega,\mathcal{A},\mu)$. Ferner sei für $X \in \mathcal{L}^p(\mu)$:

$$\|X\|_p := \left(\int |X|^p d\mu\right)^{\frac{1}{p}}.$$

3.19 Satz. $\mathcal{L}^p(\mu)$ ist ein Vektorraum und $\|.\|_p$ ist eine Halbnorm auf $\mathcal{L}^p(\mu)$.

Zur Vorbereitung benötigen wir einige Resultate:

3.20 Lemma. Sei $X \in \varepsilon^*$, dann gilt:

$$\int X d\mu = 0 \Leftrightarrow X = 0 \ \mu\text{-f.s..}$$

Beweis. $N := \{X \neq 0\} \in \mathcal{A}$, da X messbar ist.

„$\Leftarrow$“ Sei zunächst $X = 0$ μ-f.s., also $\mu(N) = 0$. Sei $Y_n = \mathbb{1}_N$. Dann ist $(Y_n)_{n\in\mathbb{N}}$ eine Folge in ε und $Y := \sup\limits_n Y_n \in \varepsilon^*$. Da $0 \leq X \leq Y$, folgt

$$0 \leq \int X d\mu \leq \int Y d\mu = \sup_{n\in\mathbb{N}} \int Y_n d\mu = \sup_{n\in\mathbb{N}}(n\mu(N)) = 0 \longrightarrow \int X d\mu = 0.$$

„$\Rightarrow$“ Sei $\int X d\mu = 0$. Sei $A_n = \{X \geq \frac{1}{n}\}$. Dann ist $(A_n)_{n\in\mathbb{N}}$ nichtfallend in $\mathcal{A}$ mit $\bigcup A_n = N$. Also folgt mit der Stetigkeit von oben: $\mu(N) = \lim\limits_{n\to\infty} \mu(A_n)$. Da $X \geq \frac{1}{n}\mathbb{1}_{A_n}$, folgt

$$0 = \int X d\mu \geq \int \frac{1}{n}\mathbb{1}_{A_n} d\mu = \frac{1}{n}\mu(A_n).$$

Also $\mu(A_n) = 0 \Rightarrow Y(\{X \neq 0\}) = 0$. □

3.21 Folgerung. Seien X, Y numerische messbare Funktionen mit $X = Y$-f.s., dann gilt:

(i) $X, Y \in \varepsilon^* \Rightarrow \int X d\mu = \int Y d\mu$

(ii) X μ-integrierbar $\Rightarrow$ Y μ-integrierbar mit $\int X d\mu = \int Y d\mu$

Beweis. (i) Sei $N := \{X \neq Y\}$ und nach Voraussetzung $\mu(N) = 0$. Nach Lemma 3.20 folgt:

$$\int X\mathbb{1}_N d\mu = 0 = \int Y\mathbb{1}_N d\mu.$$

Ferner:

$$\int X\mathbb{1}_{N^c} d\mu = \int Y\mathbb{1}_{N^c} d\mu.$$

Dann folgt die Aussage aus der Linearität des Integrals.

(ii) Folgt aus (i), da dann $X_+ = Y_+$ μ-f.s. und $X_- = Y_-$ μ-f.s..

□

3.22 Folgerung. Seien X, Y messbare Funktionen mit $|X| \leq Y$ μ-f.s., dann gilt:

(i) Y μ-integrierbar $\Rightarrow$ X μ-integrierbar

(ii) X μ-integrierbar $\Rightarrow$ $|X| < \infty$ μ-f.s.

Beweis. (i) Sei $\tilde{Y} = \max\{Y, |X|\}$. Dann ist $\tilde{Y} \in \varepsilon^*$ und $Y = \tilde{Y}$, also ist X μ-integrierbar (nach 3.10).

(ii) Sei $N = \{|X| = +\infty\}$. Dann $N \in \mathcal{A}$ und $\mathbb{1}_N \leq \frac{|X|}{\alpha}$ für alle $\alpha > 0$. Also:

$$\mu(N) = \int \mathbb{1}_N d\mu \leq \frac{1}{\alpha}\underbrace{\int |X| d\mu}_{<\infty}.$$

Mit $\alpha \nearrow +\infty$ folgt $\mu(N) = 0$. □

3.23 Satz. (Hölder-Ungleichung)
Seien $1 < p < \infty$ und q mit $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt für alle numerischen messbaren Funktionen X und Y:

$$\int |XY| d\mu \leq \left(\int |X|^p d\mu\right)^{\frac{1}{p}} \left(\int |Y|^q d\mu\right)^{\frac{1}{q}}.$$

Beweis. Œ:

- $X, Y \in \varepsilon^*$
- $\delta := \left(\int |X|^p d\mu\right)^{\frac{1}{p}} \neq 0 \neq \left(\int |Y|^q d\mu\right)^{\frac{1}{q}} =: \tau$
- $\delta < \infty, \tau < \infty$, da die Ungleichung sonst trivial ist.

Anwendung des Mittelwertsatzes auf die Funktion $f(y) = (1+y)^{\frac{1}{p}}$ liefert $(1+y)^{\frac{1}{p}} - 1 \leq \frac{1}{p}y$ für $y \geq 0$ und somit $z^{\frac{1}{p}} \leq \frac{1}{p}z + \frac{1}{q}$.
Seien $\alpha, \beta > 0$. Dann ist $\alpha\beta^{-1} > 1$ oder $\beta\alpha^{-1} \geq 1$. Œ sei $\alpha\beta^{-1} \geq 1$. Dann

$$\alpha^{\frac{1}{p}}\beta^{\frac{1}{q}} = (\frac{\alpha}{\beta})^{\frac{1}{p}}\beta \leq (\frac{1}{p}(\frac{\alpha}{\beta}) + \frac{1}{q})\beta = \frac{\alpha}{p} + \frac{\beta}{q}.$$

Insbesondere folgt: $\frac{1}{\delta\tau}|XY| \leq \frac{|X|^p}{\delta^p p} + \frac{|Y|^q}{\tau^q q}$. Somit:

$$\int |XY| d\mu \leq \delta\tau(\frac{1}{p} + \frac{1}{q}) = \left(\int |X|^p d\mu\right)^{\frac{1}{p}} \left(\int |Y|^q d\mu\right)^{\frac{1}{q}}.$$

□

3.24 Folgerung. Ist μ ein endliches Maß und gilt $p_1 > p_2$, so folgt $\mathcal{L}^{p_1}(\mu) \subset \mathcal{L}^{p_2}(\mu)$.

Beweis. Sei $X \in \mathcal{L}^{p_1}(\mu)$. Wir verwenden die Hölder-Ungleichung mit $p = \frac{p_1}{p_2}$ und $\frac{1}{q} + \frac{1}{p} = 1$. Dann

$$\int |X|^{p_2} d\mu \leq \left(\int |X|^{p_2 p} d\mu\right) \frac{1}{p} \left(\int 1^q d\mu\right)^{\frac{1}{q}} = \underbrace{\left(\int |X|^{p_1} d\mu\right)^{\frac{1}{p}}}_{<\infty} \underbrace{(\mu(\Omega))^{\frac{1}{q}}}_{<\infty} < \infty.$$

□

3.25 Satz. (Minkowski-Ungleichung)
Seien $X, Y \in \mathcal{L}^p(\mu)$. Dann ist $X + Y \in \mathcal{L}^p(\mu)$ und

$$\left(\int |X+Y|^p d\mu\right)^{\frac{1}{p}} \leq \left(\int |X|^p d\mu\right)^{\frac{1}{p}} + \left(\int |Y|^p d\mu\right)^{\frac{1}{p}}.$$

Beweis. Œ seien X, Y nicht-negativ, da $|X+Y|^p \leq (|X| + |Y|)^p$ gilt:

- $p = 1$: trivial.
- Also $p > 1$. Œ $\int |X+Y|^p d\mu > 0$ Da

$$|X+Y|^p \leq (2\max\{X, Y\})^p \leq 2^p(X^p + Y^p)$$

folgt $X + Y \in \mathcal{L}^p(\mu)$. Sei $\frac{1}{q} + \frac{1}{p} = 1$, also $q = \frac{p}{p-1}$. Dann folgt aus der Hölder-Ungleichung

$$\begin{aligned}\int |X+Y|^p d\mu &= \int X(X+Y)^{p-1} d\mu + \int Y(X+Y)^{p-1} d\mu \\ &\leq \left(\int |X|^p d\mu\right)^{\frac{1}{p}} \left(\int |X+Y|^p d\mu\right)^{\frac{p-1}{p}} + \left(\int |Y|^p d\mu\right)^{\frac{1}{p}} \left(\int |X+Y|^p d\mu\right)^{\frac{p-1}{p}}\end{aligned}$$

$$\left(\int |X+Y|^p d\mu\right)^{\frac{1}{p}} = \frac{\int |X+Y|^p d\mu}{(\int |X+Y|^p d\mu)^{\frac{p-1}{p}}} \leq \left(\int |X|^p d\mu\right)^{\frac{1}{p}} + \left(\int |Y|^p d\mu\right)^{\frac{1}{p}}$$

□

Beweis. (von Satz 3.19)
$\mathcal{L}^p(\mu)$ ist eine Teilmenge des Vektorraums der $\mathcal{A}/\mathcal{B}$-messbaren Funktionen. Nach Satz 3.25 ist $\mathcal{L}^p(\mu)$ bezüglich „+“ abgeschlossen und $\|.\|_p$ erfüllt de Dreiecksungleichung. Sei nun $X \in \mathcal{L}^p(\mu)$ und $\alpha \in \mathbb{R}$. Dann:

$$\left(\int |\alpha X|^p d\mu\right)^{\frac{1}{p}} = \left(|\alpha|^p \int |X|^p d\mu\right)^{\frac{1}{p}} = |\alpha| \left(\int |X|^p d\mu\right)^{\frac{1}{p}}.$$

□

3.26 Bemerkung. Nach Lemma 3.20 gilt: $\|X\|_p = 0 \Leftrightarrow X = 0$ μ-f.s., also $\|.\|_p$ keine Norm auf $\mathcal{L}^p(\mu)$. Betrachtet man aber die Äquivalenzklassen

$$[X] := \{Y \text{ reellwertig, messbar} \,|\, X = Y \;\mu - \text{f.s.}\},$$

so ist

$$\|[X]\|_p = \|X\|_p$$

eine Norm auf dem Vektorraum $L^p(\mu)$ aller Äquivalenzklassen von Funktionen $X \in \mathcal{L}^p(\mu)$ bzgl. $\sim$ mit $X \sim Y :\Leftrightarrow X = Y\mu$-f.s..

Frage: Sind die $L^p(\mu), p \geq 1$, Banachräume?

3.27 Definition. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von numerischen messbaren Funktionen und X numerisch und messbar.

(i) Wir sagen, $(X_n)_{n\in\mathbb{N}}$ *konvergiert* gegen X μ- *fast sicher*, falls eine Menge $N \in \mathcal{A}$ mit $\mu(N) = 0$ existiert, so dass für alle $\omega \in N^c$:
$$\lim_{n\to\infty} X_n(\omega) = X(\omega).$$

(ii) Sind die X_n und X allesamt reellwertig, so heißt $(X_n)_{n\in\mathbb{N}}$ *konvergent in* $\mathcal{L}^p(\mu)$ gegen X, falls
$$\|X_n - X\|_p \to 0 \quad (n \to \infty).$$

In beiden Fällen sind die Grenzwerte eindeutig bis auf Modifikation.

3.28 Satz. (Lemma von Fatou)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in ε^*. Dann:
$$\int (\liminf_{n\to\infty} X_n)\, d\mu \leq \liminf_{n\to\infty} \int X_n\, d\mu.$$

Beweis. Sei $Y_n = \inf_{m\geq n} X_n$. Dann ist $(Y_n)_{n\in\mathbb{N}}$ nach Satz 2.8 eine Folge in ε^*. Da die Folge $(Y_n)_{n\in\mathbb{N}}$ nichtfallend ist, impliziert der Satz von Beppo-Levi
$$\int (\liminf_{n\to\infty} X_n) d\mu = \int (\sup_{n\in\mathbb{N}} Y_n) d\mu = \sup_{n\in\mathbb{N}} \int Y_n d\mu = \sup_{n\in\mathbb{N}} \int \inf_{m\geq n} X_m d\mu \leq \sup_{n\in\mathbb{N}} \inf_{m\geq n} \int X_m d\mu = \liminf_{n\to\infty} \int X_n d\mu.$$
□

3.29 Satz. (Lebesgue, dominierte Konvergenz)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{L}^p(\mu)$ für ein $p \geq 1$ und $(X_n)_{n\in\mathbb{N}}$ konvergiere gegen X μ-fast sicher. Falls ein $Y \in \varepsilon^*$ mit $|X| \leq Y$ μ-fast sicher und $\int Y^p d\mu < \infty$ existiert, so gilt:
$$\|X_n - X\|_p \to 0 \quad (n \to \infty).$$
Im Fall $p = 1$ gilt zudem:
$$\lim_{n\to\infty} \int X_n d\mu = \int X d\mu.$$

Beweis. Da $X_n \to X$ μ-f.s., folgt, dass auch $|X|^p \leq Y^p$ μ-fast sicher. Nach Folgerung 3.22 ist $|X|^p$ μ-integrierbar. Sei
$$Y_n := |X_n - X|^p.$$
Aufgrund der Minkowski-Ungleichung ist Y_n μ-integrierbar, und es ist zu zeigen, dass $\int Y_n d\mu \to 0$ für $n \to \infty$.
Sei $Z := (|X| + Y)^p$. Dann ist ebenso Z μ-integrierbar und $Z - Y_n \geq 0$ μ-fast sicher. Aus dem Lemma von Fatou (unter Folgerung 3.21) folgt:
$$\int \liminf_{n\to\infty} (Z - Y_n)\, d\mu \leq \liminf_{n\to\infty} \int (Z - Y_n)\, d\mu = \int Z d\mu - \limsup_{n\to\infty} Y_n\, d\mu.$$
Da $X_n \to X$ μ-fast sicher, folgt $Z - Y_n \to Z$ μ-fast sicher. Also $\int \liminf_{n\to\infty} (Z - Y_n) d\mu = \int Z d\mu$. Also: $\limsup_{n\to\infty} \int Y_n d\mu \leq 0$. Da $\int Y_n d\mu \geq 0$, folgt $\lim_{n\to\infty} \int Y_n d\mu = 0$. Ist $p = 1$, so folgt aus dem bisher Gezeigten:
$$|\int X_n d\mu - \int X d\mu| \leq \int |X - X_n| d\mu \to 0 \quad (n \to \infty).$$
□

3.30 Satz. Sei $(X_n)_{n\in\mathbb{N}}$ ein Cauchy-Folge in $(\mathcal{L}^p(\mu), \|.\|_p)$. Dann existiert ein $X \in \mathcal{L}^p(\mu)$ mit $\|X_n - X\|_p \to 0$ $(n \to \infty)$. Die Räume $(L^p(\mu), \|.\|_p)$ sind also Banachräume.

Beweis. Da $(X_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist, existiert eine Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ mit $\|X_{n_{k+1}} - X_{n_k}\| \leq 2^{-k}$ für alle $k \in \mathbb{N}$. Sei $Y_k := X_{n_{k+1}} - X_{n_k}$ und $Y := \sum_{k=1}^{\infty} |Y_k|$. Aus der Minkowski-Ungleichung und dem Satz von Beppo-Levi folgt:
$$||Y||_p \leq \sum_{k=1}^{\infty} \|Y_k\|_p \leq \sum_{k=1}^{\infty} 2^{-k} = 1.$$

Da $|Y|^p$ μ-integrierbar ist, folgt $Y < \infty$ μ-f.s. (Folgerung 3.22). Also konvergiert die Reihe $\sum_{k=1}^{\infty} Y_k$ μ-fast sicher. Da $X_{n_k} = X_{n_1} + \sum_{l=1}^{k-1} Y_l$, konvergiert die Folge $(X_{n_k})_{k\in\mathbb{N}}$ μ-fast sicher gegen einen Grenzwert $X \in \mathcal{L}^p(\mu)$. Ferner folgt:

$$|X_{n_k}| \leq |X_{n_1}| + Y.$$

Nach der Minkowski-Ungleichung ist $\int (|X_{n_1}+Y)^p d\mu < \infty$. Der Satz von Lebesgue impliziert, dass $\|X_{n_k} - X\|_p \to 0 \quad (k \to \infty)$. Da $(X_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist, konvergiert die ganze Folge in $\mathcal{L}^p(\mu)$ gegen X. □

3.31 Folgerung. Falls $(X_n)_{n\in\mathbb{N}}$ in $\mathcal{L}^p(\mu)$ gegen X konvergiert, so existiert eine Teilfolge $(X_{n_k})_{k\in\mathbb{N}}$ mit $X_{n_k} \to X$ μ-fast sicher.

3.4 Integration bezüglich Produktmaßen

Es seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$, zwei Maßräume. Auf $\Omega_1 \times \Omega_2$ definieren wir die σ-Algebra

$$\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(A_1 \times A_2;\, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2),$$

die sogenannte *Produkt-σ-Algebra* von $\mathcal{A}_1$ und $\mathcal{A}_2$. Sei $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Dann betrachtet man die *Schnitte* (ω_i-Schnitte)

$$\begin{aligned} Q_{\omega_i} &:= \{\omega_2 \in \Omega_2 \,|\, (\omega_1, \omega_2) \in Q\} \qquad \omega_1 \in \Omega_1, \\ Q_{\omega_2} &:= \{\omega_1 \in \Omega_1 \,|\, (\omega_1, \omega_2) \in Q\} \qquad \omega_2 \in \Omega_2. \end{aligned}$$

3.32 Lemma. Für $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt: $Q_{\omega_1} \in \mathcal{A}_2$ und $Q_{\omega_2} \in \mathcal{A}_1$ für alle $\omega_1 \in \Omega_1, \omega_2 \in \Omega_2$.

Beweis. Sei $\mathcal{M} := \{Q \subset \Omega_1 \times \Omega_2 \,|\, Q_{\omega_1} \subset \mathcal{A}_2\}$. Dann ist $\mathcal{M}$ eine σ-Algebra auf $\Omega_1 \times \Omega_2$. Denn:

- $\Omega_1 \times \Omega_2 \in \mathcal{M}$, da $(\Omega_1, \Omega_2)_{\omega_1} = \Omega_2 \in \mathcal{A}_2$
- $Q \in \mathcal{A} \Rightarrow Q_{\omega_1} \in \mathcal{A}_2 \Rightarrow (\Omega_1 \times \Omega_2 \setminus Q)_{\omega_1} = \Omega_2 \setminus Q_{\omega_1} \in \mathcal{A}_2$
- $(Q_n)_{n\in\mathbb{N}}$ in $\mathcal{M} \Rightarrow \forall n \in \mathbb{N}\ (Q_n)_{\omega_1} \in \mathcal{A}_2 \Rightarrow (\bigcup_{n\in\mathbb{N}} Q_n)_{\omega_1} = \bigcup_{n\in\mathbb{N}} (Q_n)_{\omega_1} \in \mathcal{A}_2$

Zudem gilt: Die Mengen der Form $A_1 \times A_2$ mit $A_i \in \mathcal{A}_i$ liegen in $\mathcal{M}$. Denn:

$$(A_1 \times A_2) = \left\{\begin{matrix} A_1 & ,\omega_1 \in A_1 \\ \varnothing & ,\omega_1 \notin A_1 \end{matrix}\right\} \in \mathcal{A}_2$$

Also: $\mathcal{M} \supset \sigma(A_1 \times A_2; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2) = \mathcal{A}_1 \otimes \mathcal{A}_2$. □

Sei nun $X : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}\mathcal{A}_1 \otimes \mathcal{A}_2/\overline{\mathcal{B}}$-messbar. Wir definieren die ω_i-Schritte

$$X_{\omega_1} : \Omega_2 \to \overline{\mathbb{R}}, \omega_2 \mapsto X(\omega_1, \omega_2) \qquad \omega_1 \in \Omega_2,$$

$$X_{\omega_2} : \Omega_1 \to \overline{\mathbb{R}}, \omega_1 \mapsto X(\omega_1, \omega_2) \qquad \omega_2 \in \Omega_2.$$

3.33 Lemma. Ist X $\mathcal{A}_1 \otimes \mathcal{A}_2/\overline{\mathcal{B}}$-messbar, so sind X_{ω_1} und X_{ω_2} $\mathcal{A}_2/\overline{\mathcal{B}}$ bzw. $\mathcal{A}_1/\overline{\mathcal{B}}$-messbar.

Beweis. Für $B \in \overline{\mathcal{B}}$ gilt:

$$X_{\omega_1}^{-1}(B) = \{\omega_2 \in \Omega_2 \,|\, (\omega_1, \omega_2) \in X^{-1}(B)\} = (X^{-1}(B))_{\omega_1} \in \mathcal{A}_2$$

nach Lemma 3.32. □

Sei nun $X : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ $\mathcal{A}_1 \otimes \mathcal{A}_2$-messbar und nicht-negativ. Nach Lemma 3.33 existieren die Integrale

$$\int X_{\omega_1} d\mu_2 \text{ und } \int X_{\omega_2} d\mu_1$$

und sind Funktioenen in ω_1 bzw. ω_2.
In dieser Situation gilt:

3.34 Lemma. Die Abbildungen $\omega_1 \mapsto \int X_{\omega_1} d\mu_2$ und $\omega_2 \mapsto \int X_{\omega_2} d\mu_1$ sind $\mathcal{A}_1/\overline{\mathcal{B}}$-messbar, sofern μ_1 und μ_2 σ-endlich sind.

Beweis. (1) $X = \mathbb{1}_Q$, $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$, μ_1, μ_2 endlich: Sei

$$S_Q(\omega_1) := \int (\mathbb{1}_Q)_{\omega_1} d\mu_2 = \mu_2(Q_{\omega_1}).$$

Sei

$$\mathcal{D} = \{Q \in \mathcal{A}_1 \otimes \mathcal{A}_2 \,|\, \omega_1 \mapsto \mu_2(Q_{\omega_1}) \text{ ist } \mathcal{A}_1/\overline{\mathcal{B}} \text{ messbar}\}.$$

Dann ist $\mathcal{D}$ ein Dynkin-System, denn:

- $S_{\Omega_1 \times \Omega_2} = \mu_2(\Omega_2)$ ist konstant als Funktion in ω_1.
- $S_{\Omega_1 \times \Omega_2 \setminus Q} = S_{\Omega_1 \times \Omega_2} - S_Q$ ist messbar. Also: $Q^c \in \mathcal{D}$.
- Seien $(Q_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ paarweise disjunkt. Dann: $S_{\bigcup\limits_{n\in\mathbb{N}} Q_n} = \sum\limits_{n\in\mathbb{N}} S_{Q_n}$ ist messbar. Also: $\bigcup\limits_{n\in\mathbb{N}} Q_n \in \mathcal{D}$.

Ferner: $A_1 \times A_2 \in \mathcal{D}$ für $A_1 \in \mathcal{A}_1$ und $A_2 \in \mathcal{A}_2$, da $S_{A_1 \times A_2} = \mathbb{1}_{A_1} \mu_2(A_2)$. Aus dem Dynkin-Lemma (1.26) folgt:

$$\mathcal{D} = \mathcal{A}_1 \otimes \mathcal{A}_2.$$

(2) $X = \mathbb{1}_Q$, $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$: Sei nun $(B_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_2$ nichtfallend mit $\mu_2(B_n) < \infty$ und $\bigcup\limits_{n\in\mathbb{N}} B_n = \Omega$. Sei $\mu_{2,n}(A) := \mu_2(A \cap B_n)$. Nach (1) ist $\omega_1 \mapsto \mu_{2,n}(Q_{\omega_1})$ messbar für alle $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Aus der Stetigkeit der Maßes μ_2 von unten folgt

$$\mu_2(Q_{\omega_1}) = \sup_{n\in\mathbb{N}} \mu_{2,n}(Q_{\omega_1})$$

ist $\mathcal{A}_1/\overline{\mathcal{B}}$-messbar als abzählbares Supremum von messbaren Funktionen.

(3) X ist nicht-negative Treppenfunktion: (2) und der Linearität des Integrals.

(4) X nicht-negativ: Sei $(X^{(n)})_{n\in\mathbb{N}}$ eine nichtfallende Folge in $\varepsilon(\Omega_1 \times \Omega_2)$ mit

$$\sup_{n\in\mathbb{N}} X^{(n)} = X$$

Da $(\mathbb{1}_Q)_{\omega_1} = \mathbb{1}_{Q_{\omega_1}}$ ist $(X^{(n)}_{\omega_1})_{n\in\mathbb{N}}$ eine nicht-fallende Folge in $\varepsilon(\Omega_2)$ mit $\sup\limits_{n\in\mathbb{N}} X^{(n)}_{\omega_1} = X_{\omega_1}$. Also

$$\omega_1 \mapsto \int X_{\omega_1} \, d\mu_2 = \sup_{n\in\mathbb{N}} \int X^{(n)}_{\omega_1} \, d\mu_2$$

ist $\mathcal{A}_1/\overline{\mathcal{B}}$-messbar als abzählbares Supremum messbarer Funktionen unter Verwendung von Schritt 3. □

Sind μ_1 und μ_2 σ-endlich und X $\mathcal{A}_1 \otimes \mathcal{A}_2/\overline{\mathcal{B}}$-messbar und nicht-negativ, so können wir nach Lemma 3.34 die iterativen Integrale

$$\int \left(\int X_{\omega_1} \, d\mu_2 \right) \mu_1(d\omega_1) \text{ und } \int \left(\int X_{\omega_2} \, d\mu_1 \right) \mu_2(d\omega_2)$$

betrachten.
Frage: Sind beide iterierten Integrale gleich und kann man sie als Integral bezüglich eines geeigneten Maßes auf $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ darstellen?

3.35 Satz. Sind μ_1 und μ_2 σ-endlich, so existiert ein eindeutig bestimmtes Maß μ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \text{ für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \quad (*).$$

Es heißt *Produktmaß* von μ_1 und μ_2 und wird mit $\mu_1 \otimes \mu_2$ notiert. Für $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt:

$$\mu_1 \otimes \mu_2 \, (Q) = \int \mu_2(Q_{\omega_1}) \, \mu_1(d\omega_1) = \int \mu_1(Q_{\omega_2}) \, \mu_2(d\omega_2).$$

Beweis. Nach Lemma 3.34 ist

$$\pi(Q) := \int \mu_2(Q_{\omega_1}) \, \mu_1(d\omega_1)$$

für alle $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ definiert. Wir zeigen, dass π ein Maß ist:

- $\pi(Q) \geq 0$: klar.
- $\pi(\varnothing) = 0$, da $\varnothing_{\omega_1} = \varnothing$ und also $\mu_2(\varnothing_{\omega_1}) = 0$.

- σ-additiv: Seien $(Q_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_1 \otimes \mathcal{A}_2$ paarweise disjunkt. Dann folgt:

$$\mu_2((\bigcup_{n\in\mathbb{N}} Q_n)_{\omega_1}) = \mu_2(\bigcup_{n\in\mathbb{N}} (Q_n)_{\omega_1}) \underset{\text{von } \mu_2}{\overset{\sigma-\text{Add.}}{=}} \sum_{n\in\mathbb{N}} \mu_2((Q_n)_{\omega_1}).$$

Mit dem Satz von Beppo-Levi folgt:

$$\pi(\bigcup_{n\in\mathbb{N}} Q_n) = \sum_{n\in\mathbb{N}} \int \mu_2((Q_n)_{\omega_1})\mu_1(d\omega_1) = \sum_{n\in\mathbb{N}} \pi(Q_n).$$

Ebenso folgt, dass $\overline{\pi}$ definiert durch

$$\overline{\pi}(Q) = \int \mu_1(Q_{\omega_2})\, \mu_2(d\omega_2)$$

ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ ist. Beide Maße erfüllen $(*)$, da etwa $\mu_2((A_1 \times A_2)_{\omega_1}) = \mathbb{1}_{A_1}\mu_2(A_2)$. Es bleibt zu zeigen, dass $(*)$ ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ eindeutig festlegt. Seien π_1 und π_2 beliebige Maße, die $(*)$ erfüllen. π_1 und π_2 erben dann die σ-Endlichkeit von μ_1 und μ_2. Also folgt die Eideutigkeit aus Satz 1.21, (vi). □

3.36 Satz. (Fubini)
Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, 2$, σ-endliche Maßräume und $X : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ $\mathcal{A}_1 \otimes \mathcal{A}_2/\overline{\mathcal{B}}$-messbar. Dann:

(i) Ist X nicht-negativ, so

$$\int_{\Omega_1\times\Omega_2} X\, d(\mu_1 \otimes \mu_2) = \int \left(\int X_{\omega_1}\, d\mu_2\right) \mu_1(d\omega_1) = \int \left(\int X_{\omega_2}\, d\mu_1\right) \mu_2(d\omega_2). \qquad (F)$$

(ii) Ist X $\mu_1 \otimes \mu_2$-integrierbar, so ist X_{ω_1} für μ_1-fast alle ω_1 μ_2-integrierbar und X_{ω_2} ist für μ_2-fast alle ω_2 μ_1-integrierbar. Die μ_1- (bzw. μ_2-) fast überall definierten Funktionen

$$\omega_1 \mapsto \int X_{\omega_1} d\mu_2 \text{ und } \omega_2 \mapsto \int X_{\omega_2} d\mu_1$$

sind μ_1- (bzw. μ_2-) integrierbar und (F) gilt.

Beweis. (i) Für nicht-negative Treppenfunktionen Satz 3.35 und Linearität der Integrale. Sei X nicht-negativ und $(X^{(n)})_{n\in\mathbb{N}}$ eine nichtfallende Folge in $\varepsilon(\Omega_1 \times \Omega_2)$ mit $X = \sup_{n\in\mathbb{N}} X^{(n)}$. Dann ist $(X^{(n)}_{\omega_2})_{n\in\mathbb{N}}$ „approximierend“ für X_{ω_2} und der Satz von Beppo-Levi liefert:

$$\begin{aligned}\int\int X_{\omega_2} d\mu_1 \mu_2(d\omega_2) &= \int \left(\sup_{n\in\mathbb{N}} \int X^{(n)}_{\omega_2} d\mu_1\right) \mu_2(d\omega_2)\\ &= \sup_{n\in\mathbb{N}} \int\int X^{(n)}_{\omega_2} d\mu_1 \mu_2(d\omega_2)\\ &= \sup_{n\in\mathbb{N}} \int X^{(n)} d(\mu_1 \otimes \mu_2) = \int X d(\mu_1 \otimes \mu_2).\end{aligned}$$

(ii) Teil (i) liefert

$$\int\int |X_{\omega_1}| d\mu_2 \mu_1(d\omega_1) = \int\int |X_{\omega_2}| d\mu_1 \mu_2(d\omega_2) = \int |X| d\mu_1 \otimes \mu_2 < \infty \qquad (*).$$

Also impliziert Folgerung 3.22:

$$\int |X_{\omega_1}| d\mu_2 < \infty \text{ für } \mu_1\text{-fast alle } \omega_1 \in \Omega_1 \text{ und } \int |X_{\omega_2}| d\mu_1 < \infty \text{ für } \mu_2\text{-fast alle } \omega_2 \in \Omega_2.$$

Also ist X_{ω_1} für μ_1-fast alle $\omega_1 \in \Omega_1$ μ_2-integrierbar und

$$\omega_1 \mapsto \int X_{\omega_1} d\mu_2 = \int (X^+)_{\omega_1} d\mu_2 - \int (X^-)_{\omega_1} d\mu_2$$

ist μ_1-fast überall definiert und nach Lemma 3.34 $\mathcal{A}_1/\overline{\mathcal{B}}$ messbar. Aus $(*)$ und Satz 3.10 folgt, dass diese Funktion μ_1-integrierbar ist und Anwendung von Teil (i) auf Positiv- und Negativteil:

$$\begin{aligned}\int\int X_{\omega_1} d\mu_2 \mu_1(d\omega_1) &= \int \left(\int (X^+)_{\omega_1} d\mu_2 \mu_1(d\omega_1) - \int (X^-)_{\omega_1} d\mu_2 \mu_1(d\omega_1)\right)\\ &= \int X^+ d\mu_1 \otimes \mu_2 - \int X^- d\mu_1 \otimes \mu_2\\ &= \int X d\mu_1 \otimes \mu_2.\end{aligned}$$

□

3.37 Bemerkung. Kurzschreibweise für Doppelintegrale: $\int \left(\int X d\mu_1\right) d\mu_2 := \int \int X_{\omega_2} d\mu_1 \mu_2(d\omega_2)$.

3.38 Bemerkung. Seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, ..., n$ σ-endliche Maßräume. Induktiv sieht man, dass ein eindeutig bestimmtes Maß

$$\mu : \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes ... \otimes \mathcal{A}_n \to [0, \infty] \text{ mit } \mu(A_1 \times A_2 \times ... \times A_n) = \prod_{i=1}^{n} \mu_i(A_i) \text{ für alle } A_1 \in \mathcal{A}_1, ..., A_n \in \mathcal{A}_n$$

existiert. Es heißt Produktmaß von $\mu_1, ..., \mu_n$ und wird mit $\mu_1 \otimes ... \otimes \mu_n$ notiert. Der Maßraum $\left(\sum_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{A}_i, \bigotimes_{i=1}^{n} \mu_i\right)$ ist der *Produktraum*von $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i = 1, ..., n$. Der Satz von Fubini gilt dann in obiger Form für n-fach Integrale.

3.5 Varianz, Kovarianz und Faltung

In diesem Abschnitt sein ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ gegeben.

3.39 Definition. Seien X und Y numerische Zufallsvariablen.

(i) Ist X $\mathcal{P}$-quasi-integrierbar, so wird

$$E[X] := \int X d\mathcal{P}$$

der *Erwartungswert von X unter $\mathcal{P}$* genannt.

(ii) Ist X $\mathcal{P}$-integrierbar, dann heißt

$$Var(X) := E[(X - E[X])^2]$$

die *Varianz von X*. $\sqrt{Var(X)}$ ist die *Standardabweichung von X*.

(iii) SInd X, Y und $X \cdot Y$ $\mathcal{P}$-integrierbar, so heißt

$$E[(X - E[X])(Y - E[Y])] =: Cov(X, Y)$$

sie *Kovarianz von X und Y*. Ist $Cov(X, Y) = 0$, so nennt man X und Y *unkorreliert.*

(iv) Falls $(X, Y) \in L^2(\Omega, \mathcal{A}, \mathcal{P})$ und $Var(X) \neq 0 \neq Var(Y)$, so wird

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}}$$

die *Korrelation von X und Y* genannt.

(v) Ist $X \in L^n(\Omega, \mathcal{A}, \mathcal{P})$ für ein $n \in \mathbb{N}$, so heißt $E[X^n]$ das *n-te Moment von X* und $E[|X|^n]$ das *n-te Absolutmoment von X*.

3.40 Satz. Seien X und Y reellwertige Zufallsvariablen.

(i) $X \in L^2(\mathcal{P}) \Leftrightarrow X$ integrierbar und $Var(X) < \infty$
In diesem Fall gilt:

  (a) $Var(X) = E[X^2] - E[X]^2$

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

(ii) Sind $X, Y \in L^2(\mathcal{P})$ mit $Var(X) \neq 0 \neq Var(Y)$, so ist $|Corr(X, Y)| \leq 1$ und $|Corr(X, Y)1 = 1 \Leftrightarrow \exists (a, b) \in \mathbb{R}$ mit $a \neq 0 : Y = aX + b$ $\mathcal{P}$-fast sicher.

Beweis. (i) Ist $X \in L^2(\mathcal{P})$, so ist $X \in L^1(\mathcal{P})$ nach Folgerung 3.24. Also ist X $\mathcal{P}$-integrierbar und die Konstante $E[X] \in L^2(\mathcal{P})$, da $\mathcal{P}$ endlich ist. Aufgrund der Minkowski-Ungleichung ist dann $X - E[X] \in L^2(\mathcal{P})$, also $Var(X) < \infty$.
Ist X integrierbar und $Var(X) < \infty$, so sind $E[X] \in L^2(\mathcal{P})$ und $X - E[X] \in L^2(\mathcal{P})$. Also: $X = (X - E[X]) + E[X] \in L^2(\mathcal{P})$.

  (a) Folgt aus der Linearität des Erwartungswertes.

  (b) $Var(aX + b) = E[((aX + b) - E[aX + b])^2] = E[a^2(X - E[X])^2] = a^2 Var(X)$

(ii) Seien $\xi := \frac{X - E[X]}{\sqrt{Var(X)}}$ und $\eta := \frac{Y - E[Y]}{\sqrt{Var(Y)}}$.
Dann: $E[\xi] = E[\eta] = 0$ und $Var(\xi) = Var(\eta) = 1$.
Somit: $Corr(X, Y) = E[\xi\eta] = E[\frac{1}{2}(\xi^2 + \eta^2) - \frac{1}{2}(\xi - \eta)^2] = 1 - \frac{1}{2}E[(\xi - \eta)^2]$. Da $0 \leq E[(\xi - \eta)^2] \leq 2E[\xi^2 + \eta^2] = 2$, folgt:

- $Corr(X,Y) \in [-1,1]$
- $Corr(X,Y) = 1 \Leftrightarrow \xi = \eta$ $\mathcal{P}$-fast sicher

In diesem Fall gilt: $Y = \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}X - \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}E[X] + E[Y]$.

- $Corr(X,Y) = -1 \Leftrightarrow 2(\xi^2 + \eta^2) - (\xi - \eta)^2 = 0$ $\mathcal{P}$-fast sicher $\Leftrightarrow \xi = \eta$ $\mathcal{P}$-fast sicher

In diesem Fall gilt: $Y = -\frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}X + \frac{\sqrt{Var(Y)}}{\sqrt{Var(X)}}E[X] + E[Y]$.
Ist $Y = aX + b$ mit $a \neq 0$, so ist $\eta = \frac{a}{|a|}$, da $E[Y] = aE[X] + b$ und $\sqrt{Var(Y)} = |a|\sqrt{Var(X)}$.
Also: $Corr(X,Y) = E[\xi\eta] = \frac{a}{|a|}$. □

3.41 Satz. Seien $X_1, ..., X_n$ unabhängige reellwertige Zufallsvariablen .

(i) Sind $X_1, ..., X_n$ nicht-negativ, so gilt:

$$E\left[\prod_{i=1}^n X_i\right] = \prod_{i=1}^n E[X_i] \qquad (*).$$

(ii) Sind $X_1, ..., X_n$ integrierbar, so ist auch $\prod_{i=1}^n X_i$ integrierbar und $(*)$ gilt. Insbesondere sind $X_1, ..., X_n$ dann paarweise unkorreliert, d.h. $Corr(X_i, X_j) = 0 \,\forall\, i \neq j$.

Beweis. Sei $\mathcal{P}_{\mathrm{X}}$ die Verteilung der n-dimensionalen Zufallsvariable $\mathrm{X} = \begin{pmatrix} \mathrm{X}_1 \\ : \\ \mathrm{X}_n \end{pmatrix}$ und $\mathcal{P}_{X_i}$ die Verteilung von X_i, $i = 1, ..., n$. Unabhängigkeit bedeutet

$$\mathcal{P}_{\mathrm{X}} = \mathcal{P}_{\mathrm{X}_1} \otimes ... \otimes \mathcal{P}_{\mathrm{X}_n}.$$

(i) Aus dem Transformationssatz (Schritt 2 in Beweis) und dem Satz von Fubini folgt

$$\begin{aligned} E\left[\prod_{i=1}^n X_i\right] &= \int x_1 \cdots x_n \mathcal{P}_{\mathrm{X}}(d(x_1, ..., x_n)) \\ &= \int \cdots \int x_1 \cdots x_n \, \mathcal{P}_{\mathrm{X}_1}(dx_1) \mathcal{P}_{\mathrm{X}_n}(dx_n) = \prod_{i=1}^n \int x_i \, d\mathcal{P}_{\mathrm{X}_i}(dx_i) = \prod_{i=1}^n E[X_i] \end{aligned}$$

(ii) Mit $X_1, ..., X_n$ sind auch $|X_1|, ..., |X_n|$ unabhängig (Übung). Also folgt aus (i):

$$E\left[\left|\prod_{i=1}^n X_i\right|\right] = E\left[\prod_{i=1}^n |X_i|\right] = \prod_{i=1}^n E[|X_i|] < \infty.$$

Also ist $\prod_{i=1}^n X_i$ integrierbar. Die gleiche Argumentation wie in Teil (i) zeigt nun, dass $(*)$ gilt.
Im Spezialfall erhält man für $i \neq j$ $E[X_i X_j] - E[X_i]E[X_j] = 0$, also die paarweise Unkorreliertheit. □

3.42 Bemerkung. Die Umkehrung von (ii) des vorherigen Satzes gilt im Allgemeinen nicht. Aus Unkorreliertheit kann man nicht auf Unabhängigkeit schließen (Übung).

3.43 Satz. (Bienaymé) Sind $X_1, ..., X_n$ paarweise unkorreliert, so gilt:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i).$$

Beweis.

$$\begin{aligned} Var\left(\sum_{i=1}^n X_i\right) &= E\left[\left(\sum_{i=1}^n X_i - E\left[\sum_{i=1}^n X_i\right]\right)^2\right] = E\left[\left(\sum_{i=1}^n X_i - E[X_i]\right)^2\right] \\ &= \sum_{i=1}^n \underbrace{E[(X_i - E[X_i])^2]}_{Var(X_i)} + 2\sum_{i=1}^n \sum_{j=1}^{i-1} \underbrace{E[(X_i - E[X_i])(X_j - E[X_j])]}_{Cov(X_i, X_j) = 0} \end{aligned}$$

□

3.44 Definition. Seien X und Y unabhängige $\mathbb{R}^D$-wertige Zufallsvariablen mit Verteilungen $\mathcal{P}_X$ und $\mathcal{P}_Y$. Dann wird die Verteilung von $X + Y$ die *Faltung von* $\mathcal{P}_X$ *und* $\mathcal{P}_Y$ genannt und wird mit $\mathcal{P}_X * \mathcal{P}_Y$ notiert.

3.45 Satz. (Faltungsformel) Seien X und Y unabhängige $\mathbb{R}^D$-wertige Zufallsvariablen und $A \in \mathcal{B}_D$. Dann:

$$(\mathcal{P}_X * \mathcal{P}_Y)(A) = \int \mathcal{P}_X(-y+A)\mathcal{P}_Y(dy) = \int \mathcal{P}_Y(-x+A)\mathcal{P}_X(dx).$$

Haben $\mathcal{P}_X$ und $\mathcal{P}_Y$ Dichten f_X und f_Y bezüglich $\lambda^{\otimes D}$, so ist

$$(f_X * f_Y)(z) = \int f_X(z-y)f_Y(y)\lambda^{\otimes D}(dy) = \int f_Y(z-x)f_X(x)\lambda^{\otimes D}(dx)$$

Dichte von $\mathcal{P}_X * \mathcal{P}_Y$ bezüglich $\lambda^{\otimes D}$.

Beweis. Sei $\mathcal{P}_{(X,Y)}$ die Verteilung von $\begin{pmatrix} X \\ Y \end{pmatrix}$. Aufgrund der Unabhängigkeit von X und Y ist $\mathcal{P}_{(X,Y)} = \mathcal{P}_X \otimes \mathcal{P}_Y$. Also folgt aus dem Transformationssatz und dem Satz von Fubini:

$$(\mathcal{P}_X * \mathcal{P}_Y)(A) = \int \mathbb{1}_{\{X+Y\in A\}}\mathcal{P}_{(X,Y)}d(x,y) = \int\int \mathbb{1}_{\{X+Y\in A\}}\mathcal{P}_Y(dy)\mathcal{P}_X(dx) = \int \mathcal{P}_Y(-x+A)\mathcal{P}_X(dx).$$

Haben $\mathcal{P}_X$ und $\mathcal{P}_Y$ Dichten f_X und unf f_Y bezüglich $\lambda^{\otimes D}$, so gilt:

$$\begin{aligned}(\mathcal{P}_X * \mathcal{P}_Y)(A) &= \int \int\limits_{-x+A} f_Y(y)\lambda^{\otimes D}(dy)\, f_X(x)\lambda^{\otimes D}(dx)\\ &= \int\int\limits_A f_Y(y-x)\lambda^{\otimes D}(dy)\, f_X(x)\lambda^{\otimes D}(dx) \overset{\text{Fubini}}{=} \underbrace{\int\limits_A\int\limits_{\mathbb{R}^D} f_Y(y-x)\, f_X(x)\lambda^{\otimes D}(dx)\, \lambda^{\otimes D}(dy)}_{*}.\end{aligned}$$

Also ist $(*)$ als Funktion in Y Dichte von $\mathcal{P}_X * \mathcal{P}_Y$ bezüglich $\lambda^{\otimes D}$. □

3.46 Beispiel. Sind $X_1, ..., X_n$ unabhängig und Bernoulli(n)-verteilt, so ist $\sum_{i=1}^{n} X_i$ binomial(n, p)-verteilt.

Kapitel 4

Folgen von Zufallsvariablen

4.1 Konvergenzarten von Zufallsvariablen

4.1 Definition. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge reellwertiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ und X eine reellwertige Zufallsvariable.

(i) $(X_n)_{n\in\mathbb{N}}$ *konvergiert gegen X in Wahrscheinlichkeit*, falls

$$\forall \varepsilon > 0 : \quad \mathcal{P}(\{|X_n - X| > \varepsilon\}) \to 0 \quad n \to \infty.$$

(ii) $(X_n)_{n\in\mathbb{N}}$ *konvergiert gegen X $\mathcal{P}$-fast sicher*, falls

$$\mathcal{P}(\{\omega \in \Omega \,|\, X_n(\omega) \to X(\omega)\}) = 1.$$

(iii) $(X_n)_{n\in\mathbb{N}}$ *konvergiert gegen X im p-ten Mittel* (in L^p), falls

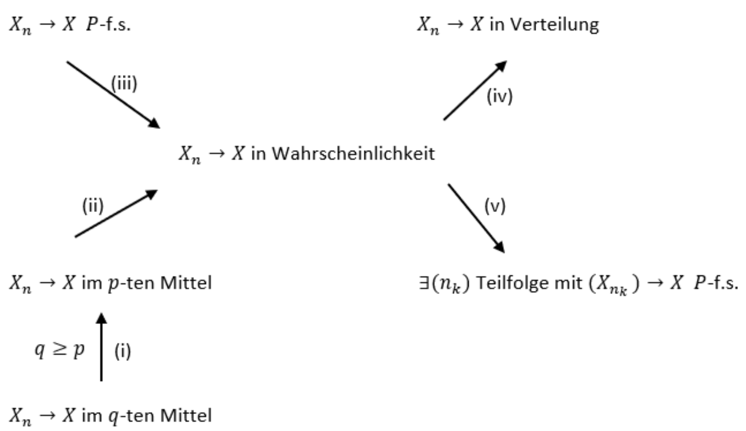
$$E[|X_n - X|^p] \to 0.$$

4.2 Definition. Seien $(X_n)_{n\in\mathbb{N}}$ reellwertige Zufallsvariablen auf dem Wahrscheinlichkeitsraum $(\Omega_n, \mathcal{A}_n, \mathcal{P}_n)_n$ und X eine $\mathbb{R}$-wertige Zufallsvariable auf $(\Omega, \mathcal{A}, \mathcal{P})$. Man sagt, $(X_n)_{n\in\mathbb{N}}$ *konvergiert gegen X in Verteilung*, falls für alle stetigen und beschränkten Fumktionen $f : \mathbb{R} \to \mathbb{R}$ gilt:

$$E^{\mathcal{P}_n}[f(X_n)] \to E^{\mathcal{P}}[f(X)]$$

Der folgende Zusammenhang besteht zwischen den Konvergenzsätzen:

4.3 Satz. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge $\mathbb{R}$-wertiger Zufallsvariablen und X $\mathbb{R}$-wertige Zufallsvariable auf $(\Omega, \mathcal{A}, \mathcal{P})$. Dann gilt:



Zum Vorbereiten benötigen wir:

4.4 Lemma. (Borel-Cantelli)
Sei $(\Omega, \mathcal{A}, \mathcal{P})$ ein Wahrscheinlichkeitsraum und $(A_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ eine Folge von Ereignissen. Dann:

(i) $\sum\limits_{n=1}^{\infty} \mathcal{P}(A_n) < \infty \Rightarrow \mathcal{P}\left(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k\right) = 0$

(ii) $\sum\limits_{n=1}^{\infty}$ und $(A_n)_{n\in\mathbb{N}}$ unabhängig $\Rightarrow \mathcal{P}\left(\bigcap\limits_{n=1}^{\infty} \bigcup\limits_{k=n}^{\infty} A_k\right) = 1$

Beweis. (i) Die Stetigkeit des Wahrscheinlichkeitsmaßes von oben impliziert

$$\mathcal{P}\left(\bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} \mathcal{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathcal{P}(A_k) = 0.$$

(ii) Mit $(A_i)_i$ ist auch $(A_i^c)_i$ unabhängig. Insbesondere gilt für alle $k \geq 0$ und $n \in \mathbb{N}$: $\mathcal{P}\left(\bigcap\limits_{m=n}^{n+k} A_m^c\right) = \prod\limits_{m=n}^{n+k} \mathcal{P}(A_m^c)$. Damit:

$$\begin{aligned}\mathcal{P}\left(\bigcap_{n\in\mathbb{N}} \bigcup_{m=n}^{\infty} A_m\right) &= \lim_{n\to\infty} \mathcal{P}\left(\bigcup_{m=n}^{\infty} A_m\right) = \lim_{n\to\infty} \lim_{k\to\infty} \mathcal{P}\left(\bigcup_{m=n}^{n+k} A_m\right) \\ &= \lim_{n\to\infty} \lim_{k\to\infty} \left(1 - \mathcal{P}\left(\bigcup_{m=n}^{n+k} A_m^c\right)\right) = \lim_{n\to\infty} \lim_{k\to\infty} \left(1 - \prod_{m=n}^{n+k} \mathcal{P}(A_m^c)\right)\end{aligned}$$

Falls $\mathcal{P}(A_m) = 1$ für unendlich viele $m \in \mathbb{N}$, so ist die Behauptung klar. Andernfalls gilt für hinreichend großes n und $k \geq 0$ wegen $\ln(1-x) \leq 0 - x$:

$$\log\left(\prod_{m=n}^{n+k} \mathcal{P}(A_m)^c\right) = \log\left(\prod_{m=n}^{n+k} (1 - \mathcal{P}(A_m))\right) = \sum_{m=n}^{n+k} \log(1 - \mathcal{P}(A_m)) \leq -\sum_{m=n}^{n+k} \mathcal{P}(A_m) \to -\infty \quad (k \to \infty)$$

Also: $\forall n \in \mathbb{N} : \lim\limits_{k\to\infty} \prod\limits_{m=n}^{n+k} \mathcal{P}(A_m^c) = 0$. Insbesondere folgt die Behauptung. □

4.5 Satz. (i) *(Markov-Ungleichung)* $\forall p \in [1, \infty[, \varepsilon > 0$ gilt:

$$\mathcal{P}(\{|X| \geq \varepsilon\}) \leq \frac{E[|X|^p]}{\varepsilon^p}$$

(ii) *(Chebyshev-Ungleichung)* Ist X $\mathcal{P}$-integrierbar und $\varepsilon > 0$, so gilt:

$$\mathcal{P}(\{|X - E[X]| \geq \varepsilon\}) \leq \frac{Var(X)}{\varepsilon^2}$$

Beweis. (i) $\mathcal{P}(\{|X| \geq \varepsilon\}) = E[\mathbb{1}_{\{|X|\geq\varepsilon\}}] \leq E\left[\frac{|X|^p}{\varepsilon^p}\right] = \frac{E[|X|^p]}{\varepsilon^p}$

(ii) Folgt aus (i) mit $p = 2$ und $X - E[X]$ statt X. □

Beweis. (von Satz 4.3)

(i) $X_n \to X$ in L^q, $q \geq p$. Mit der Hölder-Ungleichung folgt: $E[|X_n - X|^p]^{\frac{1}{p}} \leq E[|X_n - X|^q]^{\frac{1}{q}} \overset{n\to\infty}{\longrightarrow} 0$

(ii) $X_n \to X$ in L^p. Sei $\varepsilon > 0$. Dann: $\mathcal{P}(\{|X_n - X| \geq \varepsilon\}) \overset{Markov}{\leq} \frac{E[|X_n - X|^p]}{\varepsilon^p} \to 0$

(iii) $X_n \to X$ $\mathcal{P}$-fast sicher. Dazu sei

$$\{\omega \,|\, X_n(\omega) \to X(\omega)\} = \left\{\omega \,|\, \forall k \in \mathbb{N} \; \exists n_0 : |X_n(\omega) - X(\omega)| \leq \frac{1}{k} \forall n \geq n_0\right\} = \bigcap_{k\in\mathbb{N}} \bigcup_{n_0\in\mathbb{N}} \bigcap_{n\geq n_0}^{\infty} \left\{\omega \,|\, |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}.$$

Also:

$$\begin{aligned} X_n \to X \; \mathcal{P}\text{-f.s.} &\Leftrightarrow \forall k \in \mathbb{N} : \mathcal{P}\left(\bigcup_{n_0\in\mathbb{N}} \bigcap_{n\geq n_0} \left\{\omega \,|\, |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}\right) = 1 \\ &\Leftrightarrow \forall k \in \mathbb{N} : \mathcal{P}\left(\bigcup_{n_0=1}^{\infty} \bigcap_{n=n_0}^{\infty} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) = 1 \\ &\Leftrightarrow \forall k \in \mathbb{N} : \lim_{n_0\to\infty} \mathcal{P}\left(\bigcap_{n=n_0}^{\infty} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) = 1 \\ &\Leftrightarrow \forall k \in \mathbb{N} : \lim_{n_0\to\infty} \mathcal{P}\left(\left\{\sup_{n\geq n_0} |X_n - X| \leq \frac{1}{k}\right\}\right) = 1 \\ &\Leftrightarrow \forall \varepsilon > 0 : \lim_{n_0\to\infty} \mathcal{P}(\{\sup_{n\geq n_0} |X_n - X| \leq \varepsilon\}) = 1 \end{aligned}$$

Also:

$$
\begin{aligned}
X_n \to X \ \mathcal{P}\text{-f.s.} \ &\Rightarrow \forall \varepsilon > 0: \lim_{N\to\infty} \mathcal{P}(\{|X_n - X| \le \varepsilon\}) = 1 \\
&\Rightarrow \forall \varepsilon > 0: \lim_{N\to\infty} \mathcal{P}(\{|X_n - X| > \varepsilon\}) = 0 \\
&\Rightarrow X_n \to X \text{ in Wahrscheinlichkeit}
\end{aligned}
$$

(iv) $X_n \to X$ in Wahrscheinlichkeit. Zu zeigen: $X_n \to X$ in Verteilung.
Sei $f : \mathbb{R} \to \mathbb{R}$ stetig und beschränkt mit Schranke C. Wähle $N \in \mathbb{N}$ groß genug, so dass für vorgegeben $\varepsilon > 0$: $\mathcal{P}(\{|X| > N\}) \le \frac{\varepsilon}{8C}$. Sei $\delta > 0$ so gewählt, dass $|f(x) - f(y)| \le \frac{\varepsilon}{4}$ für alle $x, y \in [-N, N]$ mit $|x - y| \le \delta$. Dann:

$$
\begin{aligned}
E[|f(X_n) - f(X)|] &= E[|f(X_n) - f(X)|\mathbb{1}_{\{|X_n - X| \le \delta \text{ und } |X| \le N\}}] + E[|f(X_n) - f(X)|\mathbb{1}_{\{|X_n - X| \le \delta \text{ und } |X| > N\}}] \\
&\quad + E[|f(X_n) - f(X)|\mathbb{1}_{\{|X_n - X| > \delta\}}] \\
&\le \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + 2C \underbrace{\mathcal{P}(\{|X_n - X| > \delta\})}_{\to 0 \ (n\to\infty)}
\end{aligned}
$$

Also $E[|f(X_n) - f(X)|] \le \varepsilon$ für hinreichend großes $n \ge N$.

(v) $X_n \to X$ in Wahrscheinlichkeit. Zu zeigen: $\exists$ Teilfolge (n_k) mit $X_{n_k} \to X$ $\mathcal{P}$-fast sicher.
Setze $n_1 := 1$. Gegeben n_{k-1} sei $n_k := \inf\{n \ge n_{k-1} + 1\}$. Sei $\mathcal{P}(\{|X_\nu - X_\mu| > 2^{-k}\}) \le 2^{-k}$ für alle $\nu, \mu \ge n$.
Da $\mathcal{P}(\{|X_\nu - X_\mu| > 2^{-k}\}) \le \mathcal{P}(\{|X_\nu - X| > 2^{-(k+1)}\}) + \mathcal{P}(\{|X_\mu - X| > 2^{-(k+1)}\})$ und $X_n \to X$ in Wahrscheinlihckeit, folgt, dass $n_k < \infty$. Dann: $\sum\limits_{k=1}^{\infty} \mathcal{P}(\{|X_{n_{k+1}} - X_{n_k} <^{-k}\}) \le \sum\limits_{k=1}^{\infty} 2^{-k} < \infty$.
Das Lemma von Borel-Cantelli impliziert $\mathcal{P}(\{|X_{n_{k+1}} - X_{n_k}| > 2^{-k}$ für unendlich viele $k \in \mathbb{N}\}) = 0$.
Sei $\mathcal{N} := \left\{ \sum\limits_{k=1}^{\infty} |X_{n_{k+1}} - X_{n_k}| = +\infty \right\}$. Dann $\mathcal{P}(\mathcal{N}) = 0$.Sei

$$
\tilde{X}(\omega) \begin{cases} X_{n_1}(\omega) + \sum\limits_{k=1}^{\infty} X_{n_{k+1}}(\omega) - X_{n_k}(\omega) & , w \notin \mathcal{N} \\ 0 & , \omega \in \mathcal{N}. \end{cases}
$$

Dann gilt $X_{n_k} \to \tilde{X}$ $\mathcal{P}$-fast sicher. Also auch $X_{n_k} \to \tilde{X}$ in Wahrscheinlichkeit nach Teil (ii). Da der Grenzwert bei Konvergenz in Wahrscheinlichkeit bis auf Modifikation eindeutig ist, folgt $X = \tilde{X}$ $\mathcal{P}$-fast sicher und also auch $X_{n_k} \to X$ $\mathcal{P}$-fast sicher. □

In Beweis von Teil (v) wurde gezeigt:

4.6 Folgerung. (Cauchy-Kriterium für Konvergenz in Wahrscheinlichkeit)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von reellen Zufallsvariablen. Dann sind äquivalent:

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert in Wahrscheinlichkeit (gegen eine reelle Zufallsvariable)

(ii) $(X_n)_{n\in\mathbb{N}}$ ist eine Cauchy-Folge in Wahrscheinlichkeit, d.h.

$$
\forall \varepsilon > 0, \delta > 0 \, \exists N \in \mathbb{N} \, \forall n, m \ge N : \ \mathcal{P}(\{|X_n - X_m| > \varepsilon\}) \le \delta.
$$

4.7 Satz. (Cauchy-Kriterium)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von reellen Zufallsvariablen. Dann sind äquivalent:

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert $\mathcal{P}$-fast sicher (gegen eine reelle Zufallsvariable)

(ii) $(X_n)_{n\in\mathbb{N}}$ ist eine Cauchy-Folge bezüglich fast sicherer Konvergenz, d.h.

$$
\mathcal{P}(\{\omega \in \Omega \,|\, (X_n(\omega))_{n\in\mathbb{N}} \text{ ist Cauchy-Folge}\}) = 1.
$$

(iii) Für jedes $\varepsilon > 0$ gilt:

$$
\lim_{n\to\infty} \mathcal{P}(\{\sup_{k\ge 1} \,|\, |X_{n+k} - X_n| \le \varepsilon\}) = 1.
$$

Beweis. **(i)⇔(ii):** Klar (Analysis I und Messbarkeit des Grenzwerts).
(ii)⇔(iii): Analog zum Beweis von Satz 4.3(ii). □

4.8 Beispiel. (i) Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Bernoulli(p_n)-verteilten Zufallsvariablen. Dann:

- $X_n \to 0$ in Wahrscheinlichkeit $\Leftrightarrow$ $p_n \to 0$

Denn für alle $0 < \varepsilon < 1$ gilt:

$$\mathcal{P}(\{|X_n - 0| > \varepsilon\}) = \mathcal{P}(\{X_n = 1\}) = p_n.$$

- $X_n \to 0$ $\mathcal{P}$-fast sicher $\Leftrightarrow \sum\limits_{n=1}^{\infty} p_n < \infty$.

Denn nach dem Lemma von Borel-Cantelli gilt:

$$\mathcal{P}(\{X_n = 1 \text{ für unendlich viele } n \in \mathbb{N}\}) = \begin{cases} 1 & , \sum\limits_{n=1}^{\infty} p_n = +\infty \\ 0 & , \text{sonst} \end{cases}$$

Insbesondere folgt, dass für $p_n = \frac{1}{n}$ gilt: $X_n \to 0$ in Wahrscheinlichkeit, aber $(X_n)_{n\in\mathbb{N}}$ konvergiert nicht $\mathcal{P}$-fast sicher.

(ii) Sei $(\Omega, \mathcal{A}, \mathcal{P}) = (\mathbb{R}, \mathcal{B}, \alpha_{[0,1]})$, wobei $\alpha_{[0,1]}$ die Gleichverteilung auf $[0, 1]$ ist. Sei

$$X_n(\omega) = \begin{cases} e^n & , \omega \leq \frac{1}{n} \\ 0 & , \omega > \frac{1}{n} \end{cases}$$

Dann gilt: $X_n(\omega) \to 0$ für alle $\omega > 0$, also $\mathcal{P}$-fast sicher und somit auch in Wahrscheinlichkeit. Die Folge konvergiert nicht in $L^p(\Omega, \mathcal{A}, \mathcal{P})$ für $p \geq !$, da dann nur $X = 0$ als Grenzwert in Frage käme, aber

$$E[|X_n - 0|^p] = \frac{e^{np}}{n} \to +\infty \qquad (n \to \infty).$$

4.2 Gesetze der großen Zahlen (GGZ)

Empirische Beobachtung: Wenn ein Experiment sehr häufig unter identischen Bedingungen unabhängig wiederholt wird, so liegt das arithmetische Mittel $\frac{1}{n} \sum\limits_{i=1}^{n} x_i$der Beobachtungen typischerweise in der Nähe des Erwartungswertes des Einzelexperiments.
Denken wir die Beobachtungen x_i als Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen $(X_i)_{i=1,\ldots,n}$, so sollte also in einer geeigneten Form gelten: $\frac{1}{n} \sum\limits_{i=1}^{n} X_i \to E[X_1]$.
Hierbei heißt eine Familie von Zufallsvariablen $(X_i)_{i\in I}$ *identisch verteilt*, falls alle X_i dieselbe Verteilung haben, d.h. $\mathcal{P}_{X_i} = \mathcal{P}_{X_j}$ für alle $i, j \in I$.

4.9 Definition. Es ei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $L^1(\Omega, \mathcal{A}, \mathcal{P})$. Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das *starke* (bzw. das *schwache*) *Gesetz der großen Zahlen*, falls

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - E[X_i]) \to 0 \ \mathcal{P}\text{-fast sicher (bzw. in Wahrscheinlichkeit).}$$

4.10 Satz. (Schwaches Gesetz der großen Zahlen)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von paarweise unkorrelierten Zufallsvariablen in $L^2(\Omega, \mathcal{A}, \mathcal{P})$ mit

$$(W) \qquad \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = 0.$$

Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das schwache Gestz der großen Zahlen. Hinreichende Bedingungen für (W) sind:

(i) $(X_n)_{n\in\mathbb{N}}$ ist identisch verteilt.

(ii) $\exists C > 0 \, \forall n \in \mathbb{N} : Var(X_n) \leq C$

(iii) $\sum\limits_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty$

Beweis. Aufgrund der Tschebyscheff-Ungleichung und des Satzes von Bienaymé gilt für $\varepsilon > 0$

$$\mathcal{P}\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - E[X_i]\right| \geq \varepsilon\right\}\right) = \mathcal{P}\left(\left\{\left|\sum_{i=1}^{n} X_i - E\left[\sum_{i=1}^{n} X_i\right]\right| \geq n\varepsilon\right\}\right)$$
$$\leq \frac{Var\left(\sum\limits_{i=1}^{n} X_i\right)}{n^2\varepsilon^2} = \frac{1}{\varepsilon^2}\left(\frac{1}{n^2}\sum_{i=1}^{n} Var(X_i)\right) \to 0 \quad (n \to \infty)$$

wegen (W). Offensichtlich gilt: (i)$\Leftrightarrow$(ii)$\Leftrightarrow$$(W)$. (iii)$\Leftrightarrow$$(W)$ folgt aus dem folgenden Lemma von Kronecker mit $b_n = n^2$ und $x_n = \frac{Var(X_n)}{n^2}$. $\square$

4.11 Lemma. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$.

(i) *(Töplitz)* Sei $b_n = \sum\limits_{i=1}^{n} a_i$ für nicht-negative $(a_i)_{i\in\mathbb{N}}$, so dass $b_n \nearrow +\infty$. Falls $x_n \to x$ in $\mathbb{R}$, so gilt auch $\frac{1}{b_n}\sum\limits_{i=1}^{n} a_i x_i \to x$.

(ii) *(Kronecker)* Sei $(b_n)_{n\in\mathbb{N}}$ eine nicht-fallende Folge von reellen Zahlen mit $b_n \nearrow +\infty$. Falls $\sum\limits_{n=1}^{\infty} x_n < \infty$, so gilt $\lim\limits_{n\to\infty} \frac{1}{b_n}\sum\limits_{i=1}^{n} b_i x_i = 0$.

Beweis. (i) Sei $\varepsilon > 0$ und N_0 so gewählt, dass $|x_n - x| \leq \frac{\varepsilon}{2}$ für alle $n \geq N_0$. Wähle $N_1 \geq N_0$ so, dass $\frac{1}{b_{N_1}}\sum\limits_{i=1}^{N_0} a_i|x_i - x| \leq \frac{\varepsilon}{2}$. Dann gilt für alle $n \geq N_1$:

$$\begin{aligned}
|\frac{1}{b_n}\sum_{i=1}^{n} x_i a_i - x| &\leq \frac{1}{b_n}\sum_{i=1}^{n} a_i|x_i - x| \\
&\leq \frac{1}{b_{N_1}}\sum_{i=1}^{N_0} a_i|x_i - x| + \frac{1}{b_n}\sum_{i=N_0+1}^{n} a_i|x_i - x| \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\frac{b_n - b_{N_0}}{b_n} \leq \varepsilon
\end{aligned}$$

(ii) Setze $b_0 = S_0 = 0, S_n = \sum\limits_{i=1}^{n} x_i$ für $n \geq 1$. Partielle Summation liefert:

$$\sum_{j=1}^{n} b_j x_j = \sum_{j=1}^{n} b_j(S_j - S_{j-1}) = b_n S_n - \underbrace{b_0 S_0}_{=0} - \sum_{j=1}^{n} S_{j-1}(b_j - b_{j-1})$$

Also:

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j = S_n - \frac{1}{b_n}(b_j - b_{j-1})S_j \to 0.$$

Dach nach Teil (i) mit S_n auch $\frac{1}{b_n}\sum\limits_{j=1}^{n} S_j(b_j - b_{j-1})$ gegen den selben Grenzwert konvergiert. $\square$

Nächstes Ziel ist die Herleitung des folgenden starken Gesetz der großen Zahlen:

4.12 Satz. (Kolmogorov)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen in $L^1(\Omega, \mathcal{A}, \mathcal{P})$. Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das starke Gesetz der großen Zahlen, falls

(i) $(X_n)_{n\in\mathbb{N}}$ identisch verteilt ist.

(ii) $\sum\limits_{n=1}^{\infty} \frac{Var(X_n)}{n^2} < \infty$

Zur Vorbereitung:

4.13 Satz. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen in $L^2(\Omega, \mathcal{A}, \mathcal{P})$ mit $E[X_n] = 0$ für alle $n \in \mathbb{N}$. Sei $S_n = \sum\limits_{i=1}^{n} X_i$. Dann:

(i) *(Kolmogorosche Ungleichung)* Für alle $\varepsilon > 0$ gilt: $\mathcal{P}(\{\max\limits_{1\leq k\leq n} |S_k| \geq \varepsilon\}) \leq \frac{E[|S_n|^2]}{\varepsilon^2}$.

(ii) Falls $\sum\limits_{n=1}^{\infty} E[X_n^2] < \infty$, so konvergiert $(S_n)_{n\in\mathbb{N}}$ $\mathcal{P}$-fast sicher für $n \to \infty$.

Beweis. (i) Seien $A := \{\max\limits_{1\leq k\leq n} |S_k| \geq \varepsilon\}$ und $A_k := \{\forall I = 1, ..., k-1\, |S_i| < \varepsilon$ und $|S_k| \geq \varepsilon\}$, $k = 1, ..., n$. Dann ist A disjunkte Vereinigung der A_k. Also: $\varepsilon^2\mathcal{P}(A) = \varepsilon^2\sum\limits_{k=1}^{n}\mathcal{P}(A_k) \leq \sum\limits_{k=1}^{n} E[\mathbb{1}_{A_k}|S_k|^2]$. Aufgrund der Unabhängigkeit der (X_i) und da $E[X_i] = 0$, folgt $E\left[\mathbb{1}_{A_k} S_k \sum\limits_{i=k+1}^{n} X_i\right] = E[\mathbb{1}_{A_k} S_k] E\left[\sum\limits_{i=k+1}^{n} X_i\right] = 0$. Also:

$$E[\mathbb{1}_{A_k} S_k^2] \leq E[\mathbb{1}_{A_k} S_k^2] + 2E\left[\mathbb{1}_{A_k} S_k\left(\sum_{i=k+1}^{n} X_i\right)\right] + E\left[\mathbb{1}_{A_k}\left(\sum_{i=k+1}^{n} X_i\right)^2\right] = E[\mathbb{1}_{A_k} S_n^2].$$

Zusammen folgt: $\varepsilon^2\mathcal{P}(A) \leq \sum\limits_{k=1}^{n} E[\mathbb{1}_{A_k} S_n^2] = E[\mathbb{1}_{A_k} S_n^2] \leq E[S_n^2]$.

(ii) Aufgurnd bon Satz 4.7 reicht es zu zeigen, dass für jedes $\varepsilon > 0$ $\lim_{n\to\infty} \mathcal{P}(\{\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\}) = 0$ gilt. Dies folgt aus (i):

$$\begin{aligned}
\mathcal{P}(\{\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\}) &= \lim_{N\to\infty} \mathcal{P}(\{\max_{1\leq k\leq n} |S_{n+k} - S_n| \geq \varepsilon\}) \\
&\leq \lim_{n\to\infty} \frac{1}{\varepsilon^2} E[(\sum_{k=n}^{n+N} X_k)^2] \\
&\overset{\text{Bienaymé}}{=} \lim_{N\to\infty} \frac{1}{\varepsilon^2} \sum_{k=n}^{n+N} E[X_k^2] = \frac{1}{\varepsilon^2} \sum_{k=n}^{\infty} E[X_k^2] \to 0 \quad (n\to\infty),
\end{aligned}$$

da $\sum\limits_{k=1}^{\infty} E[X_k^2] < \infty$.

□

Beweis. (von Satz 4.12)

(ii) Sei $S_n = \sum\limits_{k=1}^{n} X_k - E[X_k]$. Lemma 4.11(ii) besagt, dass $\frac{1}{n} S_n \to 0$ $\mathcal{P}$-fast sicher, falls $\sum\limits_{k=1}^{\infty} \frac{X_k - E[X_k]}{k} < \infty$ $\mathcal{P}$-fast sicher. Letzteres folgt aus Satz 4.13(ii), da

$$\sum_{k=1}^{\infty} E[(\frac{X_k - E[X_k]}{k})^2] = \sum_{k=1}^{\infty} \frac{1}{k^2} Var(X_k) < \infty.$$

(i) Œ sei $E[X_i] = 0$ durch Übergang von X_i zu $X_i - E[X_i]$.
<u>Schritt 1:</u> Wir zeigen: $\mathcal{P}(\{|X_n| \geq n$ für unendlich viele $n\}) = 0$ $(*)$.
Da $E[|X_1|] < \infty$ folgt:

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathcal{P}(\{|X_n| > n\}) &= \sum_{n=1}^{\infty} \mathcal{P}(\{|X_1| > n\}) \\
&= \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} \mathcal{P}(\{k \leq |X_1| < k+1\}) \\
&= \sum_{k=1}^{\infty} k\mathcal{P}(\{k \leq |X_1| < k+1\}) \\
&\leq \sum_{k=1}^{\infty} E[\mathbb{1}_{\{k\leq |X_1| < k+1\}} |X_1|] = E[|X_1|] < \infty
\end{aligned}$$

Das Lemma von Borel-Cantelli impliziert nun $(*)$.
<u>Schritt 2:</u> Sei

$$\tilde{X}_n := \begin{cases} X_n & , |X_n| \leq n \\ 0 & , \text{ sonst.} \end{cases}$$

Wir zeigen: Das starke Gesetz der großen Zahlen gilt für $(X_n)_{n\in\mathbb{N}}$. $\Leftrightarrow$ Das starke Gesetz der großen Zahlen gilt für $(\tilde{X}_n)_{n\in\mathbb{N}}$.
Aus Schritt 1 folgt: $\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{n}\sum\limits_{i=1}^{n} \tilde{X}_i \to 0 (n\to\infty)$ $\mathcal{P}$-fast sicher.
Ferner: $\frac{1}{n}\sum\limits_{i=1}^{n} R[\tilde{X}_i] = \frac{1}{n}\sum\limits_{i=1}^{n} E[X_i \mathbb{1}_{\{|X_i|\leq i\}}] = \frac{1}{n}\sum\limits_{i=1}^{n} E[X_1 \mathbb{1}_{\{|X_1|\leq i\}}]$.
Da $E[X_1 \mathbb{1}_{\{|X_1|\leq n\}}] \to E[X_1] = 0$ nach dem Satz über dominante Konvergenz, folgt aus Lemma 4.11(i), dass $\frac{1}{n}\sum\limits_{i=1}^{n} E[\tilde{X}_i] \to 0$.
<u>Schritt 3:</u> Das starke Gesetz der großen Zahlen gilt für $(\tilde{X}_n)_{n\in\mathbb{N}}$.
Die Folge $(\tilde{X}_n)_{n\in\mathbb{N}}$ ist unabhängig. Ferner gilt:

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{Var(\tilde{X}_n)}{n^2} \leq \sum_{n=1}^{\infty} \frac{E[\tilde{X}_n^2]}{n^2} &\leq \sum_{n=1}^{\infty} \frac{1}{n^2} E[X_n^2 \mathbb{1}_{\{|X_n|\leq n\}}] \\
&= \sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^{n} E[X_n^2 \mathbb{1}_{\{k-1\leq |X_n| < k\}}] \\
&= \sum_{k=1}^{\infty} E[X_1^2 \mathbb{1}_{\{k-1\leq |X_1| \leq k\}}] \sum_{n=k}^{\infty} \frac{1}{n^2}.
\end{aligned}$$

Da $\sum\limits_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k}$, folgt:

$$\sum_{n=1}^{\infty} \frac{Var(\tilde{X}_n)}{n^2} \leq \sum_{k=1}^{\infty} \frac{2}{k} E[X_1^2 \mathbb{1}_{\{k-1 \leq |X_1| < k\}}] \leq 2 \sum_{k=1}^{\infty} E[|X_1| \mathbb{1}_{\{k-1 \leq |X| < k\}}] = E[|X_1|] < \infty.$$

Also ist Teil (i) anwendbar. □

4.14 Beispiel. (Ein vorteilhaftes Spiel?)
Sei $(\xi_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Bernoulli($\frac{1}{2}$)-Experimenten („Münzwürfe“). Bei einem Anfangskapital von $X_0 = 1$ wird das Kapital X_n nach der n-ten Runde in Abhängigkeit vom $(n+1)$-ten Münzwurf halbiert, falls in der $(n+1)$-ten Runde Kopf („0“) fällt, und man erhält $\frac{2}{3}X_n$ hinzu, falls in der $(n+1)$-ten Runde Zahl („1“) fällt. Dann gilt für $n \geq 1$: $X_n = Y_n X_{n-1} = \prod\limits_{i=1}^{n} Y_i$, wobei

$$Y_n = \begin{cases} \frac{1}{2} & ,\xi_n = 0, \\ \frac{5}{3} & ,\xi_n = 1. \end{cases}$$

Da mit $(\xi_i)_{i\in\mathbb{N}}$ auch $(Y_i)_{i\in\mathbb{N}}$ unabhängig ist, folgt:

$$E[X_n] = E\left[\prod_{i=1}^{n} Y_i\right] = \prod_{i=1}^{n} E[Y_i] = E[Y_1]^n = (\frac{1}{2} \cdot \frac{1}{2} + \frac{5}{3} \cdot \frac{1}{2})^n = (\frac{13}{12})^n \to +\infty \qquad (n \to \infty).$$

Sei nun $\mu := E[\log(Y_1)] = \frac{1}{2}(\log(\frac{1}{2}) + \log(\frac{5}{3})) < \frac{1}{2}(\log(\frac{1}{2}) + \log(2)) = 0$. Nach dem starken Gesetz der großen Zahlen (Satz 4.12) folgt $\frac{1}{n}\sum\limits_{i=1}^{n} \log(Y_i) \to \mu$ $\mathcal{P}$-fast sicher, also $\sum\limits_{i=1}^{n} \log(Y_i) \to -\infty$ $\mathcal{P}$-fast sicher. Damit gilt aber auch $X_n = e^{\sum_{i=1}^{n} \log(Y_i)} \to 0$ $\mathcal{P}$-fast sicher.
Hierbei strebt der erwartete Gewinn mit der Anzahl n der Runden gegen $+\infty$, aber in fast allen Szenarien strebt der Gewinn $X_n - X_0$ gegen -1.

4.15 Beispiel. (Weierstraßscher Approximationssatz)
Sei $f : [0,1] \to \mathbb{R}$ stetig. Der Weierstraßsche Approxiamtionssatz besagt:
Zu f existiert eine Folge von Polynomen f_n mit $\sup\limits_{y\in[0,1]} |f_n(y) - f(y)| \to 0 \quad (n \to \infty)$.

Sei zunächst $y \in [0,1]$ beliebig, aber fest gewählt. Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Bernoulli(y)-verteilten Zufallsvariablen. Nach dem schwachen Gesetz der großen Zahlen konvergiert $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ gegen $E[X_1] = y$ in Wahrscheinlichkeit und also auch in Verteilung (Satz 4.3). Also gilt:

$$E[f(\overline{X_n})] \to f(y) \qquad (n \to \infty).$$

Aber: $f_n(y) := E[f(\overline{X_n})] = E[(\frac{1}{n} \underbrace{\sum_{i=1}^{n} X_i}_{\text{binomial}(n,y)})] = \sum\limits_{k=0}^{n} \binom{n}{k} f(\frac{n}{k}) y^k (1-y)^{n-k}$ ist ein Polynom in y, das sogenannte *Bernsteinpolynom* n-ten Grades zu f.
Es bleibt zu zeigen, dass die Konvergenz von f_n gegen f gleichmäßig in y ist. Wie in Beweis von Satz 4.3(iv) folgt:

$$\forall \varepsilon > 0\ \exists \delta > 0: \quad |E[f(\overline{X_n})] - f(y)| \leq \frac{\varepsilon}{2} + 2 \sup_{x\in[0,1]} |f(x)| \cdot \mathcal{P}(\{|\overline{X_n} - y| > \delta\}).$$

Da für alle $y \in [0,1]$

$$\mathcal{P}(\{|\overline{X_n} - y| > \delta\}) \leq \frac{Var(\overline{X_n})}{\delta^2} = \frac{ny(1-y)}{\delta^2 n^2} \leq \frac{\frac{1}{4}}{\delta^2 n}$$

(Tschebyscheff-Ungleichung und Varianz der Binomialverteilung), folgt

$$\forall \varepsilon > 0\ \exists \delta > 0: \quad \sup_{y\in[0,1]} |f_n(y) - f(y)| \leq \frac{\varepsilon}{2} + \frac{1}{2} \sup_{x\in[0,1]} |f(x)| \frac{1}{\delta^2 n}.$$

Dies zeigt die gleichmäßige Konvergenz von (f_n) gegen f.

4.16 Beispiel. (Normale Zahlen)
Sei $g \geq 2$ eine natürliche Zahl und $\omega \in [0,1)$. Dann hat ω eine g-adische Entwicklung

$$\omega = \sum_{n=1}^{\infty} \xi_n g^{-n},$$

wobei $\xi_n = \xi_n(\omega) \in \{0, ..., g-1\}$ für alle $n \in \mathbb{N}$ und $\xi_n \in \{0, ..., g-2\}$ für unendlich viele $n \in \mathbb{N}$. Die Zahl ω heißt *einfach g-normal*, falls für alle $\varepsilon = 0, ..., g-1$

$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g} = \frac{1}{g}, \quad S_n^{\varepsilon,g} := \#\{i \leq n \,|\, \xi_i(\omega) = \varepsilon\}.$$

Ist ω einfach g-normal für alle $g \geq 2$, so wird ω *absolut normale* genannt.
Sei $\Omega = [0,1)$, $\mathcal{A}$ die Spur-σ-Algebra von $\mathcal{B}$ auf $[0,1)$ und $\mathcal{P}$ die Einschränkung des Lebesgue-Maßes uf $[0,1)$. Dann gilt für alle $\varepsilon = 0, ..., g-1$ (g beliebig aber fest):

$$\{\xi_i = \varepsilon\} = \bigcup_{k=0}^{g^{i-1}-1} \left[\frac{kg+\varepsilon}{g^i}, \frac{kg+\varepsilon+1}{g^i}\right] =: A_i^\varepsilon$$

Denn $\omega \cdot g^i = g \cdot \left(\sum\limits_{n=1}^{i-1} \xi_n(\omega) g^{i-1-n}\right) + \xi_i(\omega) + \sum\limits_{n=i+1}^{\infty} g^{i-n}$. Die Folge von Ereignissen $(A_i^\varepsilon)_{i\in\mathbb{N}}$ ist für festes $\varepsilon = 0, ..., g-1$ unabhängig. Für $g = 2, \varepsilon = 0$ wurde dies in den Übungen gezeigt. Den allgemeinen Fall überprüft man analog.
Ferner gilt

$$\mathcal{P}(A_i^\varepsilon) = \sum_{k=0}^{g^{i-1}-1} \frac{1}{g^i} = \frac{1}{g}.$$

Sei nun $X_i^\varepsilon := \mathbb{1}_{A_i^\varepsilon}$. Dann:

$$S_n^{\varepsilon,g} = \frac{1}{n} \sum_{i=1}^{n} X_i^\varepsilon$$

Die Folge $(X_i^\varepsilon)_{i\in\mathbb{N}}$ erfüllt die Voraussetzungen des starken Gesetzes der großen Zahlen. Also

$$\lim_{n\to\infty} S_n^{\varepsilon,g}(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^\varepsilon = E[X_1^\varepsilon] = \mathcal{P}(A_1^\varepsilon) = \frac{1}{g}$$

für $\mathcal{P}$-fast alle $\omega \in [0,1)$. Dies zeigt, dass Lebesgue-fast alle $\omega \in [0,1)$ absolut normal sind.

4.3 Der zentrale Grenzwertsatz von Lindeberg-Lévy

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch-verteilten Zufallsvariablen in $L^2(\Omega, \mathcal{A}, \mathcal{P})$. Seien $\mu = E[X_1]$ und $\sigma^2 = Var(X_i)$. Sei $\sigma^2 > 0$. Dann gilt für alle $\alpha \in [0, \frac{1}{2})$:

$$\sum_{n=1}^{\infty} E\left[\left(\frac{X_n - E[X_n]}{n^{1-\alpha}}\right)^2\right] = \sum_{n=1}^{\infty} \frac{Var(X_n)}{n^{2-2\alpha}} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^{2-2\alpha}} < \infty$$

Nach dem Reihenkriterium aus Satz 4.3(ii) folgt, dass $\sum\limits_{n=1}^{\infty} \frac{X_n - E[X_n]}{n^{1-\alpha}}$ $\mathcal{P}$-fast sicher konvergiert. Also impliziert das Lemma von Kronecker (Lemma 4.11(ii)), dass

$$\frac{1}{n^{1-\alpha}} \sum_{i=1}^{n} X_i - E[X_i] \to 0 \quad \mathcal{P}\text{-fast sicher.}$$

Frage: Was passiert für $\alpha = \frac{1}{2}$?

4.17 Satz. (Zentraler Grenzwertsatz von Lindeberg-Lévy)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch-verteilten Zufallsvariablen in $L^2(\Omega, \mathcal{A}, \mathcal{P})$ mit $\sigma^2 := Var(X_1) > 0$. Sei $\mu := E[X_1]$. Dann gilt für alle $y \in \mathbb{R}$:

$$\mathcal{P}\left(\left\{\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\infty} (X_i - \mu) \leq y\right\}\right) \to \Phi(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}} dx \qquad (n \to \infty)$$

Diskussion:

(a) Da Φ die Verteilungsfunktion einer Standardnormalverteilung ist, sehen wird, dass die Verteilungsfunktion von $\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)$ gegen die Verteilungsfunktion einer $\mathcal{N}(0,1)$-Verteilung konvergiert. Dies werden wir mit den Begriffen „schwache Konvergenz“ und „wesentliche Konvergenz“ in Zusammenhang bringen.

(b) Die Normierung $\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)$ ist so gewählt, dass $E\left[\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)\right]=0$ und $Var\left(\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)\right)=\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}\underbrace{Var(X_i)}_{=\sigma^2}=1$. Für $S_n=\sum\limits_{i=1}^{n}X_i$ gilt also $\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^{n}(X_i-\mu)=\frac{S_n-E[S_n]}{Std(S_n)}$ $(=\sqrt{Var(S_n)})$.

4.18 Beispiel. Eine faire Münze wird 100 Mal unabhängig geworfen. Wie groß ist die Wahrscheinlichkeit, dass mindestens 60 Mal Kopf fällt?
Die Zufallsvariable X_i modelliere den i-ten Wurf, wobei „$X_i=1$“ Kopf im i-ten Wurf und „$X_i=0$“ Zahl im i-ten Wurf bedeutet.
Die Zufallsvariablen $X_1,...,X_{100}$ sind alle Bernoulli($\frac{1}{2}$)-verteilt und unabhängig. Somit ist $S_{100}=\sum\limits_{i=1}^{100}X_i$ dann binomial($100,\frac{1}{2}$)-verteilt. Also $\mathcal{P}(\{S_{100}\geq 60\})=\sum\limits_{k=60}^{100}B_{100,\frac{1}{2}}(k)=\sum\limits_{k=60}^{100}\binom{100}{k}(\frac{1}{2})^{100}\approx 2,84\%$.

Frage: Wie gut ist die Approximation durch den zentralen Grenzwertsatz?

$$\mathcal{P}(\{S_{100}\geq 60\})=1-\mathcal{P}(\{S_{100}\leq 59\})=1-\left(\left\{\frac{\sum\limits_{i=1}^{100}X_i-\frac{1}{2}}{\sqrt{100\cdot\frac{1}{4}}}\leq\frac{59-50}{\sqrt{\frac{100}{4}}}\right\}\right)\approx 1-\int\limits_{-\infty}^{\frac{9}{5}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}dx\approx 3,59\%$$

Da $\sum\limits_{i=1}^{n}X_i$ für unabhängige Bernoulli(p)-verteilte Zufallsvariable nur ganzzahlige Werte annehmen kann, hätten wir obiger Rechnung auch jeden Wert im Intervall $[59,60[$ vermuten können. Dem tragen folgende in der Praxis oft verwendete Korrekturformeln Rechnung: Für $\alpha\in\mathbb{N}$ gilt

$$\mathcal{P}\left(\left\{\sum_{i=1}^{n}X_i\geq\alpha\right\}\right)=\mathcal{P}\left(\left\{\sum_{i=1}^{n}X_i>\alpha-\frac{1}{2}\right\}\right)=\mathcal{P}\left(\left\{\frac{(\sum\limits_{i=1}X_i-p)}{\sqrt{np(1-p)}}>\frac{\alpha-np-\frac{1}{2}}{\sqrt{np(1-p)}}\right\}\right)\approx 1-\Phi\left(\frac{\alpha-np-\frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

und $\mathcal{P}\left(\left\{\sum\limits_{i=1}^{n}X_i\leq\alpha\right\}\right)\approx\Phi\left(\frac{\alpha-np+\frac{1}{2}}{\sqrt{np(1-p)}}\right)$. Im konkreten Beispiel:

$$\mathcal{P}(\{S_{100}\geq 60\})\approx 1-\Phi\left(\frac{60-50-\frac{1}{2}}{\sqrt{\frac{100}{4}}}\right)\approx 2,87\%$$

Zum Vergleich: Die Abschätzung mittels der Cholesky-Ungleichung, auf der das schwache Gesetz der großen Zahlen basiert, liefert hier:

$$\mathcal{P}(\{S_{100}\geq 60\})\leq\mathcal{P}\left(\left\{\left|\frac{1}{100}\sum_{i=1}^{100}X_i-\frac{1}{2}\right|\geq 0,1\right\}\right)\leq\frac{(\frac{1}{2})^2}{100\cdot 0,1^2}=\frac{1}{4}$$

4.19 Beispiel. (Monte-Carlo Integration)
Sei $g:[0,1]^D\to\mathbb{R}$ stetig. Wir wollen $\int\limits_{[0,1]^D}g\,d\lambda^{\otimes D}$ berechnen (numerisch approximieren). Sei $(U_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, auf $[0,1]^D$ gleichverteilten Zufallsvariablen, also mit Dichte $f_{(x_1,..,x_d)}=\prod\limits_{d=1}^{D}\mathbb{1}_{[0,1]}\cdot x_d$. Dann:

$$E[g(U_1)]=\int\limits_{\mathbb{R}}g(x)f(x)d\lambda^{\otimes D}=\int\limits_{[0,1]^D}g(x)d\lambda^{\otimes D}$$

Aus dem starken Gesetz der großen Zahlen folgt $\frac{1}{n}\sum\limits_{i=1}^{n}g(U_i)\to E[g(U_1)]$ $\mathcal{P}$-fast sicher. Diese Methode nennt man Monte-Carlo-Verfahren.
Frage: Wie schnell konvergiert die Monte-Carlo-Methode?
Hier hilft der zentrale Grenzwertsatz: Sei

$$\sigma^2:=Var(g(U_1))=\int\limits_{[0,1]^D}g^2(x)d\lambda^{\otimes D}-\mu^2<\infty.$$

Wir nehmen an, dass $\sigma^2 > 0$. Dann gilt für beliebiges $c > 0$:

$$\mathcal{P}\left(\left\{\mu \in \left]\frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{c\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n g(U_i) + \frac{c\sigma}{\sqrt{n}}\right]\right\}\right) = \mathbb{P}\left(\left\{-c < \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n g(U_i) - \mu \leq c\right\}\right)$$
$$\to \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Ist etwa $c = 1,96$, so ist $\Phi(c) \approx 0,973$, also:

$$\mathcal{P}\left(\left\{\mu \in \left]\frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{1,96\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n g(U_i) + \frac{1,96\sigma}{\sqrt{n}}\right]\right\}\right) \to 97,5\%$$

Der gesuchte Integralwert μ liegt dann für großes n mit einer Wahrscheinlichkeit von ca. 95% in dem zufälligen Intervall der Länge $\frac{2\cdot 1,96\sigma}{\sqrt{n}}$. Mit wachsender Anzahl n der zu generierenden Zufallsvariablen $U_1, ..., U_n$ schrumpft das Intervall mt der Ordnung $\frac{1}{\sqrt{n}}$. Für kleine Dimensionen D sind deterministische Methoden effizienter. Der Vorteil des Monte-Carlo-Verfahrens liegt darin, dass die Konvergenzordnung nicht von der Dimension abhängt, also auch für sehr große D verwendet werden kann.

Zum Beweis des zentralen Grenzwertsatzes benötigen wir einige Vorbereitungen über Konvergenz von Wahrscheinlichkeitsmaßen:

4.20 Definition. Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ und $\mathcal{P}$ ein weiteres Wahrscheinlichkeitsmaß. Dann sagt man:

(i) $(\mathcal{P}_n)_{n\in\mathbb{N}}$ *konvergiert schwach gegen* $\mathcal{P}$, falls für alle stetigen und beschränkten Funktionen $f : \mathbb{R} \to \mathbb{R}$ gilt, dass
$$\lim_{n\to\infty} \int f\, d\mathcal{P}_n = \int f\, d\mathcal{P}.$$

(ii) $(\mathcal{P}_n)_{n\in\mathbb{N}}$ *konvergiert wesentlich gegen* $\mathcal{P}$, falls für alle $A \in \mathcal{B}$ mit $\mathcal{P}(\partial A) = 0$ gilt, dass
$$\lim_{n\to\infty} \mathbb{P}_n(A) = \mathbb{P}(A).$$

(iii) Seien $F_n(x) = \mathcal{P}_n(]-\infty, x])$, $F(x) = \mathcal{P}(]-\infty, x])$ die zugehörigen Verteilungen. Dann konvergiert $(F_n)_{n\in\mathbb{N}}$ wesentlich gegen F, falls für alle Stetigkeitsstellen x von F gilt, dass
$$\lim_{n\to\infty} F_n(x) = F(x).$$

Bemerkung. Die Aussage des zentralen Grenzwertsatzes besagt, dass die Verteilung der standardisierten Summe $\frac{1}{\sigma\sqrt{n}}\sum\limits_{i=1}^n X_i - \mu$ wesentlich gegen Φ konvergiert. Dies ist äquivalent zu den anderen Konvergenzarten in 4.20.

4.21 Satz. (Portmanteau-Theorem)
Seien $(\mathcal{P}_n)_{n\in\mathbb{N}}$ und $\mathcal{P}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$. $(F_n)_{n\in\mathbb{N}}$ und F seien die zugehörigen Verteilungsfunktionen. Dann sind äquivalent:

(a) $\mathcal{P}_n \to \mathcal{P}$ schwach

(b) Für alle abgeschlossenen Mengen $A \subset \mathbb{R}$: $\limsup\limits_{n\to\infty} \mathcal{P}_n(A) \leq \mathcal{P}(A)$

(c) Für alle offenen Mengen $A \subset \mathbb{R}$: $\liminf\limits_{n\to\infty} \mathcal{P}_n(A) \geq \mathcal{P}(A)$

(d) $\mathcal{P}_n \to \mathcal{P}$ im Wesentlichen

(e) $F_n \to F$ im Wesentlichen

Beweis. **(a)⇒(b)** Sei $\varrho(x, A) := \inf\{|x - y|; y \in A\}$ der Abstand von $x \in \mathbb{R}$ zur abgeschlossenen Menge A. Sei
$$g(t) := \begin{cases} 1 & , t \leq 0 \\ 1 - t & , t \in (0, 1] \\ 0 & , t > 1. \end{cases}$$

Sei $k \in \mathbb{N}$ und betrachte $A_k := \{x \in \mathbb{R} \mid \varrho(x, A) \leq \frac{1}{k}\}$. Definiere eine stetige und beschränkte Funktion f_k durch $f_k(x) := g(k\varrho(x, A))$. Aus (a) folgt:
$$\limsup_{n\to\infty} \mathcal{P}_n(A) = \limsup_{n\to\infty} \int \mathbb{1}_A d\mathcal{P}_n \leq \limsup_{n\to\infty} \int f_k d\mathcal{P}_n \overset{(a)}{=} \int f_k d\mathcal{P} \leq \int \mathbb{1}_{A_k} d\mathcal{P} = \mathcal{P}(A_k).$$

Die Stetigkeit von $\mathcal{P}$ von oben impliziert $\lim\limits_{n\to\infty} \mathcal{P}(A_k) = \mathcal{P}\left(\bigcap\limits_{k\in\mathbb{N}} A_k\right) = \mathcal{P}(A)$, da A abgeschlossen. Also mit $k \nearrow +\infty$: $\limsup\limits_{n\to\infty} \mathcal{P}_n(A) \leq \mathcal{P}(A)$

(b)⇒(c) Sei A offen. Dann ist A^c abgeschlossen. Also folgt aus (b):

$$\liminf_{n\to\infty} \mathcal{P}_n(A) = 1 - \limsup_{n\to\infty} \mathcal{P}_n(A^c) \geq 1 - \mathcal{P}(A^c) = \mathcal{P}(A)$$

(c)⇒(b) Analog.

(c)⇒(d) Mit (c) gilt auch (b). Da $A \subset \overline{A}$, gilt für beliebiges $A \in \mathcal{B}$: $\limsup_{n\to\infty} \mathcal{P}_n(A) \leq \limsup_{n\to\infty} \mathcal{P}_n(\overline{A}) \leq \mathcal{P}(\overline{A})$ (mit (b)).

Da $\mathring{A} \subset A$ (wobei $\mathring{A}$ das Innere von A ist) folgt aus (c) $\liminf_{n\to\infty} \mathcal{P}_n(A) \geq \liminf_{n\to\infty} \mathcal{P}_(\mathring{A}) \geq \mathcal{P}(\mathring{A})$.

Sei nun $\mathcal{P}(\partial A) = 0$. Da $\overline{A} = \mathring{A} \cup \partial A$, erhalten wir $\mathcal{P}(\mathring{A}) \leq \mathcal{P}(A) \leq \mathcal{P}(\overline{A}) = \mathcal{P}(A) + \underbrace{\mathcal{P}(\partial A)}_{=0} = \mathcal{P}(\mathring{A})$.

Somit: $\liminf_{n\to\infty} \mathcal{P}_n(A) \geq \mathcal{P}(\mathring{A}) = \mathcal{P}(A) = \mathcal{P}(\overline{A}) \geq \limsup_{n\to\infty} \mathcal{P}_n(A)$ Also: $\lim_{n\to\infty} \mathcal{P}_n(A) = \mathcal{P}(A)$

(d)⇒(a) Sei $f : \mathbb{R} \to \mathbb{R}$ stetig mit $|f(x)| \leq M$ für alle $x \in \mathbb{R}$ ($M > 0$). Dann ist die Menge $D = \{t \in \mathbb{R} \,|\, \mathcal{P}(\{f = t\}) \neq 0\}$ höchstens abzählbar (die Mengen $f^{-1}(\{t\})$, $t \in D$, sind disjunkt und $\mathcal{P}$ ist ein endliches Maß).

Sei Œ $-M, M \in D$. Fixiere $\varepsilon > 0$ und wähle eine Partition $\pi = (t_0, ..., t_k)$ von $[-M, M]$ (also: $-M = t_0 < t_1 < ... < t_k = M$) mit $|\pi| := \max_{i=0,...,k-1}(t_{i+1} - t_i) \leq \varepsilon$. Diese Partition sei so gewählt, dass $t_i \notin D$ für alle $i = 0, .., k-1$.

Seien die Mengen B_i definiert durch $B_i = \{t_i \leq f < t_{i+1}\}$, $i = 0, ..., k-2$, $B_{k-1} = \{t_{k-1} \leq f < t_k\}$.

Da f stetig ist, ist $f^{-1}((t_i, t_{i+1}))$ offen für alle $i = 0, ..., k-1$. Somit:

$$\mathcal{P}(\partial B_i) \leq \mathcal{P}(f^{-1}(\{t_i\}) \cup f^{-1}(\{t_{i+1}\}) = \mathcal{P}(\{f = t_i\}) + \mathcal{P}(\{f = t_{i+1}\}) = 0$$

Daher liefert (d):

$$\sum_{i=0}^{k-1} t_i \mathcal{P}_n(B_i) \to \sum_{i=0}^{k-1} t_i \mathcal{P}(B_i) \quad (n \to \infty) \quad (*)$$

Somit:

$$\begin{aligned}\left|\int f d\mathcal{P} - \int f d\mathcal{P}_n\right| &\leq \left|\sum_{i=0}^{k-1} \int \mathbb{1}_{B_i}(f - t_i) d\mathcal{P}_n\right| + \left|\sum_{i=0}^{k-1} t_i \mathcal{P}_n(B_i) - \sum_{i=1}^{N} t_i \mathcal{P}(B_i)\right| + \left|\sum_{i=0}^{N} \int \mathbb{1}_{B_i}(t_i - f) d\mathcal{P}\right| \\ &\leq 2\varepsilon + \left|\sum_{i=0}^{k-1} t_i \mathcal{P}_n(B_n) - \sum_{i=0}^{k-1} t_i \mathcal{P}(B_n)\right| \leq 3\varepsilon\end{aligned}$$

wegen $(*)$, falls $n \geq N(\varepsilon)$ hinreichend groß.

(d)⇒(e) Da für $x \in \mathbb{R}$ $\partial(-\infty, x] = \{x\}$ gilt, folgt:

$$\mathcal{P}(\partial(-\infty, x]) = \mathcal{P}(\{x\}) = 0 \Leftrightarrow F \text{ stetig in } x \quad (\text{ da} \mathcal{P}((-\infty, x)) = F(x-))$$

Somit gilt an allen Stetigkeitsstellen x von F nach (d): $F_n(x) = \mathcal{P}_n((-\infty, x]) \to \mathcal{P}((-\infty, x]) = F(x)$

(e)⇒(c) Sei A offen. Dann existiert eine abzählbare Menge von paarweise disjunkten Intervallen $I_k = (a_k, b_k)$, $a_k \leq b_k \in \mathbb{R}$, mit $A = \bigcup_{k\in\mathbb{N}} I_k$. Sei $\varepsilon > 0$.

Falls $I_k = (a_k, b_k)$ nicht leer ist, wähle man ein Teilintervall der Form $I'_k = (a'_k, b'_k] \subset (a_k, b_k)$, so dass a'_k, b'_k Stetigkeitsstellen von F sind und $\mathcal{P}(I_k) \leq \mathcal{P}(I'_k) + \varepsilon 2^{-k}$.

Das Lemma von Fatou mit $(\mathbb{N}, 2^N, \mu)$ wobei $\mu(A) = \#A$ das Zählmaß ist, liefert

$$\liminf_{n\to\infty} \mathcal{P}_n(A) \overset{\sigma\text{-add.}}{=} \liminf_{n\to\infty} \sum_{k=1}^{\infty} \mathcal{P}_n(I_k) \geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mathcal{P}_n(I_k) \geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} \mathcal{P}_n(I'_k),$$

da $\mathcal{P}_n(I'_k) = F_n(b'_k) - F_n(a'_k) \overset{(e)}{\to} F(b'_k) - F(a'_k) = \mathcal{P}(I'_k)$. Also:

$$\liminf_{n\to\infty} \mathcal{P}_n(A) \geq \sum_{k=1}^{\infty} \mathcal{P}(I'_k) \geq \sum_{k=1}^{\infty} \mathcal{P}(I'_k) \geq \sum_{k=1}^{\infty} \mathcal{P}(I_k) - \varepsilon \sum_{k=1}^{\infty} 2^{-k} \overset{\sigma\text{-add.}}{=} \mathcal{P}(A) - \varepsilon$$

Mit $\varepsilon \searrow 0$ folgt (c). □

Zum Beweis des zentralen Grenzwertsatzes verwenden wir die Methode der charakteristischen Funktionen:

4.22 Definition. (i) Sei $\mathcal{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^D, \mathcal{B}_D)$. Dann heißt die Funktion

$$\varphi(t) := \int_{\mathbb{R}^D} e^{i\langle t,x\rangle} \mathcal{P}(dx) \quad , t \in \mathbb{R},$$

die *charakteristische Funktion von* $\mathcal{P}$. Hier ist $\langle \cdot, \cdot \rangle$ das euklidische Skalarprodukt in $\mathbb{R}^D$, $i = \sqrt{-1}$ die imaginäre Einheit und

$$\int_{\mathbb{R}^D} e^{i\langle t,x\rangle} \mathcal{P}(dx) = \int \Re(e^{i\langle t,x\rangle})\mathcal{P}(dx) + i \int \Im(e^{i\langle t,x\rangle})\mathcal{P}(dx) = \int \cos(\langle t,x\rangle)\mathcal{P}(dx) + i \int \sin(\langle t,x\rangle)\mathcal{P}(dx).$$

(ii) Für eine $\mathbb{R}^D$-wertige Zufallsvariable X ist

$$\varphi_X(t) := E[e^{i\langle t,x\rangle}] = \int_{\mathbb{R}^D} e^{i\langle t,x\rangle} d\mathcal{P}_X \quad , t \in \mathbb{R}.$$

4.23 Beispiel. (i) Ist $X \sim$Bernoulli(p), so ist $\varphi_X(t) = e^{it}p + (1-p)$, $t \in \mathbb{R}$.

(ii) Sei $X \sim \mathcal{N}(0,1)$. Aus der Übung ist bekannt, dass $\forall n \in \mathbb{N}$: $E[X^{2n}] = \frac{1}{2^n}\frac{(2n)!}{n!}$, $E[X^{2n-1}] = 0$. Also:

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{2tx} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(itx)^k}{k!} e^{-\frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \sum\nolimits_{k=0}^{\int} \frac{(it)^k}{k!} \int_{-\infty}^{\infty} x^k e^{-\frac{x^2}{2}} dx \qquad \text{(mit majorisierter Konvergenz)} \\
&= \sum_{k=0}^{\infty} \frac{(it)^{2k}}{(2k)!} \frac{1}{2^k} \frac{(2k)!}{k!} = \sum_{k=0}^{\infty} \left(\frac{(it)^2}{2}\right)^k \frac{1}{k!} = e^{-\frac{-t^2}{2}}
\end{aligned}$$

Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist $Y = \frac{1}{\sigma}(X_\mu) \sim \mathcal{N}(0,1)$. Also ist $\varphi_X = E[e^{it(\sigma Y + \mu)}] =^{it\mu - \frac{\sigma^2 t^2}{2}}$.

(iii) Sei $X \sim Poiss(\lambda)$, dann:

$$\varphi_X(t) = \sum_{k=0}^{\infty} e^{itk} \cdot \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^{it})^k}{k!} = e^{\lambda(e^{it}-1)}$$

4.24 Lemma. Sei $X \in L^2$. Dann gilt

$$\varphi_X(t) = 1 + itE[X] - \frac{1}{2}t^2 E[X^2] - \frac{1}{2}t^2 \varepsilon(t)$$

mit $|\varepsilon(t)| \leq 3E[X^2]$ und $\lim\limits_{t \to 0} \varepsilon(t) = 0$.

Beweis. Da $\varphi_X(t) = E[e^{itX}] = E[\cos(tx)] + iE[\sin(tx)]$ existieren mit Taylorentwicklung von cos bzw. sin um 0 $\vartheta_1(\omega), \vartheta_2(\omega)$ mit $|\vartheta_1(\omega)| \leq 1, |\vartheta_2(\omega) \leq 1$ und

$$e^{itX(\omega)} = 1 + itX(\omega) - \frac{1}{2}[\cos(\vartheta_1(\omega)tX(\omega)) + i\sin(\vartheta_2(\omega)tX(\omega))]X^2(\omega).$$

Sei $\varepsilon(t) = E[x^2(\cos(\vartheta_1 tx) + i\sin(\vartheta_2 tx) - 1)]$, dann:

$$\varphi_X(t) = 1 + itE[X] - \frac{1}{2}t^2 E[X^2] - \frac{1}{2}t^2 \varepsilon(t)$$

Offensichtlich gilt $|\varepsilon(t)| \leq 3E[X^2]$. Schließlich zeigt der Satz von Lebesgue, dass $\varepsilon(t) \to 0$, da $(0)(\vartheta_1 tx) + i\sin(\vartheta_2 tx) - 1 \to 0$ $\mathcal{P}$-fast sicher. $\square$

4.25 Satz. Sei $(X_n)_{n\in\mathbb{N}}$ unabhängige, identisch verteilte Folge von Zufallsvariablen in L^2 mit $E[X_1] = \mu$ und $Var(X_1) = \sigma^2$. Sei $\Gamma_n = \frac{1}{\sqrt{n\sigma^2}} \sum\limits_{k=1}^{n} (X_k - \mu)$. Dann gilt für alle $t \in \mathbb{R}$:

$$\varphi_{\Gamma_n}(t) \to e^{-\frac{t^2}{2}} \quad (n \to \infty)$$

Beweis. Es gilt:

$$\begin{aligned}
\varphi_{\Gamma_n}(t) &= E\left[e^{it\frac{1}{\sqrt{n}}\sum\limits_{k=1}^{n}\frac{X_k-\mu}{\sigma}}\right] \\
&= E\left[\prod_{k=1}^{n} e^{i\frac{t}{\sqrt{n}}\frac{X_k-\mu}{\sigma}}\right] \overset{\text{Unabh.}}{=} \prod_{k=1}^{n} E\left[e^{i\frac{t}{\sqrt{n}}\frac{X_k-\mu}{\sigma}}\right] \\
&= E\left[e^{i\frac{t}{\sqrt{n}}\frac{X_1-\mu}{\sigma}}\right]^n = \left(\varphi_{\frac{X_1-\mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right)\right)^n
\end{aligned}$$

Aus Lemma 4.24 folgt mit $X = \frac{X_1-\mu}{\sigma}$:

$$\varphi_{\frac{X_1-\mu}{\sigma}}(\frac{t}{\sqrt{n}}) = 1 + i\frac{t}{\sqrt{n}}\underbrace{E\left[\frac{X_1-\mu}{\sigma}\right]}_{=0} - \frac{1}{2}\frac{t^2}{n}\underbrace{E\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]}_{=1} - \frac{1}{2}\frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)$$

Also folgt für alle $t \in \mathbb{R}$:

$$\varphi_{\frac{X_1-\mu}{\sigma}}(\frac{t}{\sqrt{n}}) = 1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{1}{n}\right)$$

Also:

$$\varphi_{\Gamma_n}(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{(1+\frac{1}{n})^n \to e} e^{-\frac{1}{2}t^2}$$

□

Somit folgt der zentrale Grenzwertsatz aus folgendem wichtigen Zusammenhang zwischen schwacher Konvergenz und charakteristischen Funktionen:

4.26 Satz. (Stetigkeitssatz)
Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ und seien $(\varphi_n)_{n\in\mathbb{N}}$ die zugehörigen charakteristischen Funktionen. Dann gilt:

(i) Konvergiert $(\mathcal{P}_n)_{n\in\mathbb{N}}$ schwach gegen ein Wahrscheinlichkeitsmaß $\mathcal{P}$, so gilt für alle $t \in \mathbb{R}$:

$$\varphi_n(t) \to \varphi_{\mathcal{P}}(t)$$

(ii) Konvergiert $(\varphi_n)_{n\in\mathbb{N}}$ für alle $t \in \mathbb{R}$ gegen $\varphi(t)$ und ist φ stetig in 0, so ist φ die charakteristische Funktion eines Wahrscheinlichkeitsmaßes $\mathcal{P}$ und $(\mathcal{P}_n)_{n\in\mathbb{N}}$ konvergiert schwach gegen $\mathcal{P}$.

Beweis. (Beweis von Satz 4.17 (ZGWS) mittels 4.26)
Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ die Folge von Verteilungsfunktionen von $\frac{1}{\sigma\sqrt{n}}\sum\limits_{k=1}^{n}(X_k-\mu)$ und $(\varphi_n)_{n\in\mathbb{N}}$ die zugehörige Folge der charakteristischen Funktionen. Nach 4.25 und Beispiel 4.23 konvergiert $(\varphi_n)_{n\in\mathbb{N}}$ punktweise gegen die charakteristische Funktion von $\mathcal{N}(0,1)$. Satz 4.26(ii) impliziert, dass $\mathcal{P}_n \to \mathcal{N}(0,1)$ schwach, was nach dem Portmanteau-Theorem die wesentliche Konvergenz der Verteilungen nach sich zieht. □

Beweis. (von 4.26(i))
$(\mathcal{P}_n)_{n\in\mathbb{N}}$ konvergiere schwach gegen $\mathcal{P}$. Da $\Re(e^{itx}) = \cos(tx)$ und $\Im(e^{itx}) = \sin(tx)$ für festes $t \in (R)$ als Funktionen in x stetig und beschränkt sind, folgt:

$$\varphi_n(t) = \int \cos(tx)\mathcal{P}_n(dx) + i\int \sin(tx)\mathcal{P}_n(dx) \to \int \cos(tx)\mathcal{P}(dx) + i\int \sin(tx)\mathcal{P}(dx) = \varphi_{\mathcal{P}}(t)$$

□

<u>Beweisstrategie von 4.26(ii):</u>

(1) Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ mit zugehörigen Verteilungsfunktionen $(F_n)_{n\in\mathbb{N}}$ Dann liefert der *Satz von Helly* (4.38):
   - Es gibt eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ und eine Funktion $G : \mathbb{R} \to \mathbb{R}$ mit $F_{n_k}(x) \overset{k\to\infty}{\longrightarrow} G(x)$ für alle Stetigkeitsstellen x von G.
   - G ist rechtsstetig, nichtfallend, aber i.A. gilt nur, dass $\lim\limits_{x\to-\infty} G(x) \geq 0$ und $\lim\limits_{x\to\infty} \leq 1$.

(2) Die Konvergenz der charakteristischen Funktionen stellt sicher, dass die Folge $(\mathcal{P}_n)_{n\in\mathbb{N}}$ "´straff“ ist (Lemma 4.34), so dass G tatsächlich eine Verteilungsfunktion ist (Satz von Prokorov).

(3) Aus einem Eindeutigkeitssatz für charakteristische Funktionen (4.28) folgt dann, dass die Konvergenz nicht nur entlang der Teilfolge $(F_{n_k})_{k\in\mathbb{N}}$, sondern auch entlang der Folge $(F_n)_{n\in\mathbb{N}}$ gilt.

(4) Das Portmanteau-Theorem erlaubt es dann, von der wesentlichen Konvergenz der Verteilungen auf die schwache Konvergenz der Wahrscheinlichkeitsmaße zu schließen.

4.4 Mehr zu charakteristischen Funktionen

4.27 Satz. (Umkehrformel)
Sei $\mathcal{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit Verteilung F und charakteristischer Funktion φ. Dann gilt:

(i) Für alle Stetigkeitsstellen $a < b$ von F gilt:

$$F(b) - F(a) = \lim_{c\to\infty} \frac{1}{2\pi} \int\limits_{[-c,c]} \frac{e^{ita} - e^{-itb}}{it} \varphi(t)\, d\lambda$$

(ii) Falls φ λ-integrierbar ist, d.h. $\int\limits_{\mathbb{R}} |\varphi|\, d\lambda < \infty$, dann hat F eine Dichte f bezüglich λ, welche gegeben ist durch

$$f(x) = \frac{1}{2\pi} \int\limits_{\mathbb{R}} e^{-itx} \varphi(t)\, d\lambda.$$

Beweis. (i) Sei $a < b$ Stetigkeitsstellen von F. Mit dem Satz von Fubini folgt:

$$\Phi_c := \frac{1}{2\pi} \int\limits_{-c}^{c} \frac{e^{-ita} - e^{itb}}{it} \varphi(t) \lambda(dt) = \int\limits_{\mathbb{R}} \underbrace{\int\limits_{-c}^{+c} \frac{1}{2\pi} \frac{e^{-ita} - e^{itb}}{it} e^{itx} \lambda(dt)}_{\psi_c(x)} \mathcal{P}(dx)$$

Da $\frac{\cos(yt)}{t}$ punktsymmetrisch als Funktion in t ist, folgt

$$\psi_c(x) = \frac{1}{2\pi} \int\limits_{-c(x-a)}^{+c(x-a)} \frac{\sin(t(x-a)) - \sin(t(x-b))}{t} \lambda(dt) = \frac{1}{2\pi} \int\limits_{-c(x-a)}^{+c(x-a)} \frac{\sin(v)}{v} dv - \frac{1}{2\pi} \int\limits_{-c(x-b)}^{+c(x-b)} \frac{\sin(v)}{v} dv.$$

Die Funktion $t \mapsto g(t) := \int\limits_{-t}^{t} \frac{\sin(v)}{dv} dv$ ist beschränkt mit $\lim\limits_{t\to\infty} g(t) = \pi$ (Königsberger Analysis I). Also:

$$\lim_{c\to\infty} \psi_c(x) = \left\{ \begin{array}{ll} 0 & , x < a \text{ oder } x > b \\ 1 & , x \in (a,b) \\ \frac{1}{2} & , x \in \{a,b\} \end{array} \right\} =: \psi(x)$$

Somit liefert der Satz über dominierte Konvergenz:

$$\Phi_c = \int\limits_{-\infty}^{\infty} \psi_c(x) \mathcal{P}(dx) \to \int\limits_{-\infty}^{\infty} \psi(x) \mathcal{P}(dx) = \mathcal{P}((a,b]) + \underbrace{\frac{1}{2}\mathcal{P}(\{a\}) - \frac{1}{2}\mathcal{P}(\{b\})}_{=0,\text{ da } a,b \text{ Stetigkeitsstellen}} = F(b) - F(a)$$

(ii)

$$\begin{aligned}
\int\limits_a^b f(x)\lambda(dx) &= \int\limits_a^b \left(\int\limits_{-\infty}^{\infty} \frac{1}{2\pi} e^{-itx} \varphi(t) \lambda(dt) \right) \lambda(dx) \\
&\overset{\text{Fubini}}{=} \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \varphi(t) \left(\int\limits_a^b e^{-itx} dx \right) dt \\
&= \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \varphi(t) \frac{1}{it} (e^{-ita} - e^{-itb}) dt \\
&= \lim_{c\to\infty} \frac{1}{2\pi} \int\limits_{-c}^{c} \varphi(t) (e^{-ita} - e^{itb}) \frac{1}{it} \lambda(dt) \\
&\overset{\text{(i)}}{=} F(b) - F(a) \text{ für alle Stetigkeitsstellen } a < b \text{ von } F
\end{aligned}$$

Somit $F(x) = \int\limits_{-\infty}^{x} f(y)\lambda(dy)$ für alle $x \in \mathbb{R}$. $\square$

4.28 Satz. (Eindeutigkeitssatz)
Seien $\mathcal{P}$ und $\mathcal{Q}$ Wahrscheinlichkeitsmaße auf $(\mathbb{R}, \mathcal{B})$, die die gleiche charakteristische Funktion habn. Dann gilt:

$$\mathcal{P}(A) = \mathcal{Q}(A) \text{ für alle } A \in \mathcal{B}$$

Beweis. Nach Satz 4.27 gilt $\mathcal{Q}((a,b]) = \mathcal{P}((a,b])$ für alle außer höchstens abzählbar viele $a < b \in \mathbb{R}$. Aufgrund der Stetigkeit der Wahrscheinlichkeitsmaße $\mathcal{P}$ und $\mathcal{Q}$ von oben und von unten, gilt die Gleichheit für alle $a < b \in \mathbb{R}$. Der Eindeutigkeitsteil des Satzes von Carathéodory liefert dann die Behauptung. □

Bemerkung. Der Eindeutigkeitssatz gilt auch für Wahrscheinlichkeitsmaße auf $(\mathbb{R}^D, \mathcal{B}_D)$. Daraus folgt dann:

4.29 Satz. Ein Vektor $X = (X_1, ..., X_D)$ von reellwertigen Zufallsvariablen ist genau dann unabhängig, falls für alle $t = (t_1, ..., t_D) \in \mathbb{R}D$ gilt: $\varphi_X(t) = \prod\limits_{d=1}^{D} \varphi_{X_d}(t_d) \quad (*)$

Beweis. Es gelte $(*)$. Sei $\mathcal{P}_X$ de Verteilung von X und $\mathcal{P}_{X_d}$ die Verteilung von X_d $(d = 1, ..., D)$. Dann gilt für alle $t \in \mathbb{R}^D$:

$$\int_{\mathbb{R}^D} e^{i\langle t,x\rangle} \mathcal{P}_X(dx) = \varphi_X(t) \overset{(*)}{=} \prod_{d=1}^{D} \varphi_{X_d}(t_d) = \prod_{d=1}^{D} \int_{\mathbb{R}} e^{it_d x_d} \mathcal{P}_{X_d}(dx_d) \overset{\text{Fubini}}{=} \int_{\mathbb{R}^D} e^{i\langle t,x\rangle} (\mathcal{P}_{X_1} \otimes ... \otimes \mathcal{P}_{X_d})(dx)$$

Der Eindeutigkeitssatz impliziert $\mathcal{P}_X = \mathcal{P}_{X_1} \otimes ... \otimes \mathcal{P}_{X_d}$. Also ist $(X_1, ..., X_D)$ unabhängig. Die umgekehrte Richtung folgt direkt aus Satz 2.21. □

Charakteristische Funktionen sind nützlich zur Berechnung von Momenten:

4.30 Satz. Sei $E[|X|^n] < \infty$ für ein $n \in \mathbb{N}$. Dann existiert für alle $r \leq n$ die r-te Ableitung der charakteristischen Funktion φ von X und

$$\varphi^{(r)}(t) = \int_{\mathbb{R}} (ix)^r e^{itx} \mathcal{P}_X(dx).$$

Insbesondere: $\frac{1}{i^r}\varphi^{(r)}(0) = E[X^r]$

Beweis. $r = 1$: $\frac{\varphi(t+h)-\varphi(t)}{h} = E\left[e^{itX}\left(\frac{e^{ihX}-1}{h}\right)\right]$
Da $\left|\frac{e^{ihX}-1}{h}\right| \leq |X|$ und $E[|X|] \leq E[|X|^n]^{\frac{1}{n}}$, liefert der Satz über dominierte Konvergenz:

$$\frac{d}{dt}\varphi(t) = E\left[e^{itX}\left(\lim_{n\to 0} \frac{e^{ihX}-1}{h}\right)\right] = E[e^{itX} iX] = \int ixe^{itx} \mathcal{P}_X(dx)$$

Die höheren Ableitungen folgen induktiv auf analoge Weise. □

4.31 Beispiel. Sei X Poisson (λ)-verteilt. Dann:

$$\begin{aligned}
\varphi_X(t) &= \exp\{\lambda(e^{it} - 1)\} \text{ (siehe Beipsiel 4.23)} \\
\varphi'_X(t) &= \lambda i e^{it} \varphi_X(t) \\
\varphi''_X(t) &= (\lambda i e^{it})^2 \varphi_X(t) + \lambda i^2 e^{it} \varphi_X(t)
\end{aligned}$$

Also $E[X] = \lambda$ und $E[X^2] = \lambda^2 + \lambda$ und somit $Var(X) = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Aus Satz 4.30 und Taylorentwicklung von 0 folgt als Verallgemeinerung von Lemma 4.24:

4.32 Folgerung. Sei $E[|X|^n] < \infty$ für ein $n \in \mathbb{N}$. Dann gilt $\varphi_X(t) = \sum\limits_{r=0}^{n} \frac{(it)^r}{r!} E[X^r] + \theta_{n,X}(t)$ mit $\theta_{n,X}(t) = \int\limits_0^t \frac{(t-u)^{n-1}}{(n-1)!}(\varphi_X^{(n)}(u) - \varphi_X^{(n)}(0))du$, also $|\theta_{n,X}(t)| \leq \frac{t^n}{n!} \underbrace{\sup_{\vartheta\in[0,1]} |\varphi_X(\vartheta t) - \varphi_X(0)|}_{\to 0 \quad (n\to 0)}$.

4.5 Straffheit

Fragestellung: Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathcal{B})$. Wann existiert ein Wahrscheinlichkeitsmaß $\mathcal{P}$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ mit $\mathcal{P}_{n_k}\to\mathcal{P}$ schwach $(k\to\infty)$?
Dies führt zu folgender Definition:

4.33 Definition. Eine Menge $\mathbb{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathcal{B})$ wird *relativ kompakt* genannt, falls jede Folge von Wahrscheinlichkeitsmaßen in $\mathbb{P}$ eine Teilfolge enthält, die schwach gegen ein Wahrscheinlichkeitsmaß konvergiert.

4.34 Beispiele. (i) Sei $\mathbb{P}=\mathcal{P}_1,...,\mathcal{P}_n$ endlich, so ist $\mathbb{P}$ relativ kompakt, denn jede Folge in $\mathbb{P}$ muss eine konstante Teilfolge beinhalten.

(ii) Sei $\mathbb{P}=\{\mathcal{P}_n\,|\,n\in\mathbb{N}\}$ mit

$$\mathcal{P}_n(A):=\begin{cases}1 & ,n\in A\\ 0 & \text{, sonst.}\end{cases}$$

Wir betrachten die Folge $(\mathcal{P}_n)_{n\in\mathbb{N}}$. Sei $\mathcal{P}_{n_k}$ eine konvergente Teilfolge mit $\mathcal{P}_{n_k}\to\mathcal{P}$ schwach. Dann muss nach dem Portmanteau-Theorem gelten $\mathcal{P}_{n_k}((-\infty,x])\to\mathcal{P}((-\infty,x])$ für alle $x\in\mathbb{R}\setminus D$, wobei D höchstens abzählbar ist. Da $\mathcal{P}_{n_k}((-\infty,x])=0$ für $n_k>x$, folgt $\mathcal{P}((-\infty,x])=0$ für alle $x\in\mathbb{R}$ und somit $\mathcal{P}(\mathbb{R})=0$ im Widerspruch zu $\mathcal{P}(\mathbb{R})=1$. Also ist $\mathbb{P}$ nicht relativ kompakt. Beim Grenzübergang „verschwindet“ hier die gesamte Masse in $+\infty$.

4.35 Definition. Eine Menge $\mathbb{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathcal{B})$ wird *straff* genannt, falls $\forall\varepsilon>0$ eine kompakte Menge $K\subset\mathbb{R}$ existiert mit

$$\sup_{\mathcal{P}\in\mathbb{P}}\mathcal{P}(\mathbb{R}\setminus K)\leq\varepsilon.$$

4.36 Lemma. Jede relativ kompakte Menge $\mathbb{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathcal{B})$ ist straff.

Beweis. Sei $\mathbb{P}$ relativ kompakt, aber nicht straff. Dann existiert ein $\varepsilon>0$, so dass für alle kompakten Mengen $K\subset\mathbb{R}$: $\sup_{\mathcal{P}\in\mathbb{P}}\mathcal{P}(\mathbb{R}\setminus K)>\varepsilon$. Somit gilt für alle $n\in\mathbb{N}$

$$\sup_{\mathcal{P}\in\mathbb{P}}\mathcal{P}(\mathbb{R}\setminus]-n,n[)\geq\sup_{\mathcal{P}\in\mathbb{P}}(\mathbb{R}\setminus[-n,n])>\varepsilon.$$

Also existiert für alle $n\in\mathbb{N}$ ein Wahrscheinlichkeitsmaß $\mathcal{P}_n\in\mathbb{P}$ mit $\mathcal{P}_n(\mathbb{R}\setminus]-n,n[)\searrow\varepsilon$. Da $\mathbb{P}$ relativ kompakt ist, existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ und ein Wahrscheinlichkeitsmaß $\mathcal{P}$ mit $\mathcal{P}_{n_k}\to\mathcal{P}$ schwach. Das Portmanteau-Theorem impliziert für beliebiges $n\in\mathbb{N}$:

$$\varepsilon\leq\limsup_{k\to\infty}\mathcal{P}_{n_k}(\mathbb{R}\setminus]-n_k,n_k[)\leq\limsup_{k\to\infty}\mathcal{P}_{n_k}(\mathbb{R}\setminus]-n,n[)\overset{\text{Portmanteau}}{\leq}\mathcal{P}(\mathbb{R}\setminus]-n,n[)$$

Für $n\to\infty$ folgt $\varepsilon\leq\mathcal{P}(\emptyset)=0$. $\square$

Zum Lemma 4.36 gilt auch die Umkehrung:

4.37 Satz. (Prokorov)
Eine Familie $\mathbb{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R},\mathcal{B})$ ist genau dann relativ kompakt, wenn $\mathbb{P}$ straff ist.

Zur Vorbereitung des Beweises:
Es beizeichne $\mathcal{J}$ die Meenge der *verallgemeinerten Verteilungsfunktionen*, d.h. $G\in\mathcal{J}$ genau dann, wenn:

- $G:\mathbb{R}\to\mathbb{R}$ ist rechtsstetig
- G ist nichtfallend
- $0\leq G(-\infty)\leq G(\infty)\leq 1$

4.38 Satz. (Helly)
Die Menge $\mathcal{J}$ ist schwach folgenkomapkt, d.h. für jede Folge $(G_n)_{n\in\mathbb{N}}\subset\mathcal{J}$ existiert ein $G\in\mathcal{J}$ und eine Teilfolge $(n_k)_{k\in\mathbb{N}}$, so dass $G_{n_k}(x)\to G(x)$ an allen Stetigkeitsstellen x von G.

Beweis. Sei $T=\{x_1,x_2,...\}$ eine abzählbare dichte Teilmenge von $\mathbb{R}$. Da $(G_n(x_1))_{n\in\mathbb{N}}$ eine Folge in $[0,1]$ ist, existiert eine Teilfolge $N^{(1)}=(n_1^{(1)},n_2^{(1)},...)$ und ein $g_1\in[0,1]$ mit $(G_{n_k}^{(1)})(x_1)\to g_1,k\to\infty$. Sukzessive konstruiert man für alle $m\in\mathbb{N}$ Teilfolgen $N^{(m+1)}$ von $N^{(m)}$ und $g_{m+1}\in[0,1]$, sodass $G_{n_k^{(m+1)}}(x_{m+1})=g_{m+1}$ $(k\to\infty)$. Damit definiert man auf T die Funktion $G_T(x_m):=g_m$ $(x_m\in T)$. Sei N die Diagonalfolge $N=(n_1^{(1)},n_2^{(2)},...)$. Dann gilt für alle $m\in\mathbb{N}$:

$$G_{n_j^{(j)}}(x_m)\to g_m=G_T(x_m)\quad(j\to\infty)\quad(*)$$

Definiere $G(x)=\inf\{G_T(y)\,|\,y\in T\text{ und }y>x\}$. Wir zeigen:

(i) $G \in \mathcal{J}$

(ii) $G_{n_j^{(j)}}(x) \overset{j\to\infty}{\to} G(x)$ an allen Stetigkeitsstellen x von G

zu (i) G ist nichtfallend nach Konstruktion und $0 \leq G(-\infty) \leq G(\infty) \leq 1$, da $G_T(y) \in [0,1]$ für alle $y \in T$. Bleibt die Rechtsstetigkeit zu zeigen. Sei x_k eine Folge in $\mathbb{R}$ mit $x_k \searrow x$. Wegen der Monotonie von G existiert $g := \lim_{k\to\infty} G(X_k)$ und $G(x) \leq g$. Wir nehmen an, dass $G(x) < g$. Dann existiert nach Definition von G ein $y \in T$ mit $y > x$ und $G_T(y) < g$. Da $x < x_k < y$ für hinreichend großes $k \in \mathbb{N}$, folgt $g = \lim_{k\to\infty} G(x_k) \leq G_T(y) < g$. Also muss $G(x) = g$ gelten.

zu (ii) Sei z_0 eine Stetigkeitsstellen von G. Für $z_0 < y$, $y \in T$, gilt nach $(*)$:

$$\limsup_{j\to\infty} G_{n_j^{(j)}}(z_0) \leq \limsup_{j\to\infty} G_{n_j^{(j)}}(y) = G_T(y)$$

Also: $\limsup_{j\to\infty} G_{n_j^{(j)}}(z_0) \leq \inf\{G_T(y) \,|\, y \in T \text{ mit } y > z_0\}$.
Sei andererseits $z_1 \in \mathbb{R}$ und $\tilde{y} \in T$ mit $z_1 < \tilde{y} < z_0$. Dann:

$$G(z_1) \leq G_T(\tilde{y}) \overset{*}{=} \lim_{j\to\infty} G_{n_j^{(j)}}(\tilde{y}) \leq \liminf_{j\to\infty} G_{n_j^{(j)}}(z_0)$$

Da G stetig in z_0 ist, folgt mit $z_1 \nearrow z_0$: $G(z_0) \leq \liminf_{j\to\infty} G_{n_j^{(j)}}(z_0)$
Zusammen: $\lim_{j\emptyset\infty} G_{n_j^{(j)}}(z_0) = G(z_0)$ □

Beweis. (von Satz 4.37)
Sei $\mathbb{P}$ straff. Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{P}$. Sei $(F_n)_{n\in\mathbb{N}}$ die zugehörige Folge von Verteilungsfunktionen. Nach dem Satz von Helly existiert eine Teilfolge $(n_k)_{k\in\mathbb{N}}$ und eine verallgemeinerte Verteilungsfunktion $G \in \mathcal{J}$ mit $F_{n_k}(x)$ punktweise konvergent gegen $G(x)$ an allen Stetigkeitsstellen x von G. Aufgrund des Portmanteau-Theorems reicht es zu zeigen, dass G ein echte Verteilungsfunktion ist, d.h. dass $G(-\infty) = 0$ und $G(\infty) = 1$ gilt.
Sei $\varepsilon > 0$. Aufgrund der Straffheit existiert eine kompakte Menge K mit $\mathcal{P}_{n_k}(K) \geq 1 - \varepsilon\ \forall k \in \mathbb{N}$. Dann gilt für alle Stetigkeitsstellen $a < b$ von G mit $K \subset (a,b]$:

$$1 - \varepsilon \leq \mathcal{P}_{n_k}(K) \leq \mathcal{P}_{n_k}((a,b]) = F_{n_k}(b) - F_{n_k}(a) \overset{k\to\infty}{\to} G(b) - G(a)$$

Also $1 - \varepsilon \leq G(\infty) - G(-\infty)$ und mit $\varepsilon \searrow 0$: $G(+\infty) - G(-\infty) = 1$. Da $G(-\infty) \geq 0$, $G(\infty) \leq 1$ folgt $G(\infty) = 1$ und $G(-\infty) = 0$.
Die andere Richtung wurde bereits in Lemma 4.36 gezeigt. □

4.39 Satz. Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge in einer straffen Menge $\mathbb{P}$ von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$. Jede schwach konvergente Teilfolge $(\mathcal{P}_{n_k})_{k\in\mathbb{N}}$ konvergiere gegen das selbe Wahrscheinlichkeitsmaß $\mathcal{P}$. Dann konvergiert schon die ganze Folge schwach gegen $\mathcal{P}$.

Beweis. Annahme: Die Folge $(\mathcal{P}_n)_{n\in\mathbb{N}}$ konvergiert nicht schwach gegen $\mathcal{P}$. Dann existiert eine beschränkte Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $\int f d\mathcal{P}_n \nrightarrow \int f d\mathcal{P}$. Es existieren also ein $\varepsilon > 0$ und eine Teilfolge $(n_k^{(1)})_{k\in\mathbb{N}}$ mit

$$\left| \int f d\mathcal{P}_{n_k^{(1)}} - \int f d\mathcal{P} \right| \geq \varepsilon \text{ für alle } k \in \mathbb{N} \quad (*).$$

Nach dem Satz von Prokorov hat $(\mathcal{P}_{n_k^{(1)}})_{k\in\mathbb{N}}$ eine schwach konvergente Teilfolge $(\mathcal{P}_{n_k^{(2)}})_{k\in\mathbb{N}}$, deren Grenzwert nach Annahme $\mathcal{P}$ sein muss. Also $\int f d\mathcal{P}_{n_k^{(2)}} \to \int f d\mathcal{P}$ im Widerspruch zu $(*)$. □

Beweis. (von Satz 4.26(ii))
Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$, so dass die zugehörige Folge $(\varphi_n)_{n\in\mathbb{N}}$ von charakteristischen Funktionen punktweise gegen eine in 0 stetige Funktion φ konvergiert.
Wir zeigen zunächst, dass $\{\mathcal{P}_n \,|\, n \in \mathbb{N}\}$ straff ist. Dazu sei $a > 0$. Da $\Re(\varphi_n(t)) = \int \cos(tx)\mathcal{P}_n(dx)$ folgt mit dem Satz von Fubini:

$$\begin{aligned}
\frac{1}{a}\int_0^a (1 - \Re(\varphi_n(t)))\lambda(dt) &= \int_{\mathbb{R}} \frac{1}{a}\left(\int_0^a 1 - \cos(tx)\lambda(dt\right)\mathcal{P}_n(dx) \\
&= \int_{\mathbb{R}} \left(1 - \frac{\sin ax}{ax}\right)\mathcal{P}_n(dx) \\
&\geq \inf_{|y|\geq 1}\left(1 - \frac{\sin y}{y}\right)\int \mathbb{1}_{\{|ax|\geq 1\}}\mathcal{P}_n(dx) \\
&\geq \frac{1}{7}\mathcal{P}_n(\{|X| \geq \frac{1}{a}\}),
\end{aligned}$$

da $\inf\limits_{|y|\geq 1}\left(1-\frac{\sin y}{y}\right)=1-\frac{\sin(1)}{1}\geq \frac{1}{7}$. Somit:

$$\mathcal{P}_n\left(\mathbb{R}\setminus\left(-\frac{1}{a},\frac{1}{a}\right)\right)\leq \frac{1}{7}\int\limits_0^a(1-\Re(\varphi_n(t)))\lambda(dt)\xrightarrow{n\to\infty}\frac{7}{a}\int\limits_0^a(1-\Re(\varphi(t)))\lambda(dt)$$

Da φ stetig in 0 ist, folgt

$$\lim_{a\searrow 0}\frac{7}{a}\int\limits_0^a(1-\Re(\varphi(t)))\lambda(dt)=7(1-\Re(\varphi(0)))=0,$$

da $\varphi_n(0)=\int 1d\mathcal{P}_n=1$ für alle $n\in\mathbb{N}$, also $\varphi(0)=1$. Für $\varepsilon>0$ existiert ein $a>0$ mit

$$\sup_{n\in\mathbb{N}}\mathcal{P}_n\left(\mathbb{R}\setminus\left[-\frac{1}{a},\frac{1}{a}\right]\right)\leq\sup_{n\in\mathbb{N}}\mathcal{P}_n\left(\mathbb{R}\setminus\left(-\frac{1}{a},\frac{1}{a}\right)\right)\leq\varepsilon.$$

Also ist $\{\mathcal{P}_n\,|\,n\in\mathbb{N}\}$ straff. Aufgrund des Satzes von Prokorov existiert eine gegen ein Wahrscheinlichkeitsmaß schwach konvergente Teilfolge von $(\mathcal{P}_n)_{n\in\mathbb{N}}$. Sei $(\mathcal{P}_{n_k})_{k\in\mathbb{N}}$ eine gegen ein Wahrscheinlichkeitsmaß $\mathcal{P}$ konvergente Teilfolge. Aufgrund von Satz 4.26(i) gilt für alle $t\in\mathbb{R}$ $\lim\limits_{k\to\infty}\varphi_{n_k}(t)=\varphi_{\mathcal{P}}(t)$, wobei $\varphi_{\mathcal{P}}$ die charakteristische Funktion von $\mathcal{P}$ ist. Insbesondere folgt, dass $\varphi(t)=\varphi_{\mathcal{P}}(t)$ für alle $t\in\mathbb{R}$, so dass die charakteristische Funktion aller schwachen Grenzwerte $\mathcal{P}$ von Teilfolgen von $(\mathcal{P}_n)_{n\in\mathbb{N}}$ gleich sind. Der Eindeutigkeitssatz 4.28 impliziert nun, dass alle schwach konvergenten Teilfolgen von $(\mathcal{P}_n)_{n\in\mathbb{N}}$ gegen dasselbe Wahrscheinlichkeitsmaß $\mathcal{P}$ konvergieren, dessen charakteristische Funktion φ ist. Nach Satz 4.39 konvergiert dann die Folge $(\mathcal{P}_n)_{n\in\mathbb{N}}$ selbst schwach gegen $\mathcal{P}$. □

4.6 Der zentrale Grenzwertsatz von Lindeberg-Feller

4.40 Definition. Eine Folge unabhängiger Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $L^2(\mathcal{P})$ mit $Var(X_n)>0$ *genügt dem zentralen Grenzwertsatz*, falls die Folge $(\mathcal{P}_{S_n})_{n\in\mathbb{N}}$ der Verteilungen von

$$S_n:=\frac{\sum\limits_{k=1}^n X_k-E[X_k]}{\sqrt{Var(X_1+...+X_n)}}$$

schwach gegen eine $\mathcal{N}(0,1)$-Verteilung konvergiert.

Notation:

$$\sigma_n^2:=Var(X_n)$$
$$s_n^2:=Var(X_1+...+X_n)=\sum_{k=1}^n Var(X_k)=\sum_{k=1}^n\sigma_k^2$$
$$\mu_n:=E[X_n]$$

Feller-Bedingung:

$$\text{(F)}\quad\lim_{n\to\infty}\max_{j=1,\dots,n}\frac{\sigma_j}{s_n}=0$$

Feller-Bedingung:

$$\text{(L)}\quad\forall\varepsilon>0:\ \lim_{n\to\infty}\underbrace{\frac{1}{s_n^2}\sum_{j=1}^k\int\limits_{\{|x-\mu|\geq\varepsilon s_n\}}(x-\mu_j)^2\mathcal{P}_{X_j}(dx)}_{L_n(\varepsilon)}=0$$

4.41 Satz. (Lindeberg/Feller)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $L^2(\mathcal{P})$ mit $Var(X_n)>0$. Dann sind äquivalent:

(i) $(X_n)_{n\in\mathbb{N}}$ erfüllt die Lindeberg-Bedingung (L).

(ii) $(X_n)_{n\in\mathbb{N}}$ genügt dem zentralen Grenzwertsatz und der Feller-Bedingung (F).

4.42 Beispiel. (i) Sind die $(X_n)_{n\in\mathbb{N}}$ unter den Voraussetzungen von Satz 4.41 identisch verteilt, so folgt

$$L_n(\varepsilon)=\frac{1}{\sigma^2}\int\limits_{\{|x-\mu|\geq\varepsilon\sigma\sqrt{n}\}}(x-\mu)^2\mathcal{P}_{X_1}(dx)\to 0\quad(n\to\infty)$$

mit $\mu:=\mu_1$ und $\sigma^2=\sigma_1^2$. Der Satz von Lindeberg-Feller beinhaltet also den zentralen Grenzwertsatz von Linderbeg/Levy als Spezialfall.

(ii) Exisitert eine konstante $K < \infty$ mit $|X_n| \leq K$ $\mathcal{P}$-fast sicher für alle $n \in \mathbb{N}$ und falls $s_n^2 \to \infty$ für $n \to \infty$, so erfüllt $(X_n)_{n\in\mathbb{N}}$ die Lindeberg-Bedingung. Denn zunächst erfüllt $(X_n)_{n\in\mathbb{N}}$ die *Lyapunoff-Bedingung*:

$$(\Lambda) \quad \exists \delta > 0 : \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n E[|X_j - \mu_j|^{2+\delta}] = 0$$

Denn:

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^n E[|X_j - \mu_j|^{2+\delta}] \leq \frac{(2k)^\delta}{s_n^{2+\delta}} \underbrace{\sum_{j=1}^n E[(X_j - \mu_j)^2]}_{s_n^2} = \frac{(2k)^\delta}{s_n^\delta} \to 0 \quad (n \to \infty)$$

Die Lyapunoff-Bedingung impliziert aber die Lindeberg-Bedingung, denn:

$$|x - \mu_j| \geq \varepsilon s_n \Rightarrow |x - \mu_j|^{2+\delta} \geq |x - \mu_j|(\varepsilon s_n)^\delta$$

Also:

$$L_n(\varepsilon) \leq \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n \int_{\mathbb{R}} |x - \mu_j|^{2+\delta} \mathcal{P}_{X_j}(dx) = \frac{1}{\varepsilon^\delta} \frac{1}{s_n^{2+\delta}} E[|X_j - \mu_j|^{2+\delta}] \xrightarrow{(\Lambda)} 0 \quad (n \to \infty)$$

Beweis. (von Satz 4.41)
Wir zeigen nur (i)$\Rightarrow$(ii). Für (ii)$\Rightarrow$(i) siehe zum Beispiel Bauer.

Œ sei $E[X_n] = 0$ für alle $n \in \mathbb{N}$ Sei $Y_{n,j} := \frac{X_j}{s_n}$. Dann gilt $\sum\limits_{j=1}^n Y_{n,j} = \frac{\sum_{j=1}^n X_j}{\sqrt{Var(X_1+...+X_n)}}$.
Aufgrund des Stetigkeitssatzes reicht es zu zeigen, dass für alle $t \in \mathbb{R}$ $\varphi_{\sum_{j=1}^n Y_{n,j}}(t)$ für $n \to \infty$ gegen die charakteristische Funktion der $\mathcal{N}(0,1)$-Verteilung konvergiert. Es gilt:

$$\varphi_{\sum_{j=1}^n Y_{n,j}}(t) \overset{\text{Unabh.}}{=} \prod_{j=1}^n \varphi_{Y_{n,j}}(t) = \prod_{j=1}^n \varphi_{X_j}(\frac{t}{s_n})$$

Sei Z $\mathcal{N}(0,1)$-verteilt und Z_j $\mathcal{N}(0,\sigma_j^2)$-verteilt. Dann:

$$\varphi_Z(t) \overset{\text{Bsp. 4.23}}{=} e^{-\frac{t^2}{2}} = e^{-\frac{1}{s_n^2}\sum_{j=1}^n \sigma_j^2 \frac{t^2}{2}} = \prod_{j=1}^n e^{-\sigma^2 \frac{(\frac{t}{s_n})^2}{2}} = \prod_{j=1}^n \varphi_{Z_j}(\frac{t}{s_n})$$

Seien $a_1, ..., a_n$ und $b_1, ..., b_n$ komplexe Zahlen vom Betrag kleiner gleich 1. Dann:

$$\left|\prod_{v=1}^n a_v - \prod_{v=1}^n b_v\right| = |(a_1 - b_1)a_2 \cdots a_n + b_1(a_2 - b_2)a_3 \cdots a_n + ... + b_1 \cdots b_{n-1}(a_n - b_n)| \leq \sum_{v=1}^n |a_v - b_v|$$

Also:

$$\left|\varphi_Z(t) - \varphi_{\sum_{j=1}^n Y_{n,j}}(t)\right| \leq \left|\prod_{j=1}^n \varphi_{Z_j}\left(\frac{t}{s_n}\right) - \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{s_n}\right)\right| \leq \sum_{j=1}^n \left|\varphi_{Z_j}\left(\frac{t}{s_n}\right) - \varphi_{X_j}\left(\frac{t}{s_n}\right)\right| = (*)$$

Da $E[X_j] = E[Z_j] = 0$ und $E[X_j^2] = E[Z_j^2] = \sigma^2$, liefert Folgerung 4.32:

$$(*) \leq \frac{t^2}{2} \frac{1}{s_n^2} \sum_{j=1}^n \sup_{0\leq\theta\leq 1} \left|\varphi''_{X_j}(\theta \frac{t}{s_n}) - \varphi''_{X_j}(0)\right| + \sup_{0\leq\theta\leq 1} \left|\varphi''_{Z_j}(\theta \frac{t}{s_n}) - \varphi''_{Z_j}(0)\right| = (**)$$

Nach Satz 4.30 gilt:

$$\left|\varphi''_{X_j}(\theta \frac{t}{s_n}) - \varphi''_{X_j}(0)\right| \leq E[X_j^2 |e^{-i\theta \frac{t}{s_n} X_j} - 1] = (***)$$

Für festes $t \in \mathbb{R}$ sei $\delta > 0$. Dann existiert ein $\varepsilon > 0$ mit $|e^{i\theta t x} - 1| \leq \delta$ für alle $|x| \leq \varepsilon$ und $\theta \in [0,1]$. Also:

$$(***) \leq \delta E[X_j^2] + \int_{\{|x|\geq s_n\varepsilon\}} x^2 |e^{i\theta \frac{t}{s_n} x} - 1| \mathcal{P}_{X_j}(t) \leq \delta\sigma_j^2 + 2 \int_{\{|x|\geq s_n\varepsilon\}} x^2 \mathcal{P}_{X_j}(dx)$$

Die gleiche Abschätzung gilt für Z_j statt X_j. Also folgt mit $(**)$ und $(***)$:

$$\left|\varphi_Z(t)-\varphi_{\sum\limits_{j=1}^n Y_{n,j}}(t)\right| \leq t^2\left(\delta+\underbrace{\frac{1}{s_n^2}\sum_{j=1}^n\int\limits_{\{|x|\geq s_n\varepsilon\}} x^2\mathcal{P}_{X_j}(dx)}_{\mathrm{I}_n}+\underbrace{\frac{1}{s_n^2}\sum_{j=1}^n\int\limits_{\{|x|\geq s_n\varepsilon\}} x^2\mathcal{P}_{Z_j}(dx)}_{\mathrm{II}_n}\right)$$

Nun gilt $(\mathrm{I}_n)\to 0$ $(n\to\infty)$ wegen der Lindeberg-Bedingung. Da $\delta>0$ beliebig ist, folgt die Gültigkeit des zentralen Grenzwertsatzes für $(X_n)_{n\in\mathbb{N}}$, sofern

$$\mathrm{II}_n=\frac{1}{s_n^2}\sum_{j=1}^n\int\limits_{\{|x|\geq s_n\varepsilon\}} x^2\mathcal{P}_{Z_j}(dx)\to 0 \quad (n\to\infty).$$

Wir zeigen dies zunächst unter der Annahme der Feller-Bedingung (F). Dann sei $\alpha_n:=\max\limits_{j=1,..,n}\frac{\sigma_j}{s_n}$. Die Feller-Bedinung besagt, dass $\alpha_n\to 0$ $(n\to\infty)$. Dann:

$$(\mathrm{II}_n)=\frac{1}{s_n^2}\sum_{j=1}^n\sigma_j^2\int\limits_{\{|x|\geq\frac{s_n}{\sigma_j}\varepsilon\}} x^2\mathcal{P}_Z(dx)\leq\frac{1}{s_n^2}\underbrace{\sum_{j=1}^n\sigma_j^2}_{s_n^2}\int\limits_{\{|x|\geq\varepsilon\alpha_n^{-1}\}} x^2\mathcal{P}_Z(dx)\to 0,$$

da $\alpha_n^{-1}\to+\infty$ nach (F). Es bleibt zu zeigen, dass (L) die Feller-Bedingung (F) impliziert. Dies folgt aus ($\varepsilon>0$ beliebig)

$$\sigma_j^2=\int x^2d\mathcal{P}_{X_j}(dx)\leq\varepsilon^2s_n^2+\int\limits_{\{|x|\geq\varepsilon s_n\}} x^2d\mathcal{P}_{X_j}(dx)$$

Also:

$$\max_{j=1,\dots,n}\frac{\sigma_j^2}{s_n^2}\leq\varepsilon^2+\underbrace{\frac{1}{s_n^2}\sum_{j=1}^n\int\limits_{\{|x|\geq\varepsilon s_n\}} x^2d\mathcal{P}_{X_j}(dx)}_{\to 0\ (n\to\infty)\text{ wegen (L)}}$$

Da ε beliebig war, folgt (F). □

4.43 Bemerkung. Unter den Voraussetzungen des Satzes 4.41 sei $\varsigma_n:=E\left[|X_n-E[X_n]|^3\right]\leq\infty$ für alle $n\in\mathbb{N}$. Dann gilt folgende *Ungleichung von Berry und Esseen*:

$$\sup_{x\in\mathbb{R}}|F_{S_n}(x)-\Phi(x)|\leq c_0\left(\sum_{j=1}^n\varsigma_n\right)\left(\sum_{j=1}^n\sigma_j^2\right)^{-\frac{3}{2}}$$

für eine Konstante $0,4097\leq c_0\leq 0,5600$, die unabhängig von $(X_n)_{n\in\mathbb{N}}$ ist (Berry, 1941; Esseen, 1956; Shevtsova, 2010). Im Falls, dass $(X_n)_{n\in\mathbb{N}}$ unabhängig und identisch verteilt sind, gilt:

$$\sup_{x\in\mathbb{R}}|F_{S_n}(x)-\Phi(x)|\leq c_0\frac{\varsigma_1}{\sigma_1^3\sqrt{n}}$$

mit $c_0\leq 0,4748$ (Shevtsova, 2011). Hier ist Φ die Verteilungsfunktion der $\mathcal{N}(0,1)$-Verteilung. Sei speziell $(X_n)_{n\in\mathbb{N}}$ unabhängig Bernoulli(p)-verteilt. Dann:

$$E\left[|X_n-E[X_n]|^3\right]=(1-p)^3p+p^3(1-p)=p(1-p)(p^2+(1-p)^2)$$

Also:

$$\frac{\varsigma_1}{\sigma_1^3}=\frac{p^2+(1-p^2)}{\sqrt{p(1-p)}}\to+\infty \text{ für } p\to 1 \text{ oder } p\to 0$$

Für sehr kleines bzw. großes p braucht man also sehr viele Wiederholungen n, um die Approximation der Binomial(n,p)-Verteilung durch die Normalverteilung zu rechtfertigen. Hier bietet sich in der Praxis die Verwendung der Poisson-Approximation (Satz 1.35) häufig an.

4.7 Multivariable Normalverteilung

Motivation: Sei X eine $\mathcal{N}(0,1)$-verteilte Zufallsvariable mit $\mu \in \mathbb{R}$ und $\sigma > 0$. Wir wissen aus Beispiel 4.23, dass $\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$. Im Fall $\sigma = 0$ ist dies die charakteristische Funktion von $X := \mu$. Wir betrachten dies als degeneriert normalverteilt Zufallsvariable mit Erwartungswert μ und Varianz 0. Im 1-dimensionalen Falls ist X also genau dann normalverteilt, wenn $\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t}$, $\mu \in \mathbb{R}, \sigma \geq 0$.

4.44 Definition. Sei $\mu \in \mathbb{R}^D$ und Σ eine positiv semidefinite $D \times D$-Matrix. Das Wahrscheinlichkeitsmaß $\mathcal{P}$ auf $(\mathbb{R}^D, \mathcal{B}_D)$ heißt *multivariate Normalvertielung mit Parametern* μ, Σ ($\mathcal{N}(\mu, \Sigma$-Verteilung), falls die charakteristische Funktion von $\mathcal{P}$ gegeben ist durch $t \mapsto e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t, \Sigma t\rangle}$ $(t \in \mathbb{R}^D)$.
Eine $\mathbb{R}^D$-wertige Zufallsvariable heißt *multivariat normalverteilt*, falls die Verteilung von X eine multivariate Normalverteilung ist.

4.45 Satz. Sei $\mu \in \mathbb{R}^D$ und Σ eine positiv semidefinite $d \times D$-Matrix. Dann existiert die multivariate $\mathcal{N}(\mu, \Sigma)$-Verteilung als Maß $\mathcal{P}$ auf $(\mathbb{R}^D, \mathcal{B}_D)$.

Beweis. Es ist zu zeigen, dass ein Wahrscheinlichkeitsmaß $\mathcal{P}$ existiert mit

$$\int\limits_{\mathbb{R}^D} e^{i\langle t,x\rangle}\mathcal{P}(dx) = e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Sigma t\rangle} \text{ für alle } t \in \mathbb{R}^D.$$

Fall 1: Σ ist positiv definit.
Dann ist Σ invertierbar. Wir betrachten die Funktion

$$f_{\mu,\Sigma}(x) := \frac{1}{\sqrt{(s\pi)^D \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right),\ x \in \mathbb{R}^D.$$

Dazu sei $\mathcal{O}$ eine orthogonale Matrix und $\Delta = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_D \end{pmatrix}$ eine Diagonalmatrix mit $\mathcal{O}^T \Sigma \mathcal{O} = \Delta$. Da Σ positiv definit ist, folgt $\lambda_d > 0$ für alle $d = 1, ..., D$. Also $\Delta^{-1} = \mathcal{O}^T \Sigma^{-1} \Delta$. Substitution $x - \mu = \mathcal{O}u$ führt für jedes $v \in \mathbb{R}^D$ zu

$$\begin{aligned}
\int\limits_{\mathbb{R}^D} e^{i\langle \mathcal{O}v, x-\mu\rangle} f_{\mu,\Sigma}(x)\lambda^{\otimes D}(dx) &= \int\limits_{\mathbb{R}^D} e^{i\langle \mathcal{O}v, \mathcal{O}u\rangle} f_{\mu,\Sigma}(\mathbb{O}u+\mu)\lambda^{\otimes D}(du) \\
&= \int\limits_{\mathbb{R}^D} e^{i\langle v,u\rangle} \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathcal{O}u)^T \Sigma^{-1}(\mathcal{O}u)\right) \lambda^{\otimes D}(du) \\
&= \int\limits_{\mathbb{R}^D} e^{i\langle u,v\rangle} \frac{1}{\sqrt{(2\pi)^D \det(\Delta)}} \exp\left(-\frac{1}{2}u^T \underbrace{\mathcal{O}^T \Sigma^{-1} \mathcal{O}}_{\Delta^{-1}} u\right) \lambda^{\otimes D}(du) \\
&= \int\limits_{\mathbb{R}^D} \prod_{d=1}^{D} \frac{1}{\sqrt{2\pi\lambda_d}} \exp\left(-\frac{u_d^2}{2\lambda_d}\right) e^{iv_d u_d} \lambda^{\otimes D}(d(u_1, ..., u_D)) \\
&\overset{\text{Fubini}}{=} \prod_{d=1}^{D} \int\limits_{\mathbb{R}} \underbrace{\frac{1}{\sqrt{2\pi\lambda_d}} \exp\left(-\frac{u_d^2}{2\lambda_d}\right)}_{:=f_d(u_d)} e^{iv_d u_d} \lambda(du_d) = (*)
\end{aligned}$$

Da f_d Dichte einer 1-dimensionale $\mathcal{N}(0, \lambda_d)$-Verteilung ist, folgt aus Beispiel 4.23, dass

$$(*) = \prod_{d=1}^{D} e^{-\frac{1}{2}\lambda_d v_d^2} = \exp\left(-\frac{1}{2}v^T \Delta v\right) = \exp\left(-\frac{1}{2}(\mathcal{O}v)^T \Sigma \mathcal{O}v\right)$$

Insbesondere folgt mit $v = 0 \in \mathbb{R}^D$:

$$\int\limits_{\mathbb{R}^D} f_{\mu,\Sigma}(x)\lambda^{\otimes D}(dx) = 1$$

(d.h. $\mathcal{P}(A) = \int\limits_A f_{\mu,\Sigma}(x)\lambda^{\otimes D}(dx)$. ist ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^D, \mathcal{B}_D)$). Sei nun $t \in \mathbb{R}$ und $v = \mathcal{O}^T t$ (d.h. $\mathcal{O}v = t$), dann:

$$\int\limits_{\mathbb{R}^D} e^{i\langle t,x\rangle}\mathcal{P}(dx) = e^{i\langle t,\mu\rangle - \frac{1}{2}\langle t,\Sigma t\rangle}$$

Also existiert im positiv definiten Fall die $\mathcal{N}(\mu, \Sigma)$-Verteilung und hat $f_{\mu,\Sigma}$ als Dichte bezüglich des Lebesgue-Maßes $\lambda^{\otimes D}$.
Fall 2: Σ ist positiv semidefinit.
Satz $\Sigma_\varepsilon := \Sigma + \varepsilon \mathrm{I}_D$, wobei I_D die Einheitsmatrix in $\mathbb{R}^D$ ist. Dann ist Σ_ε für jedes $\varepsilon > 0$ positiv definit. Nach Fall 1 ist

$$\varphi^\varepsilon(t) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\langle t, \Sigma_\varepsilon t\rangle\right)$$

charakteristische Funktion eines Wahrscheinlichkeitsmaßes. Es gilt aber für alle $t \in \mathbb{R}^D$:

$$\lim_{n\to\infty} \varphi^{\frac{1}{n}}(t) = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}\langle t, \Sigma t\rangle\right)$$

Da die rechte Seite eine in 0 stetige Funktion definiert, folgt aus dem Stetigkeitssatz 4.26 (der auch im $\mathbb{R}^D$ gilt), dass die rechte Seite charakteristische Funktion eines Wahrscheinlichkeitsmaßes ist. □

4.46 Bemerkung. Wir haben im Beweis gezeigt, dass im Fall einer positiv definiten Matrix Σ die $\mathcal{N}(\mu, \Sigma)$-Verteilung $\mathcal{P}_{(\mu,\Sigma)}$ durch die Dihte

$$f_{\mu,\Sigma}(x) = \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left(\frac{1}{2}(x-\mu)^T \Sigma^{-1}\right), x \in \mathbb{R}^D$$

bezüglich des Lebesguemaßes auf $\mathbb{R}^D$ gegeben ist. Insbesondere ist die $\mathcal{N}(0, \mathrm{I}_D)$-Verteilung das D-fache Produktmaß von 1-dimensionalen $\mathcal{N}(0,1)$-Verteilungen.

4.47 Satz. Eine $\mathbb{R}^D$-wertige Zufallsvariable ist genau dann multivariat normalverteilt, wenn für alle $a \in \mathbb{R}^D$ $\langle a, X\rangle$ 1-dimensional normalverteilt ist.

Beweis. Sei X $\mathcal{N}(\mu, \Sigma)$-verteilt und $a \in \mathbb{R}^D$. Dann gilt für alle $t \in \mathbb{R}$:

$$E\left[e^{it\langle a,X\rangle}\right] = E\left[e^{i\langle ta,X\rangle}\right] = e^{it\langle a,\mu\rangle - \frac{1}{2}a^T\Sigma a}$$

Also ist $\langle a, X\rangle$ $\mathcal{N}(\langle a, \mu\rangle, a^T\Sigma a)$-verteilt.
Sei nun $\langle a, X\rangle$ normalverteilt für alle $a \in \mathbb{R}^D$. Setze

$$\mu := \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_D] \end{pmatrix}, \quad \Sigma = (Cov(X_i, X_j))_{i,j=1,\dots,D}.$$

Dann gilt für alle $t \in \mathbb{R}^D$:

$$E\left[e^{it\langle t,X\rangle}\right] = e^{i[\langle t,X\rangle] - \frac{1}{2}Var(\langle t,X\rangle)} = \exp\left(i\langle t, \mu\rangle - \frac{1}{2}t^T\Sigma t\right)$$

Also ist X $\mathcal{N}(\mu, \Sigma)$-verteilt. □

4.48 Bemerkung. Im Beweis wurde gezeigt:

(i) Ist $X = \begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}$ $\mathcal{N}(\mu, \Sigma)$-verteilt und $a \in \mathbb{R}^D$, so ist $\langle a, X\rangle$ $\mathcal{N}(\langle a, \mu\rangle, a^T\Sigma a)$-verteilt. Insbesondere ist X_d $\mathcal{N}(\mu_d, \sigma_{d,d})$-verteilt, wobei $\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_D \end{pmatrix}$ und $\Sigma = (\sigma_{i,j})$.

(ii) Ist $X = \begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}$ $\mathcal{N}(\mu, \Sigma)$-verteilt, so ist $\mu = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_D] \end{pmatrix}$ und $\Sigma = (Cov(X_i, X_j))_{i,j=1,\dots,D}$.
Der Vektor μ heißt *Erwartungsvektor*, die Matrix Σ *Kovarianzmatrix* von X. Analog zum ersten Teil des Beweises kann man zeigen: Ist $A \in \mathbb{R}^{K\times D}$, so ist die $\mathbb{R}^K$-wertige Zufallsvariable AX normalverteilt mit Erwartungsvektor $A\mu$ und Kovarianzmatrix $A\Sigma A^T$.

4.49 Bemerkung. Seien X_1, X_2 normalverteilt, so ist $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ nicht notwendigerweise multivariat verteilt. Zum Beispiel gilt:
Sei X_1 $\mathcal{N}(0,1)$-verteilt und μ Bernoulli($\frac{1}{2}$)-verteilt und unabhängig von X_1. Sei $X_2 := (2\mu - 1)\cdot X_1$. Dann ist X_2 ebenfalls $\mathcal{N}(0,1)$-verteilt, denn

$$\varphi_{X_2}(t) = E\left[e^{it(2\mu-1)X_1}\right] = \frac{1}{2}\left(\varphi_{X_1}(t) + \varphi_{X_1}(-t)\right) = \varphi_{X_1}(t).$$

Aber:

$$\mathcal{P}(\{X_1 + X_2 = 0\}) = \mathcal{P}(\{\mu = 0\}) = \frac{1}{2}$$

Somit ist $X_1 + X_2$ nicht normalverteilt. Nach Satz 4.47 kann auch $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ dann nicht multivariat 2-dimensional normalverteilt sein.

4.50 Satz. Sei $X = \begin{pmatrix} X_1 \\ \vdots \\ X_D \end{pmatrix}$ multivariat normalverteilt. Dann ist äquivalent:

(i) $X_1, ..., X_D$ sind unabhängig.

(ii) $X_1, ..., X_D$ sind paarweise unkorreliert.

Beweis. (ii)⇒(i): Sei X $\mathcal{N}(\mu, \Sigma)$-verteilt. Sind $X_1, ..., X_D$ paarweise unkorreliert, so hat Σ Diagonalform. Also:

$$\varphi_X(t) = \prod_{d=1}^{D} e^{it_d\mu_d - \frac{1}{2}t_d^2\sigma_{d,d}} = \prod_{d=1}^{D} \varphi_{X_d}(t_d), \quad t \in \mathbb{R}^D$$

Somit sind $X_1, ..., X_D$ unabhängig nach Satz 4.29. Die andere Richtung ist klar. □

<u>Ziel:</u> Multivariate Version des zentralen Grenzwertsatzes
Dazu:

- Sei $(\mathcal{P}_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}^D, \mathcal{B}_D)$ und $\mathcal{P}$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^D, \mathcal{B}_D)$. Man sagt, $(\mathcal{P}_n)_{n\in\mathbb{N}}$ *konvergiert schwach* gegen $\mathcal{P}$, falls für alle stetige und beschränkte $f : \mathbb{R}^D \to \mathbb{R}$ gilt:
$$\int f\, d\mathcal{P}_n \to \int f\, d\mathcal{P} \quad (n \to \infty)$$
- Eine Folge $(X_n)_{n\in\mathbb{N}}$ von $\mathbb{R}^D$-wertigen Zufallsvariablen *konvergiert in Verteilung* gegen eine $\mathbb{R}^D$-wertige Zufallsvariable X, falls $(\mathcal{P}_{X_n})_{n\in\mathbb{N}}$ schwach gegen $\mathcal{P}_X$ konvergiert.
- Es gilt die offensichtliche mehrdimensionale Version des Stetigkeitssatzes 4.26.

Verteilungskonvergenz im $\mathbb{R}^D$ kann mit folgendem Trick auf den 1-dimensionalen Fall zurückgeführt werden:

4.51 Satz. (Cramér-Wold-Device)
Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von $\mathbb{R}^D$-wertigen Zufallsvariablen und X eine $\mathbb{R}^D$-wertige Zufallsvariable. Dann sind äquivalent:

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert gegen X in Verteilung.

(ii) $\forall a \in \mathbb{R}^D$: $(\langle a, X_n\rangle)_{n\in\mathbb{N}}$ konvergiert gegen $\langle a, X\rangle$ in Verteilung.

Beweis. **(i)⇒(ii)** Sei $f : \mathbb{R}^D \to \mathbb{R}$ stetig und beschränkt. Dann ist für alle $a \in \mathbb{R}^D$ die Funktion

$$f_a : \mathbb{R} \to \mathbb{R}, x \mapsto f(\langle a, x\rangle)$$

stetig und beschränkt. Also:

$$\int f\, d\mathcal{P}_{\langle a, X_n\rangle} = E[f(\langle a, X_n\rangle)] = \int f_a\, d\mathcal{P}_{X_n} \to \int f_a\, d\mathcal{P}_X = \int f\, d\mathcal{P}_{\langle a, X\rangle}$$

(ii)⇒(i) Da $x \mapsto \Re(e^{itx})$ und $x \mapsto \Im(e^{itx})$ stetig und beschränkt sind, folgt für alle $t \in \mathbb{R}^D$:

$$\varphi_{X_n}(t) = E\left[e^{i\langle t, X_n\rangle}\right] \to E\left[e^{i\langle t, X\rangle}\right] = \varphi_X(t)$$

Aus dem Stetigkeitssatz folgt, dass $(\mathcal{P}_{X_n})_{n\in\mathbb{N}}$ schwach gegen $\mathcal{P}_X$ konvergiert. □

4.52 Satz. (Multivariater zentraler Grenzwertsatz)
Sei $(X_n)_{n\in\mathbb{N}} = (X_{n,1}, ..., X_{n,D})_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen im $\mathbb{R}^D$, so dass $X_{n,d} \in L^2(\mathcal{P})$ für alle $d = 1, ..., D$. Sei

$$\mu = (E[X_{1,1}], ..., E[X_{1,D}])^T \text{ und } \Sigma = (Cov(X_{1,i}, X_{1,j}))_{i,j=1,...,D}.$$

Dann gilt, dass $\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \mu)$ konvergiert gegen eine $\mathcal{N}(0, \Sigma)$-verteilte Zufallsvariable.

Beweis. Im Fall $D = 1$ und $\Sigma > 0$ ist die Aussage äquivalent zu

$$\frac{1}{\sqrt{\Sigma n}} \sum_{k=1}^{n} X_k - \mu \text{ konvergiert in Verteilung gegen eine } \mathcal{N}(0,1)\text{-verteilte Zufallsvariable,}$$

also zum zentralen Grenzwertsatz von Lindeberg-Levy.
Im Fall $D = 1$, $\Sigma = 0$ gilt $X_k = \mu$ $\mathcal{P}$-fast sicher, also auch $\frac{1}{\sqrt{n}} \sum\limits_{k=1}^{n} (X_k - \mu) \to 0$ $\mathcal{P}$-fast sicher. Dies impliziert die Verteilungskonvergenz gegen die $\mathcal{N}(0,0)$-verteilte Zufallsvariable 0. Also stimmt die Aussage für $D = 1$.
Sei nun $D > 2$ und Œ $\mu = 0 \in \mathbb{R}^D$. Sei Z $\mathcal{N}(0,\Sigma)$-verteilt und setze

$$X_n^{(a)} = \langle a, X_n \rangle \text{ und } S_n^{(a)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_n^{(a)}, a \in \mathbb{R}^D.$$

Dann $E[X_1^{(a)}] = \langle a, \mu \rangle$ und $Var(X_n^{(a)}) = a^T \Sigma a$.
Der Fall $D = 1$ liefert, dass $(S_n^{(a)})_{n\in\mathbb{N}}$ in Verteilung gegen $\langle a, Z \rangle$ konvergiert, da $\langle a, Z \rangle$ $\mathcal{N}(0, a^T \Sigma a)$-verteilt ist. Da $S_n^{(a)} = \langle a, \frac{1}{\sqrt{n}} \sum\limits_{k=1}^{n} (X_k - \mu) \rangle$, folgt $\langle a, S_n \rangle \to \langle a, Z \rangle$ in Verteilung für alle $a \in \mathbb{R}^D$. Die Verteilungskonvergenz von $(S_n)_{n\in\mathbb{N}}$ gegen Z folgt nun aus dem Cramér-Wold-Device. □

Ausblick

A Gesetz vom iterieren Logarithmus

Problemstellung: Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen in $L^2(\mathcal{P})$ mit $E[X_1] = 0$ und $Var(X_1) = 1$. Für $n \in \mathbb{N}$ sei

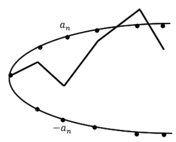
$$S_n = \sum_{k=1}^{n} X_k.$$

$(S_n)_{n\in\mathbb{N}}$ wird eine *Irrfahrt* genannt.
Man kann S_n als Position eines Teilchens zur Zeit n auffassen und ist daran interessiert, das Langzeitverhalten des Teilchens möglichst präzise zu beschreiben.
Gesucht ist also eine Folge $(a_n)_{n\in\mathbb{N}}$ von reellen Zahlen mit $\limsup\limits_{n\to\infty} \frac{S_n}{a_n} = 1$. Aus dem zentralen Grenzwertsatz folgt, dass die richtige Skalierung für Konvergenz in Verteilung $a_n = \sqrt{n}$ ist. Aus dem Kolmogorovschen 0-1-Gesetz (Stochastik II) folgt aber, dass

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = +\infty \;\; \mathcal{P}\text{-fast sicher.}$$

Andererseits zeigt dieselbe Argumentation wie zu Beginn von Abschnitt 4.3, dass

$$\lim_{n\to\infty} \frac{S_n}{\sqrt{n\log(n)}} = 0 \;\; \mathcal{P}\text{-fast sicher.}$$



Satz. (Hartman-Wintnersche Gesetz vom iterierten Logarithmus)
In der oben beschriebenen Situation gilt:

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log(n)}} = 1 \;\; \mathcal{P}\text{-fast sicher}$$

Der Beweis kann mit dem Lemma von Borel-Cantelli recht elementar geführt werden, falls X_1 standardnormalverteilt ist (s. Shiryaer, Section IV.4). In Stochastik II werden wir den Beweis für eine „normalverteilte Irrfahrt in stetiger Zeit“ (Brownsche Bewegung) führen und dann ein Verfahren zur Einbettung beliebiger Irrfahrten in diskreter Zeit in eine Brownsche Bewegung kennenlernen (Skorokhodsche Einbettungssatz).

B Der Satz von Radon-Nikodym

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $X \in \varepsilon^*$. In Satz 3.13 haben wir gezeigt, dass

$$\nu(A) = \int_A X d\mu, \;\; A \in \mathcal{A}$$

ein Maß auf $(\Omega, \mathcal{A})$ definiert. Dieses Maß ν ist *absolut stetig* bezüglich μ in dem Sinn, dass $\forall A \in \mathcal{A}$:

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

Der Satz von Radon-Nikodym beinhaltet siw Umkehrung:

Satz. (Radon-Nikodym)
Seien μ, ν σ-endliche Maße auf $(\Omega, \mathcal{A})$, so dass ν absolut stetig bezüglich μ ist. Dann existiert $X \in \varepsilon^*$ mit

$$\nu(A) = \int_A X d\mu \text{ für alle } A \in \mathcal{A}.$$

X ist bis auf μ-Nullmenge eindeutig bestimmt und heißt *Radon-Nikodym-Dichte* von ν bezüglich μ.

<u>Spezialfall:</u> Seien $A_1, .., A_\nu \in \Omega$ mit $A_k \neq \varnothing$, $A_k \cap A_j = \varnothing$ und $\bigcup_{k=1}^{N} A_k = \Omega$. Sei $\mathcal{A} = \left\{ \bigcup_{k \in R} A_k \mid R \subset \{1, ..., K\} \right\}$ und seien μ, ν endlich. Setze

$$X(\omega) = \sum_{k=1}^{K} \mathbb{1}_{A_k}(\omega) \left(\mathbb{1}_{\{\mu(A_k) \neq 0\}} \frac{\nu(A_k)}{\mu(A_k)} \right)$$

Dann gilt für alle $A = \bigcup_{k \in R} A_k \in \mathcal{A}$:

$$\begin{aligned}
\nu(A) = \sum_{k \in R} \nu(A_k) &\overset{\text{absolut}}{\underset{\text{stetig}}{=}} \sum_{k \in R} \nu(A_k) \mathbb{1}_{\{k \mid \mu(A_k) \neq 0\}} \\
&= \sum_{k \in R} \mathbb{1}_{\{k \mid \mu(A_k) \neq 0\}} \frac{\nu(A_k)}{\mu(A_k)} \int \mathbb{1}_{A_k} d\mu \\
&= \sum_{k \in R} \int X \mathbb{1}_{A_k} d\mu = \int_A X d\mu
\end{aligned}$$

Der Fall allgemeiner σ-Algebren kann sukzessive auf diesen einfachen Fall zurückgeführt werden. Dafür benötigen wir Martingalkonvergenzsatz aus Stochastik II („Martingale“ sind die stochastischen Modelle für „faire Spiele“).

C Der Kolmogorovsche Ausweitungssatz

Sei $X = (X_n)_{n \in \mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathcal{P})$. Zu X definiert man die *endlich-dimensionalen Randverteilungen* durch

$$\mathcal{P}_n(A) := \mathcal{P}\left(\left\{ \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix} \in A \right\} \right), \quad n \in \mathbb{N}, \quad A \in \mathcal{B}_n.$$

Sei erfüllen offensichtlich die folgende Konsistenzbedingung:

$$\mathcal{P}_{n+1}(A \times \mathbb{R}) = \mathcal{P}_n(A) \text{ für alle } n \in \mathbb{N} \text{ und } A \in \mathcal{B}_n \quad \text{(K)}$$

Es gilt:

Satz. (Kolmogorovsche Ausweitungssatz)
Für jedes $n \in \mathbb{N}$ sei $\mathcal{P}_n$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^n, \mathcal{B}_n)$. Die Familie $(\mathcal{P}_n)_{n \in \mathbb{N}}$ erfülle die Konsistenzbedingung (K). Dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ und eine Folge $(X_n)_{n \in \mathbb{N}}$ von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathcal{P})$, die $(\mathcal{P}_n)_{n \in \mathbb{N}}$ als endlich-dimensionale Randverteilungen hat.

<u>Beispiel:</u> Sei Q ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ und sei $\mathcal{P}_n = Q^{\otimes n}$. Dann:

$$\mathcal{P}_{n+1}(A \times \mathbb{R}) = Q^{\otimes n} \otimes Q(A \times \mathbb{R}) = Q^{\otimes n}(A) \underbrace{Q(\mathbb{R})}_{=1} = \mathcal{P}_n(A), \quad n \in \mathbb{N}, \quad A \in \mathcal{B}_n.$$

Die Familie $(\mathcal{P}_n)_{n \in \mathbb{N}}$ erfüllt die Konsistenzbedingung. Es existiert also ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathcal{P})$ und eine Folge $(X_n)_{n \in \mathbb{N}}$ von reellwertigen Zufallsvariablen auf $(\Omega, \mathcal{A}, \mathcal{P})$, die $Q^{\otimes n}$ als n-dimensionales Randverteilung hat (für alle $n \in \mathbb{N}$). Insbesondere sind $(X_1, ..., X_n)$, $n \in \mathbb{N}$, unabhängig und jedes X_k ist Q-verteilt. Also existieren Folgen von unabhängigen, identisch verteilten Zufallsvariablen zu vorgegebener Verteilung Q.

<u>Idee:</u> Fasse die Folge $X = (X_n)_{n \in \mathbb{N}}$ als Abbildung

$$X : \Omega \to \mathbb{R}^N = \text{„Raum der reellwertigen Folgen“}$$

auf und studiere die Existenz (und Eindeutigkeit) von Wahrscheinlichkeitsmaßen auf $\mathbb{R}^N$ ausgestattet mit einer geeigneten σ-Algebra.