Stochastik 1 Skript

10. Juli 2018

Literatur

- Shiryaer: Probabillity, Springer
- Bauer: Wahrscheinlichkeitstheorie, de Gruyter
- Klenke: Wahrscheinlichkeitstheorie, Springer

1 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 nichtleere Menge aller (möglichen) Versuchsausgänge.
Ω wird Stichprobenraum oder 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 "Warten, bis zum ersten Mal eine 6 fällt" betrachten. Hier bietet sich als Stichprobenraum $\Omega = \mathbb{N} = \{1, 2, ...\}$ an, da beliebig häufig andere Zahlen auftreten können, bis zum ersten Mal eine 6 fällt.

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

1.2 Definition

Sei $\Omega \neq \emptyset$ höchstens abzählbar und $p : \Omega \to \mathbb{R}$ mit

(i) $p(\omega) \geq 0 \ \forall \ \omega \in \Omega$

(ii) $\displaystyle\sum_{\omega \in \Omega} p(\omega) = 1$

Dann wird p eine Zähldichte genannt und das Paar (Ω, p) ein diskretes Zufallsexperiment.
Wir interpretieren $p(\omega)$ als die Wahrscheinlichkeit für das Auftreten der Ereignis ω.
Oft ist man an Wahrscheinlichkeiten von Mengen von Ergebnnissen 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:

- $\emptyset$: 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
$B \backslash A$	B tritt ein, aber A nicht
A^C	A tritt nicht ein

A und B werden unvereinbar genannt, falls $A \cap B = \emptyset$.
Gegeben ein diskretes Zufallsexperiment (Ω, p) betrachten wir die Abbildung

$$P : 2^\Omega \to \mathbb{R} \quad A \to \sum_{\omega \in A} p(\omega)$$

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

1.3 Motivation

Man beachte, dass 1.1 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 2,4 bzw. 6, d.h.

$$\frac{\#\text{gerade Zahl}}{n} = \frac{\#2}{n} + \frac{\#4}{n} + \frac{\#6}{n}$$

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 $\Omega \neq \emptyset$ höchstens abzählbar und $P : 2^\Omega \to \mathbb{R}$. Dann sind äquivalent:

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

(ii) P erfüllt folgende Eigenschaften:

  (a) $P(A) \geq 0 \; \forall \; A \in 2^\Omega$
  (b) $P(\Omega) = 1$
  (c) P ist σ-additiv, d.h. $\forall$ Folgen $(A_n)_{n\in\mathbb{N}} \subset 2^\Omega$ gilt:
  $A_n \cap A_m = \emptyset \; \forall \; n, m \in \mathbb{N}, n \neq m \Rightarrow P(\bigcup_{n=1}^\infty A_n) = \sum_{n=1}^\infty P(A_n)$

Beweis:

- (i)$\Rightarrow$(ii): a) und b) folgen direkt aus 1.1 und der Definition der Zähldichte.
  Zu c): Seien $(A_n)_{n\in\mathbb{N}}$ paarweise disjunkt. Dann:

$$\sum_{n=1}^\infty P(A_n) = \sum_{n=1}^\infty \sum_{\omega \in A_n} p(\omega) = \sum_{\omega \in \bigcup_{i=1}^\infty A_n} p(\omega) = P\left(\bigcup_{n=1}^\infty A_n\right)$$

  (Doppelreihensatz oder Großer Umordnungssatz)
  wobei die Vertauschung der Summationsreihenfolge aufgrund der Nichtnegativität der Zähldichte durch den großen Umordnungsssatz gerechtfertigt ist.

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

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

  also ist P eine Zähldichte, da $p(\omega) = P(\{\omega\}) \geq 0$ nach a) und

$$\sum_{\omega \in \Omega} p(\omega) = P(\Omega) = 1$$

Erinnerung: Der große Umordnungssatz besagt(s. Königsberger Analysis 1):

Sei $\mathbb{N} = \bigcup_{j\in\mathbb{N}} T_j$ für disjunkte Mengen $(T)_{j\in\mathbb{N}}$ und $(a_n)_{n\in\mathbb{N}}$ eine Folge von reellen Zahlen mit $\sum_{n=1}^{\infty} |a_n| < \infty$.
Sei $b_j = \begin{cases} 0, T_j = \emptyset \\ \sum_{n\in T_j} a_n, T_j \neq \emptyset \end{cases}$
Dann gilt: $\sum_{n=1}^{\infty} a_n = \sum_{j=1}^{\infty} b_j$.

1.5 Definition

Sei $\Omega \neq \emptyset$ höchstens abzählbar.

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

(ii) ein Tripel $\left(\Omega, 2^\Omega, P\right)$, wobei P ein diskretes Wahrscheinlichkeitsmaß auf $\left(\Omega, 2^\Omega\right)$ist, heißt diskreter Wahrscheinlichkeitsraum.

1.2: Allgemeine Wahrscheinlichkeitsräume

Häufig treten Zufallsexperimente mit überabzählbarem Stichprobenraum auf.

1.6 Beispiel

(i) Wird ein Münzwurf beliebig oft ausgeführt, so bietet sich als Stichprobenraum

$$\Omega = \{0,1\}^{\mathbb{N}} = \left\{(\omega_n)_{n\in\mathbb{N}} | \omega_n \in \{0,1\} \ \forall\, n \in \mathbb{N}\right\} \text{ an.}$$

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

Bei überabzählbaren Stichprobenräumen ist die Potenzmenge häufig zu groß, um dort Wahrscheinlichkeitsmaße 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 $P : 2^\Omega \to \mathbb{R}$ mit $P(A) \geq 0 \ \forall\, A \in 2^\Omega, P(\Omega) = 1$, die die folgende Invarianzeigenschaft erfüllt:

$$\forall\, n \geq 1 \text{ und } A \in 2^\Omega \text{ gilt: } P(T_n A) = P(A) \text{ wobei:}$$

$T_n : \Omega \to \Omega, (\omega_1, \omega_2, ...) \to (\omega_1, ..., \omega_{n-1}, 1 - \omega_n, \omega_{n+1}, ...)$ Beweis: Wir nehmen an, ees gibt eine solche Abbildung.
Für $\omega, \omega' \in \Omega$ Sei $\omega \sim \omega' :\Leftrightarrow \ \exists\, n_0 : \ \forall\, n \geq n_0 : \omega_1 = \omega'_1$ Man prüft leicht nach, dass eine Äquivalenzrelation definiert. Nach dem Auswahlaxiom existiert eine Menge $A^* \subset \Omega$, die aus jeder Äquivalenzklasse genau ein Element enthält.
Sei $\rho = \{S \subset \mathbb{N}(\text{S hat höchstens endlich viele Elemente})\}$
ρ ist abzählbar unendlich als disjunkte Vereinigung der endlichen Mengen

$$\{S \in \rho | \max S = m\}, \ m \in \mathbb{N} \text{ und } \{\emptyset\}$$

Für $S = \{n_1, ..., n_k\} \in \rho$ sei $T_s := T_{n_1} \circ \cdots \circ T_{n_k}$ mit der Konvention $T_\emptyset = \text{id}$.
Dann gilt:

(a) $\Omega = \bigcup_{S\in\rho} T_S A^*$, denn zu jedem $\omega \in \Omega$ existiert ein $\omega' \in A^*$ mit $\omega \sim \omega'$ und also $\omega = T_S \omega' \in T_S A^*$ für ein $S \in \rho$

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

$$\omega \sim T_S \omega = T_{S'} \omega' \sim \omega'$$

Nach Konstruktion von A^* folgt $\omega = \omega'$ (Da A^* nur ein Element aus jeder Äquivalenzklasse enthält) und also $S = S'$

(c) $P(T_S A^*) = P(A^*)\ \forall\ S \in \rho$. Dies folgt induktiv aus der Invarianzeigenschaft.

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

$$P(\Omega) = P\left(\bigcup_{S\in\rho} T_S A^*\right) = \sum_{S\in\rho} P(T_S A^*) = \sum_{S\in\rho} P(A^*) = \left\{\begin{matrix} 0 & , P(A^*) = 0 \\ \infty & , P(A^*) \neq 0 \end{matrix}\right\} \in \{0, \infty\}$$

im Widerspruch zu $P(\Omega) = 1$
Der Ausweg besteht darin, nicht mehr allen Teilmengen vom Ω Wahrscheinlichkeiten zuordnen zu wollen, sondern nur noch "geeigneten" Teilmengen.

1.8 Definition

Sei $\Omega \neq \emptyset$ Ein Mengensystem $\alpha \subset 2^\Omega$ wird σ-Algebra genannt, falls gilt:

- $\Omega \in \alpha$
- $A \in \alpha \Rightarrow A^C \in \alpha$
- $(A_n)_{n\in\mathbb{N}} \subset \alpha \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \alpha$

1.9 Beispiel

(i) 2^Ω ist eine σ-Algebra

(ii) $\{\emptyset, \Omega\}$ ist eine σ-Algebra, die sogenannte triviale σ-Algebra

1.10 Lemma

Seien $(\alpha)_{i\in\mathbb{N}}$ σ-Algebren auf Ω, wobei I eine nicht-leere Menge ist.
Dann ist $\bigcap_{i\in I} \alpha_i$ eine σ-Algebra auf Ω
Beweis:

- $\Omega \in \alpha\ \forall\ i \in I$. Also $\Omega \in \bigcap_{i\in I} \alpha_i$
- $A \in \alpha_i\ \forall\ i \in I \Rightarrow A^C \in \alpha_i\ \forall\ i \in I$
- $(A_n)_{n\in\mathbb{N}} \subset \alpha_i\ \forall\ i \in I \Rightarrow \bigcap_{n\in\mathbb{N}} A_n \in \alpha_i\ \forall\ i \in I$

1.11 Definition

Sei $\Omega \neq \emptyset$ und $M \subset 2^\Omega$ ein System von Teilmengen von Ω. Dann wird

$$\sigma(M) = \bigcap\{\alpha\ \sigma\text{-Algebra auf}\Omega \text{ mit } M \subset \alpha\}$$

die von M erzeugte σ-Algebra genannt Lemma 1.10 stellt sicher, dass $\sigma(M)$ eine σ-Algebra auf Ω definiert.

1.12 Beispiel

Sei $\Omega = \mathbb{R}^D$ Wir notieren mit
$[a, b) = [a_1, b_1) \times \cdots \times [a_D, b_D), a = (a_1, \ldots, a_D) \in \mathbb{R}^D;\ b = (b_1, \ldots, b_D) \in \mathbb{R}^D$
die nach rechts halboffenen Intervalle in $\mathbb{R}^D$. (Konv: $b_i < a_i \Rightarrow [a_i, b_i) = \emptyset$) Entsprechend sind die Notationen $(a, b], [a, b], (a, b)$ zu lesen. Dann erzeugen die Mengensysteme

$$J_1 = \left\{(a, b] | a, b \in \mathbb{R}^D\right\} \quad J_3 = \left\{(a, b) | a, b \in \mathbb{R}^D\right\}$$
$$J_2 = \left\{[a, b) | a, b \in \mathbb{R}^D\right\} \quad J_4 = \left\{[a, b] | a, b \in \mathbb{R}^D\right\}$$

allesamt die gleiche σ-Algebra, die Borelsche σ-Algebra auf $\mathbb{R}^D$ genannt wird und die wir mit $\mathcal{B}_D$ notieren.

1.13 Definition

(i) Sei $\Omega \neq \emptyset$ und α eine σ-Algebra auf Ω. Dann wird (Ω, α) ein Messraum genannt.

(ii) Sei (Ω, α) ein Messraum, $P : \alpha \to \mathbb{R}$ wird Wahrscheinlichkeitsmaß auf (Ω, α) genannt, falls:

  (a) $P(\Omega) = 1$

  (b) $P(A) \geq 0 \; \forall \, A \in \alpha$

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

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

(iii) Ist (Ω, α) ein Messraum und P ein W-Maß auf (Ω, α), so wird (Ω, α, P) ein Wahrscheinlichkeitsraum genannt.

1.14 Bemerkung

Sei P ein W-Maß auf (Ω, α) und $A \in \alpha$. Dann folgt aus der σ-Additivität $1 = P(\Omega) = P\big(A \cup A^C \cup \emptyset \cup \emptyset \cup \ldots\big) = P(A) + P\big(A^C\big) + \sum_{i=1}^{\infty} P(\emptyset)$ Also folgt $P(\emptyset) = 0$ und $P(A) \leq 1 \; \forall \, A \in \alpha$

1.15 Definition

Sei (Ω, α) ein Messraum. Eine Abbildung $\mu : \alpha \to \mathbb{R} \cup \{\infty\}$ wird Maß auf (Ω, α), falls Eigenschaften b), c) aus der Definition eines W-Maßes erfüllt sind und statt a) gilt: $\mu(\emptyset) = 0$
(Alle Eigenschaften bis auf Normierung erfüllt, statt Normierung gilt $\mu(\emptyset) = 0$) Das Tripel (Ω, α, μ) heißt dann ein Maßraum.
Eim Maß μ heißt endlich, falls $\mu(\Omega) < \infty$, und σ-endlich, falls eine Folge $(B_n)_{n\in\mathbb{N}}$ in α existiert mit $\mu(B_n) < \infty \; \forall \, n \in \mathbb{N}$ und $\Omega = \bigcup_{n\in\mathbb{N}} B_n$

1.3 Konstruktion von Maßen

- Motivation: Sei $\Omega \neq \emptyset$. Häufig ist uns klar, wie wir ein Maß auf gewillen Teilmengen von Ω definieren wollen. So kann man z.B. den Intervallen in $\mathbb{R}$ die Intervalllänge als Maß zuorden.
- Frage: Unter welchen Bedingungen ist es möglich, ein auf gewissen Teilmangen festgelegtes Maß auf eine σ-Algebra auszudehnen?

Wir betrachten zunächst 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 $\lim_{x\to\infty} F(x) = 1$ und $\lim_{x\to-\infty} F(x) = 0$, dann heißt F eine Verteilungsfunktion.
Sei $J_1 = \{(a, b] | a \leq b \in \mathbb{R}\}$ die Menge aller linksoffenen Intervalle in $\mathbb{R}$.
Ziel ist es, eine σ-Algebra α mit $J_1 \subset \alpha$ zu konstruieren und ein Maß μ auf α 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 Sinn folgender Definition bildet:

1.16 Definition

Sei $\Omega \neq \emptyset$ und $\mathcal{R} \subset 2^{\Omega}$. $\mathcal{R}$ wird ein Ring auf Ω genannt, falls

- $\Phi \in \mathcal{R}$
- $A, B \in \mathcal{R} \Rightarrow A \backslash B \in \mathcal{R}$ 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. $\mu_0 : \mathcal{R} \to \mathbb{R} \cup \{\pm\infty\}$ wird Inhalt genannt, falls:

- $\mu_0(\emptyset) = 0$
- $\mu_0(A) \geq 0 \;\forall\, A \in \mathcal{R}$
- $A, B \in \mathcal{R}$ 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:

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

, 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 $\mu_0((a,b]) = F(b) - F(a) \;\forall\, a \leq b$

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

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

Jeder Inhalt auf $\mathcal{F}$ ist also durch die Werte auf den linkshalboffenen 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 disjunkte Vereinigung aus Mengen aus J_1, also $(a,b] = \bigcup_{n=1}^{N-1}(a_n, a_{n+1}]$ mit $a = a_1 \leq a_2 \leq \cdots \leq a_n = b$. Dann:

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

Sei nun $A \in \mathcal{F}$ und seien zwei Zerlegungen von A als Vereinigung disjunkter Mengen in J_1 gegeben.

$$A = \bigcup_{k=1}^{K} I_k = \bigcup_{m=1}^{M} \tilde{I}_m \quad \text{(I)}$$

Dann folgt aus a)

$$\mu_0(I_k) = \mu_0(I_k \cap A) = \mu_0\left(\bigcup_{m=1}^{M} \big(I_k \cap \tilde{I}_m\big)\right) \underset{\text{(I)}}{=} \sum_{m=1}^{M} \mu_0\big(I_k \cap \tilde{I}_m\big)$$

$$\text{und } \mu_0\big(\tilde{I}_m\big) = \sum_{k=1}^{K} \mu_0\big(I_k \cap \tilde{I}_m\big)$$

Somit:

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

Also ist die Abbildung $\mu_0 : \mathcal{F} \to \mathbb{R}, A \to \sum_{k=1}^{K} \mu_0(I_k)\,; I_1, \dots, I_k \in J_1$ disjunkt mit $\bigcup_{k=1}^{K} I_k = A$ wohldefiniert.
μ_0 ist Inhalt:

- $\mu_0((0,0]) = F(0) - F(0) = 0$

- $\mu_0(A) \geq 0$: Aufgrund der Konstruktion reicht es, $A \in J_1$ zu betrachten. Für $A \in J$ folgt die Nichtnegativität aus der Monotonie von F_k
- Additivität: Seien $A, B \in \mathcal{F}$ disjunkt mit Darstellungen $A = \bigcup_{k=1}^{K} I_k$; $B = \bigcup_{m=1}^{M} \tilde{I}_m$.
  Dann ist $A \cup B = \left(\bigcup_{k=1}^{K} I_k\right) \cup \left(\bigcup_{m=1}^{M} \tilde{I}_m\right)$ eine disjunkte Vereinigung ALso:

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

■

Um zu zeigen, dass das eben definierte Maß 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 Prämaß

$$\Updownarrow$$

(b) μ_0 ist stetig von 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:

$$\mu_0\left(\bigcup_{k=1}^{K} A_n\right) = \lim_{n \to \infty} \mu_0(A_n)$$

$$\Downarrow$$

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

$$\mu_0\left(\bigcap_{k=1}^{K} A_n\right) = \lim_{n \to \infty} \mu_0(A_n)$$

$$\Updownarrow$$

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

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

Ist μ_0 endlich, d.h. $\mu_0(A) < \infty \ \forall \ A \in \mathcal{R}$, 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 $\emptyset$ zu zeigen.
Sei $(A_n)_{n \in \mathbb{N}}$ eine Folge in $\mathcal{F}$ mit $A_n \supset A_{n+1}$ mit $\lim_{n \to \infty} \mu_0(A_n) > 0$.
Dann ist zu zeigen, dass $\bigcap_{n \in \mathbb{N}} \neq \emptyset$.

- Schritt 1:
  Zu jedem $\varepsilon > 0$ und jedem $A \in \mathcal{F}$ existiert ein $B \in \mathcal{F}$ mit $\bar{B} \subset A$ und $\mu_0(B) \geq \mu_0(A) - \varepsilon$.
  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: $\bar{B} \subset A$ und $\mu_0(B) = F(b) - F(c) \Rightarrow F(b) - F(a) = \mu_0(A)$ für $c \downarrow a$ wegen der Rechtsstetigkeit von F.
  Insbesondere ist $\mu_0(B) \geq \mu_0(A) - \varepsilon$ falls c hinreichend nahe an a ist.

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

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

- Schritt 3:
  Wir zeigen, dass $\bigcap_{n\in\mathbb{N}} \bar{C}_n \neq \emptyset$.
  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 nichtleer und also $\bar{C}_n \neq \emptyset\ \forall\, n \in \mathbb{N}$.
  Da die Mengen A_n beschränkt sind und $\bar{C}_n \subset A_n$, ist $\big(\bar{C}_n\big)_{n\in\mathbb{N}}$ eine Folge von kompakten Mengen in $\mathbb{R}$ mit $\bar{C}_n \supset \bar{C}_{n+1}\ \forall\, n \in \mathbb{N}$. Wähle $x_n \in \bar{C}_n$ 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 alle $m \in \mathbb{N}$ und $n_k \geq m$ gilt: $x_{n_k} \in \bar{C}_m$. Die Kompaktheit von $\bar{C}_m$ impliziert, dass $x \in C_m\ \forall\, m \in \mathbb{N}$. Insbesondere:

  $$\bigcap_{m\in\mathbb{N}} \bar{C}_m \neq \emptyset$$

  Schritt 3 schließt den Beweis ab, da

  $$\emptyset \neq \bigcap_{m\in\mathbb{N}} \bar{C}_m \subset \bigcap_{m\in\mathbb{N}} A_m$$

  ■

Gegeben sei eine Maßerzeugende Funktion F auf $\mathbb{R}$ 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 $\mathcal{B}$ ausweiten. Dazu:

1.21 Satz(Ausweitungssatz von Caratheodory)

Sei μ_0 ein Prämaß auf einem Ring $\mathcal{R}$ auf Ω. Dann:

(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 σ-endlich, so ist das Maß in i) eindeutig bestimmt.

Beweisskizze später
Als Folgerung ergibt sich:

1.22 Satz

Sei $F : \mathbb{R} \to \mathbb{R}$ maßerzeugend. Dann existiert genau ein Maß μ_F auf $\mathcal{B}$ 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 $F(x) = G(x) + c \; \forall \; x \in \mathbb{R}$.
Beweis:
Da $J_1 \subset \mathcal{F} \subset \mathcal{B}$ folgt $\mathcal{B} = \sigma(J_1) \subset \sigma(\mathcal{F}) \subset \sigma\mathcal{B} = \mathcal{B}$.

2 Kap2

3 Kap3

28.05.2018

3.1 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:

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

3.2 Satz (Minkovski-Ungleichung)

Sei $X, Y \in \mathcal{L}^p(\mu)$. Dann ist $X + Y \in \mathcal{L}^p(\mu)$

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

Beweis:
Œ seien X, Y nicht negativ, da $|X+Y|^p \leq (|X| + |Y|)^p$
$p = 1$ trivial, dass Œ$p > 1$
Œ$\int |X+Y|^p \,\mathrm{d}\mu > 0$
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:

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

Beweis von Satz 3.20:
$\mathcal{L}^p(\mu)$ ist eine Teiklmenge des Vektorraums der $\mathcal{A}|\mathcal{B}$-messbaren Funktionen. Nach Satz 3.263.2 ist $\mathcal{L}^p(\mu)$ bzgl $'+'$ abgeschllossen und $\|\cdot\|_p$ erfüllt die Dreiecksungleichung.
Sei nun $X \in \mathcal{L}^p(\mu)$ uns $\alpha \in \mathbb{R}$. Dann:

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

3.3 Bemerkung:

Nach Lemma 3.21 gilt: $\|X\|_p = 0 \Leftrightarrow X = 0$ μ-fast sicher.
$\|\cdot\|_p$ ist keine Norm auf $\mathcal{L}^p(\mu)$. Betrachtet man aber die Äquivalenzklassen

$$[X] := \{Y \text{ reellwertig, messbar} | X = Y \, \mu\text{-fast sicher}\}$$

so ist

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

eine Norm auf dem Vektorraum $L^p(\mu)$ aller Äquivalenzklassen von Funktion $X \in \mathcal{L}^p(\mu)$ bzgl. mit

$$X \sim Y :\Leftrightarrow X = Y \ \mu\text{-fast sicher}$$

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

3.4 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.

3.5 Satz(Lemma von Fatou)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{E}^*$. Dann:

$$\int \liminf_{n\to\infty} \mathrm{d}\mu \leq \liminf_{n\to\infty} \int X_n \, \mathrm{d}\mu$$

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 $\mathcal{E}^*$. Da die Folge $(Y_n)_{n\in\mathbb{N}}$ nicht fallend ist, impliziert der Satz von Beppo-Levi:

$$\int \liminf_{n\to\infty} \mathrm{d}\mu = \int \sup_{n\in\mathbb{N}} Y_n \, \mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int Y_n \, \mathrm{d}\mu = \sup_{n\in\mathbb{N}} \int \inf_{m\geq n} X_m \, \mathrm{d}\mu \leq \sup_{n\in\mathbb{N}} \inf_{m\geq n} \int X_n \, \mathrm{d}\mu = \liminf_{n\to\infty} \int X_n \, \mathrm{d}\mu$$

3.6 Satz(Lebesgue, dominante 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 $Y \in \mathcal{E}^*$ mit $|X_n| \leq Y$ μ-fast sicher und $\int Y^p \, \mathrm{d}\mu < \infty$, so gilt:

$$\|X_n - X\| \to 0 \quad (n \to \infty).$$

Im Fall $p = 1$ gilt zudem:

$$\lim_{n\to\infty} \int X_n \, \mathrm{d}\mu = \int X \, \mathrm{d}\mu \, .$$

Beweis:
Da $X_n \to X$ μ-fast sicher, folgt, dass auch $|X|^p \leq Y^p$ μ-fast sicher. Nach Folgerung 3.23 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 $\int Y_n \,\mathrm{d}\mu \to 0 \quad (n \to \infty)$
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 (und Folgerung 3.22) folgt:

$$\int \liminf_{n\to\infty} (Z - Y_n) \,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int (Z - Y_n) \,\mathrm{d}\mu = \int Z \,\mathrm{d}\mu - \limsup_{n\to\infty} \int Y_n \,\mathrm{d}\mu$$

Da $X_n \to X$ μ-fast sicher, folgt $Z - Y_n \to Z$ μ-fast sicher. Also:

$$\int \liminf_{n\to\infty} (Z - Y_1) \mathrm{d}\mu = \int Z \mathrm{d}\mu$$

Also:

$$\limsup_{n\to\infty} \int Y_n \mathrm{d}\mu \leq 0$$

Da $\int Y_n \mathrm{d}\mu \geq 0$, folgt $\lim_{n\to\infty} \int Y_n \mathrm{d}\mu = 0$.
Dann folgt aus dem Gezeigten:

FEHLT

3.7 Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge in $\left(\mathcal{L}^p(\mu), \|\cdot\|_p\right)$. Dann existiert ein $X \in \mathcal{L}^p(\mu)$ mit $\|X_n - X\|_p \to 0 \quad (n \to \infty)$.
Die Räume $\left(L^p(\mu), \|\cdot\|_p\right)$ 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

$$\left\|X_{n_{k+1}} - X_{n_k}\right\| \leq 2^{-k} \; \forall \, k \in \mathbb{N}$$

Sei $Y_k := X_{n_{k+1}} - X_{n_k}$ und $Y := \sum_{k=1}^{\infty} |Y_k|$.
Aus der Mirkowski-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-fast sicher. Da $X_{n_k} = X_{n_1} + \sum_{l=1}^{k-1} Y_l$, konvergiert fie 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 Mirkowski-Ungleichung ist $\int (|X_{n_k} + Y|)^p \,\mathrm{d}\mu < \infty$. Der Satz von Lebesgue impliziert, dass $\|X_{n_k} - X\|_p \to \infty \quad (k \to \infty)$. Da $(X_n)_{n\in\mathbb{N}}$ eine Cauchy-Folge ist, konvergiert die ganze Folge in $\mathcal{L}^p(\mu)$ gegen X.
Im Beweis wurde gezeigt:

3.8 Folgerung

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

3.4 Integration bzgl. Produktmaßen

Es seien $(\Omega, \mathcal{A}_i, \mu_i), i = 1, 2$, zwei Maßräume. Auf $\Omega \times \Omega$ 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 sog. Progukt-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)

$$Q_{\omega_1} := \{\omega_2 \in \Omega | (\omega_1, \omega_2) \in Q\}\,, \quad \omega_1 \in \Omega$$
$$Q_{\omega_2} := \{\omega_1 \in \Omega | (\omega_1, \omega_2) \in Q\}\,, \quad \omega_2 \in \Omega$$

3.9 Lemma

Für $Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ gilt: $Q_{\omega_1} \in \mathcal{A}_2$ und $Q_{w_2} \in \mathcal{A}_1$ 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 $\mathcal{M}$ σ-Algebra auf $\Omega_1 \times \Omega_2$. Denn:

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

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

$$(A_1 \times A_2) = \begin{cases} A_2, \ \omega_1 \in A_1 \\ \emptyset, \ \omega_1 \notin A_1 \end{cases} \in \mathcal{A}_2$$

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 $X : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}} \quad \mathcal{A}_1 \otimes \mathcal{A}_2 | \bar{\mathcal{B}}$ - messbar. Wir definieren die ω_i-Schnitte

$$X_{\omega_1} : \Omega_2 \to \bar{\mathbb{R}}, \omega_2 \to X(\omega_1, \omega_2), \omega_1 \in \Omega_2$$
$$X_{\omega_2} : \Omega_1 \to \bar{\mathbb{R}}, \omega_1 \to X(\omega_1, \omega_2), \omega_2 \in \Omega_2$$

3.10 Lemma

Ist X $\mathcal{A}_1 \otimes \mathcal{A}_2 | \bar{\mathcal{B}}$ -messbar, so sind X_{ω_1} und X_{ω_2} $\mathcal{A}_2 \bar{\mathcal{B}}$ bzw. $\mathcal{A}_1 | \bar{\mathcal{B}}$-messbar.

Beweis: Für $B \in \bar{\mathcal{B}}$ gilt:

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

nach Lemma 3.9
Sei nun $X : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mathcal{A}_1 \otimes \mathcal{A}_2$-messbar und nichtnegativ. Nach Lemma 3.10 existieren die Integrale:

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

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

3.11 Lemma

Die Abbildungen $\omega_1 \to \int X_{\omega_1} \,\mathrm{d}\mu_2$ und $\omega_2 \to \int X_{\omega_2} \,\mathrm{d}\mu_1$ sind $\mathcal{A}_1 | \bar{\mathcal{B}}$ bzw. $\mathcal{A}_2 | \bar{\mathcal{B}}$-messbar, sofern μ_1 und μ_2 σ-endlich sind.
Beweis:

(1) $X = \mathbb{1}_Q, Q \in \mathcal{A}_1 \otimes \mathcal{A}_2,\ \mu_1, \mu_2$ endlich;
Sei $S_Q(\omega_1) = \int (\mathbb{1}_Q)_{\omega_1} \,\mathrm{d}\mu_2 = \mu_2(Q_{\omega_1})$
Sei $\mathcal{D} = \left\{Q \in \mathcal{A}_1 \otimes \mathcal{A}_2 | \omega_1 \to \mu_2(Q_{\omega_1}) \text{ist} \mathcal{A}_1 | \bar{\mathcal{B}} \text{messbar}\right\}$ Dann ist $\mathcal{D}$ ein Dynkin-System, denn:

- $S_{\Omega_1 \times \Omega_2} = \mu(\Omega_2)$ ist konstant als Funktion in ω_1
- $S_{\Omega_1 \times \Omega_2} \backslash Q = S_{\Omega_1 \times \Omega_2} - S_Q$ ist messbar. Also $Q^C \in \mathcal{D}$ für $Q \in \mathcal{D}$
- $Seien (Q_n)_{n\in\mathbb{N}}$ in $\mathcal{D}$ paarweise disjunkt. Dann:

$$S_{\bigcup_{n\in\mathbb{N}}} = \sum_{n\in\mathbb{N}} S_{Q_n} \text{ ist messbar. Also: } \bigcup_{n\in\mathbb{N}} Q_n \in \mathcal{D}$$

Ferner: $A_1 \times A_2 \in \mathcal{D} f \ddot{u} r A_1 \in \mathcal{A}_1 und A_2 \in \mathcal{A}_2$, da $S_{A-1 \times A_2} = \mathbb{1}_{A_1} \mu(A_2)$

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

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

ist $\mathcal{A}_1|\bar{\mathcal{B}}$-messbar als abzählbares System von messbaren Funktionen.

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

(4) X nicht negativ: Sei $\left(X^{(n)}\right)_{n\in\mathbb{N}}$ eine nichtfallende Folge in $\mathcal{E}(\Omega_1 \times \Omega_2)$ mit $\sup_{n\in\mathbb{N}} X^{(n)} = X$
Da $(\mathbb{1}_Q)_{\omega_1} = \mathbb{1}_{Q_{\omega_1}} = \mathbb{1}_{Q_{\omega_1}}$ ist $\left(X^{(n)}_{\omega_1}\right)_{n\in\mathbb{N}}$ eine nichtfallende Folge in $\mathcal{E}(\Omega_2)$ mit $\sup X^{(n)}_{\omega_1} = X_{\omega_1}$.
Also:

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

ist $\mathcal{A}|\bar{\mathcal{B}}$-messbar als abzählbares Supremum messbarer Funktionen unter Verwendung von Schritt 3.

Sind μ_1 und μ_2 σ-endlich und X $\mathcal{A}_1 \otimes \mathcal{A}_2|\bar{\mathcal{B}}$-messbar und nicht-negativ, so können wir nach Lemma 3.35 3.11 die iterierten Integrale

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

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

3.12 <u>Satz</u>

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

$$\mu(A_1 \times A_2) = \mu(A_1)\mu_2(A_2) \ \forall\ A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 (*)$$

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:

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

Beweis:
Nach Lemma 3.35 3.11 ist $\pi(Q) := \int \mu_2(Q_{\omega_1})\mu_1(\mathrm{d}\omega_1) \ \forall\ Q \in \mathcal{A}_1 \otimes \mathcal{A}_2$ definiert.
Wir zeigen, dass π ein Maß ist:

- $\pi(Q) \geq 0$: klar
- $\pi(\emptyset) = 0$, da $Q_{\omega_1} = \emptyset$ und also $\mu_2(Q_{\omega_1}) = 0$
- σ-additiv: Seien $(Q_n)_{n\in\mathbb{N}}$ in $\mathcal{A}_1 \otimes \mathcal{A}_2$ paarweise disjunkt. Dann folgt:

$$\mu_2\left(\left(\bigcup_{n\in\mathbb{N}} Q_n \right)_{\omega_1} \right) = \mu_2\left(\bigcup_{n\in\mathbb{N}} (Q_n)_{\omega_1} \right) \underset{\sigma\text{-Additivität}}{=} \sum_{n\in\mathbb{N}} \mu_2\left((Q_n)_{\omega_1}\right)$$

Mit dem Satz von Beppo-Levi folgt:

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

ebenso folgt, dass $\bar{\pi}$ definiert durch

$$\bar{\pi}(Q) = \int \mu_1(q_{\omega_2})\mu_2(\mathrm{d}\omega_2)$$

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 Eindeutigkeit aus Satz 1.21 1.21 (ii)

3.13 Satz(Fubini)

Seien $(\Omega, \mathcal{A}, \mu)$, $i = 1, 2$ σ-endliche Maßräume (d.h. μ_1, μ_2 σ-endlich) und $X : \Omega_1 \times \Omega_2 \to \bar{\mathbb{R}}$ $\mathcal{A}_1 \otimes \mathcal{A}_2 | \bar{\mathcal{B}}$-messbar. Dann:

(i) ist X nichtnegativ, so:

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

(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

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

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

Beweis:

(i) Für nichtnegative Treppenfunktionen: Satz 3.36 und linearität der Integrale. Sei X nichtnegativ und $(X_n)_{n \in \mathbb{N}}$ eine nichtfallende Folge in $\mathcal{E}(\Omega_1 \times \Omega_2)$ mit $X = \sup_{n \in \mathbb{N}} X^{(n)}$.
Dann ist $\left(X_{\omega_2}^{(n)} \right)_{n \in \mathbb{N}}$ approximierend für X_{ω_2} und der Satz von Beppo-Levi liefert:

$$\int \int X_{\omega_2} \,\mathrm{d}\mu_1 \,\mu(\mathrm{d}\omega_2) = \int \left(\sup_{n \in \mathbb{N}} \int X_{\omega_2}^{(n)} \,\mathrm{d}\mu_1 \right) \mu_2(\mathrm{d}\omega_2)$$
$$= \sup_{n \in \mathbb{N}} \int \int X_{\omega_2}^{(n)} \,\mathrm{d}\mu_1 \,\mu_2(\mathrm{d}\omega_2) = \sup_{n \in \mathbb{N}} \int X^{(n)} \,\mathrm{d}(\mu_1 \otimes \mu_2) = \int X \,\mathrm{d}(\mu_1 \otimes \mu_2)$$

(ii) Teil (i) liefert:

$$\int \int |X_{\omega_1}| \,\mathrm{d}\mu_2 \,\mu_1(\mathrm{d}\omega_1) = \int \int |X_{\omega_2}| \,\mathrm{d}\mu_1 \,\mu_2(\mathrm{d}\omega_2) = \int |X| \,\mathrm{d}(\mu_1 \otimes \mu_2) < \infty$$

Also impliziert Folgerung 3.23 **??**:

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

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

$$\omega_1 \to \int X_{\omega_1} \,\mathrm{d}\mu_2 = \int \left(X^+ \right)_{\omega_1} \mathrm{d}\mu_2 - \int \left(X^- \right)_{\omega_1} \mathrm{d}\mu_2$$

ist μ-fast überall definiert und nach Lemma 3.35 3.11 $\mathcal{A}|\bar{\mathcal{B}}$-messbar. Aus (*) und Satz 3.10 **??** folgt, dass diese Funktion μ_1-integrierbar ist, und Anwendung von Teil (i) auf Positiv- und Negativteil:

$$\int \int X_{\omega_1} \,\mathrm{d}\mu_2 \,\mu_1 \,\mathrm{d}\omega_1 = \int \int \left(X^+ \right)_{\omega_1} \mathrm{d}\mu_2 \,\mu_1 \,\mathrm{d}\omega_1 - \int \int \left(X^- \right)_{\omega_1} \mathrm{d}\mu_2 \,\mu_1 \,\mathrm{d}\omega_1$$
$$= \int X^+ \,\mathrm{d}\mu_1 \otimes \mu_2 - \int X^- \,\mathrm{d}\mu_1 \otimes \mu_2 = \int X \,\mathrm{d}\mu_1 \otimes \mu_2$$

3.14 Bemerkung: Kurzschreibweise für Doppelintegrale

$$\int \left(\int X \, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 := \int \int X_{\omega_2} \, \mathrm{d}\mu_1 \, \mu_2(\mathrm{d}\omega_2)$$

3.15 Bemerkungen

Seien $(\Omega_i, \mathcal{A}_1, \mu_i), i = 1, ..., n$ σ-endliche Maßräume. Induktiv sieht man, dass ein eindeutig bestimmtes Maß

$$\mu : \mathcal{A}_1 \otimes \mathcal{A}_2 \otimes ... \otimes \mathcal{A}_n \to [0, \infty]$$

mit

$$\mu(A_1 \times ... \times A_n) = \prod_{i=1}^{n} \mu(A_i)$$

Es heißt Produktmaß von $\mu_1, ..., \mu_n$ und wird mit $\mu_1 \otimes ... \otimes \mu_n$ notiert. Der Maßraum

$$\left(\prod_{i=1}^{n} \Omega_i, \bigotimes_{i=1}^{n} \mathcal{A}_i, \bigotimes_{i=1}^{n} \mu_i \right) \text{ ist der Produktraum von } (\Omega_i, \mathcal{A}_1, \mu_i), i = 1, ..., n$$

3.4: Varianz, Kovarianz und Faltung

In diesem Abschnitt sei ein W-Raum $(\Omega, \mathcal{A}, P)$ gegeben.

3.16 Definition

Seien X und Y numerische Zufallsvariablen.

(i) Ist X P-quasi-integrierbar, so wird

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

der Erwartungswert von X unter P genannt.

(ii) Ist X P-integrierbar, dann heißt

$$\mathrm{Var}(X) := E\left[(X - E[X])^2 \right]$$

die Varianz von X und $\sqrt{\mathrm{Var}(X)}$ die Standardabweichung von X.

(iii) Sind X, Y und $X \cdot Y$ P-integrierbar, so heißt

$$\mathrm{Cov}(X, Y) := E[(X - E[X])(Y - E[Y])]$$

die Kovarianz von X und Y. Ist $\mathrm{Cov}(X, Y) = 0$, so nennt man X und Y unkorreliert.

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

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

die Variation von X und Y genannt.

(v) Ist $X \in L^2(\Omega, \mathcal{A}, P)$ für ein $n \in \mathbb{N}$, so heißt $E[X]^2$ das n-te Moment vom X und $E[|X|^n]$ das n-te Absolutmoment von X.

3.17 Satz

Seien X und Y reellwertige Zufallsvariablen.

(i) $X \in L^2(P) \Leftrightarrow X$ integrierbar und $\mathrm{Var}(X) < \infty$
In diessem Fall gilt:
  (a) $\mathrm{Var}(X) = E\left[X^2\right] - E[X]^2$
  (b) $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$, $a, b \in \mathbb{R}$

(ii) Sind $X, Y \in L^2(P)$ mit $\mathrm{Var}(X) \neq 0 \neq \mathrm{Var}(Y)$, so:

$$|\mathrm{Corr}(X,Y)| \leq 1 \text{ und}$$
$$|\mathrm{Corr}(X,Y)| = 1 \;\Leftrightarrow\; \exists (a,b) \in \mathbb{R}^2 \text{ mit } a \neq 0 : Y = aX + b \; P-\text{fast sicher.}$$

Beweis:

(i) Sei $X \in L^2(P)$, so ist $X \in L^1(P)$ nach Folgerung 3.25 3.1. Also ist X P-integrierbar und die Konstante $E[X] \in L^2(P)$, da P endlich ist. Aufgrund der Mirkowski-Ungleichung ist dann $X - E[X] \in L^2(P)$, also $\mathrm{Var}(X) < \infty$.
Ist X integrierbar und $\mathrm{Var}(X) < \infty$, so sind $E[X] \in L^2(P)$ und $X - E[X] \in L^2(P)$.
Also: $X = (X - E[X]) + E[X] \in L^2(P)$.
  (a) folgt aus der Linearität des Erwartungswertes.
  (b) $\mathrm{Var}(aX + b) = E\left[((aX + b) - E[aX + b])^2\right] = E\left[a^2((X - E[X])^2\right] = a^2 \mathrm{Var}(X)$

(ii) Seien $\xi := \frac{X - E[X]}{\sqrt{\mathrm{Var}(X)}}$, $\eta := \frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}}$.
Dann $E[\xi] = E[\eta] = 0$ und $\mathrm{Var}(\xi) = \mathrm{Var}(\eta) = 1$.
Somit:

$$\mathrm{Corr}(X,Y) = E[\xi\eta] = E\left[\frac{1}{2}\left(\xi^2 + \eta^2\right) - \frac{1}{2}(\xi - \eta)^2\right] = 1 - \frac{1}{2} E\left[(\xi - \eta)^2\right]$$

Da $0 \leq E\left[(\xi - \eta)^2\right] \leq 2E\left[\xi^2 + \eta^2\right] = 4$, folgt:

- $\mathrm{Corr}(X,Y) \in [-1, 1]$
- $\mathrm{Corr}(X,Y) = 1 \Leftrightarrow \xi = \eta$ P-fast sicher in diesem Fall gilt:

$$Y = \frac{\sqrt{\mathrm{Var}(Y)}}{\sqrt{\mathrm{Var}(X)}} X - \frac{\sqrt{\mathrm{Var}(Y)}}{\sqrt{\mathrm{Var}(X)}} E[X] + E[Y]$$

- $\mathrm{Corr}(X,Y) = -1 \Leftrightarrow 2\left(\xi^2 + \eta^2\right) - (\xi - \eta)^2 = 0$ P-fast sicher
In diesem Fall gilt:

$$Y = -\frac{\sqrt{\mathrm{Var}(Y)}}{\sqrt{\mathrm{Var}(X)}} + \frac{\sqrt{\mathrm{Var}(Y)}}{\sqrt{\mathrm{Var}(X)}} E[X] + E[Y]$$

Ist $Y = aX + b$ mit $a \neq 0$, so $\mu = \frac{a}{|a|}$, da $E[Y] = aE[X] + b$ und $\sqrt{\mathrm{Var}(Y)} = |a|\sqrt{\mathrm{Var}(X)}$
Also: $\mathrm{Corr}(X,Y) = E[\eta\xi] = \frac{a}{|a|}$.

3.18 Satz

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

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

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

(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. $\operatorname{Corr}(X_i, X_j) = 0 \; \forall \; i \neq j$

Beweis:

Sei P_X die Verteilubng der n-dimensionalen Zufallsvariable $X = \begin{pmatrix} X_1 \\ . \\ . \\ . \\ X_n \end{pmatrix}$ und P_{X_i} die Verteilung von $X_i, i = 1, ..., n$ Unabhängigkeit bedeutet.

$$P_X = P_{X_1} \otimes ... \otimes PX_n$$

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

$$E\left[\prod_{i=1}^n X_i\right] = \int x_1...x_n P_X(\mathrm{d}(x_1...x_n) = \int \cdots \int x_1...x_n P_{X_1}(\mathrm{d}x_1)...P_{X_n}(\mathrm{d}x_n) = \prod_{i=1}^n \int x_i \, \mathrm{d}P_{X_i}(\mathrm{d}x_i) = \prod_{i=1}^n E[X_i]$$

(ii) Mit $X_1, ..., X_n$ sind auch $|X_1|, ..., |X_n|$ unabh. (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.

3.19 Bemerkung

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

3.20 Satz(Bienaymée)

Sind $X_1, ..., X_n$ paarweise unkorreliert, so gilt:

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

Beweis:

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

3.21 Definition

Seien X und Y unabhängige $\mathbb{R}^D$-wertige Zufallsvariablen mit Verteilungen P_X und P_Y.
Dann wird die Verteilung von $X + Y$ die Faltung von P_X und P_Y genannt und mit $P_X * P_Y$ notiert.

3.22 Satz(Faltungsformel)

Seien X und Y unabhängige $\mathbb{R}^D$-wertige Zufallsvariablen und $A \in \mathcal{B}_D$. Dann

$$(P_X * P_Y)(A) = \int P_X(-y + A) P_Y(\mathrm{d}y) = \int P_Y(-x + A) P_X(\mathrm{d}x)$$

haben P_X und P_Y Dichten f_X und f_Y bzgl $\lambda^{\otimes D}$, so ist

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

Dichte von $P_X * P_Y$ bzgl $\lambda^{\otimes D}$

Beweis: Sei $P_{(X,Y)}$ die Verteilung von $\begin{pmatrix} X \\ Y \end{pmatrix}$. Aufgrund der Unabhängigkeit von X und Y ist $P_{(X,Y)} = P_X \otimes P_Y$. Also folgt aud dem Transformationssatz und dem Satz von Fubini:

$$P_X * P_Y(A) = \int \mathbb{1}_{\{x+y \in A\}} P_{(X,Y)}(\mathrm{d}(x,y)) = \int \Big(\int \mathbb{1}_{\{x+y\in A\}} P_Y(\mathrm{d}y) \Big) P_X(\mathrm{d}x) = \int P_Y(-x+A) P_X(\mathrm{d}x)$$

Haben P_X und P_Y Dichten f_X und f_Y bzgl $\lambda^{\otimes D}$, so folgt:

$$\begin{aligned}(P_X * P_Y)(A) &= \int \Big(\int_{x+A} f_Y(y) \lambda^{\otimes D}(\mathrm{d}y) \Big) f_X(x) \lambda^{\otimes D}(\mathrm{d}x) = \int \Big(\int_A f_Y(y-x) \lambda^{\otimes D}(\mathrm{d}y) \Big) f_X(x) \lambda^{\otimes D}(\mathrm{d}x) \\ &\underset{\text{Fubini}}{=} \int_A \Big(\int_{\mathbb{R}^D} f_X(y-x) f_X(x) \lambda^{\otimes D}(\mathrm{d}x) \Big) \lambda^{\otimes D}(\mathrm{d}y)\end{aligned}$$

Also ist (*) als Funktion in y Dichte von $P_X * P_Y$ bzgl $\lambda^{\otimes D}$.

3.23 Beispiel

Sind $X_1, ..., X_n$ unabhängig und Bernoulli(p)-verteilt, so ist $\sum_{i=1}^n X_i$ binomial(n, p)-verteilt (Übung 1.1)

4 4. Folgen von Zufallsvariablen

4.1 Konvergenzarten von Zufallsvariablen

4.1 Definition

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

(i) $(X_n)_{n\in\mathbb{N}}$ konvergiert gegen X in Wahrscheinlichkeit, falls für alle $\varepsilon > 0$

$$P(\{|X_n - X| > \varepsilon\}) \to 0 \ (n \to \infty)$$

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

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

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

$$E[|X_n - X|^p] \to 0 \ (n \to \infty) \quad (1 \le p < \infty)$$

4.2 Definition

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

$$E^{P_n}[f(X_n)] \to E^P[f(X)] \ (n \to \infty)$$

Der folgende Zusammenhang besteht zwischen diesen Konvergenzarten:

4.3 Satz

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

$$\begin{gathered} X_n \to X P\text{-fast sicher (f.s.)} \\ \underset{\text{(iii)}}{\Rightarrow} X_n \to X \text{in Wahrscheinlichkeit (Wkt)} \\ \underset{\text{(iv)}}{\Rightarrow} X_n \to X \text{in Verteilung} \end{gathered}$$

$$\begin{gathered} X_n \to X \text{im } q\text{-ten Mittel, } q \geq p \\ \underset{\text{(i)}}{\Rightarrow} X_n \to X \text{im p-ten Mittel} \\ \underset{\text{(ii)}}{\Rightarrow} X_n \to X \text{in Wahrscheinlichkeit (Wkt)} \\ \underset{\text{(v)}}{\Rightarrow} \exists\, (n_k)\text{Teilfolge: } X_{n_k} \to X P - \text{f.s.} \end{gathered}$$

Zur Vorbereitung benötigen wir:

4.4 Lemma(Borel-Cantell)

Sei $(\Omega, \mathcal{A}, P)$ und $(A_n)_{n\in\mathbb{N}}$ eine Folge in $\mathcal{A}$. Dann:

(i)

$$\sum_{i=1}^{\infty} P(A_i) < \infty \Rightarrow P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 0$$

(ii)

$$\sum_{i=1}^{\infty} P(A_i) = \infty \text{ und } (A_n)_{n\in\mathbb{N}} \text{ unabhängig } \Rightarrow P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = 1$$

Beweis:

(i) Die Stetigkeit des W-Maßes von oben impliziert:

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

(ii) Mit $(A)_{i\in\mathbb{N}}$ ist auch $\left(A^C\right)_{i\in\mathbb{N}}$ unabhängig. Insbesondere gilt für alle $k \geq 0$ und $n \in \mathbb{N}$:

$$P\left(\bigcap_{m=n}^{n+k} A_m^C\right) = \prod_{m=n}^{n+k} P\left(A_m^C\right)$$

Damit:

$$P\left(\bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty}\right) = \lim_{n\to\infty} \lim_{k\to\infty} P\left(\bigcap_{m=n}^{n+k} A_m\right)$$

$$= \lim_{n\to\infty} \lim_{k\to\infty} \left(1 - P\left(\bigcap_{m=n}^{n+k} A_m\right)\right) = \lim_{n\to\infty} \lim_{k\to\infty} \left(1 - \prod_{m=n}^{n+k} P\left(A_m^C\right)\right)$$

falls $P(A_m) = 1$ für unendlich viele $m \in \mathbb{N}$, so ist die Aussage klar.
anderenfalls gilt für hinreichend großes n und $k \geq 0$ wegen $\log(1-x) \leq -x$

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

Also:

$$\forall\, n \in \mathbb{N} : \lim_{k\to\infty} \prod_{m=n}^{n+k} P\big(A_m^C\big) = 0. \text{ Insbesondere folgt: } P\left(\bigcap_{n\in\mathbb{N}} \bigcup_{k=n}^{\infty} A_k\right) = 1$$

4.5 Satz

(i) Markov-Ungleichung: Für $p \in [1, \infty)$ und $\varepsilon > 0$ gilt:

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

(ii) Tschelbyscheff-Ungleichung: Ist X P-integrierbar und $\varepsilon > 0$, so gilt:

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

Beweis:

(i) $P(\{|X| \geq \varepsilon\}) = E\big[\mathbb{1}_{\{|X|\geq\varepsilon\}}\big] \leq E\Big[\frac{|X|^p}{\varepsilon^p}\Big] = \frac{E[|X|^p]}{\varepsilon^p}$

(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, $q \geq p$. Zz: $X_n \to X$ in L^p.
Mit der Hölder-Ungleichung folgt (Beweis wie in Folgerung 3.1):

$$0 \leq E[|X_n - X|^p]^{\frac{1}{p}} \leq E[|X_n - X|^q]^{\frac{1}{q}} \to 0 \ (n \to \infty)$$

(ii) $X_n \to X$ in L^p. Zz: $X_n \to X$ in Wahrscheinlichkeit.
Sei $\varepsilon > 0$. Dann:

$$P(\{|X_n - X| \geq \varepsilon\}) \underset{\text{Markov-Ungleichung}}{\leq} \frac{E[|X_n - X|^p]}{\varepsilon^p} \to 0 \ (n \to \infty)$$

(iii) $X_n \to X$ P-fast sicher. Zz: $X_n \to X$ in Wkt.
Dazu:

$$\begin{aligned}\{\omega | X_n(\omega) \to X(w)\} &= \left\{\omega |\ \forall\, k \in \mathbb{N}\ \exists\, N_0 \in \mathbb{N}\ \forall\, n \geq N_0 : |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\} \\ &= \bigcap_{k\in\mathbb{N}} \bigcup_{N_0\in\mathbb{N}} \bigcap_{n=N_0}^{\infty} \left\{\omega : |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}\end{aligned}$$

Also:

$$\begin{aligned} & X_n \to X \ P\text{-fast sicher} \\ \Leftrightarrow\ & \forall\, k \in \mathbb{N} : P\left(\bigcup_{N_0\in\mathbb{N}} \bigcap_{n\geq 0} \left\{\omega : |X_n(\omega) - X(\omega)| \leq \frac{1}{k}\right\}\right) = 1 \\ \Leftrightarrow\ & \forall\, k \in \mathbb{N} : \lim_{N_0\to\infty} P\left(\bigcap_{n=N_0} \left\{|X_n - X| \leq \frac{1}{k}\right\}\right) = 1 \\ \Leftrightarrow\ & \forall\, k \in \mathbb{N} : \lim_{N_0\to\infty} P\left(\left\{\sup_{n\geq N_0} |X_n - X| \leq \frac{1}{k}\right\}\right) = 1 \\ \Leftrightarrow\ & \forall\, \varepsilon > 0 : \lim_{N_0\to\infty} P\left(\left\{\sup_{n\geq N_0} |X_n - X| \leq \varepsilon\right\}\right) = 1 \end{aligned}$$

Also: $X_n \to X$ P-fast sicher $\Rightarrow\ \forall\, \varepsilon > 0 : \lim_{N\to\infty} P(\{|X_N - X| \leq \varepsilon\}) = 1$
$\Rightarrow\ \forall\ \lim_{N\to\infty} P(\{|X_N - X| > \varepsilon\}) = 0 \Rightarrow X_n \to X$ in Wahrscheinlichkeit.

(iv) $X_n \to X$ in Wahrscheinlichkeit. Z.z.: $X_n \to X$ 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 vorgegebenes $\varepsilon > 0$:

$$P(\{|X| > N\}) \leq \frac{\varepsilon}{8C}$$

Sei δ so gewählt, dass $|f(x) - f(y)| \leq \frac{\varepsilon}{4}$ für alle $x, y \in [-N, N]$ mit $|x - y| \leq \delta$. Dann:

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

Also:

$$E[|f(X_n) - f(X)|] \leq \varepsilon \text{ für hinreichend großes } n \geq N_0$$

(v) $X_n \to X$ in Wahrscheinlichkeit. Z.z.: $\exists$ Teilfolge (n_k) mit $X_{n_k} \to X$ P-fast sicher.
Setze $n_1 = 1$. Gegeben n_{k-1} sei $n_k := \inf\big\{n \geq n_{k-1} + 1 | P\big(\big\{|X_\nu - X_\mu| > 2^{-k}\big\}\big) \text{ für alle } \nu, \mu \geq n\big\}$
Da

$$P\big(\big\{|X_\nu - X_\mu| > 2^{-k}\big\}\big) \geq P\Big(\Big\{|X_\nu - X| > 2^{-(k+1)}\Big\}\Big) + P\Big(\Big\{|X_\mu - X| > 2^{-(k+1)}\Big\}\Big)$$

und $X_n \to X$ in Wahrscheinlichkeit, folgt, dass $n_k < \infty$. Dann:

$$\sum_{k=1}^{\infty} P\big(\big\{\big|X_{n_{k+1}} - X_{n_k}\big| > 2^{-k}\big\}\big) \leq \sum_{k=1}^{\infty} 2^{-k} < \infty$$

Das Lemma von Borel-Cantelli impliziert, dass

$$P\big(\big\{\big|X_{n_{k+1}} - X_{n_k}\big| > 2^{-k} \text{ für unendlich viele } k \in \mathbb{N}\big\}\big) = 0$$

Sei $\mathcal{N} := \big\{\sum_{k=1}^{\infty} \big|X_{n_{k+1}} - X_{n_k}\big| = +\infty\big\}$. Dann: $P(\mathcal{N}) = 0$. Sei

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

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

Im Beweis von Teil (v) wurde mitgezeigt:

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.

$$\forall\, \varepsilon > 0, \delta > 0 \;\exists\, N \in \mathbb{N} \;\forall\, n, m \geq N P(|X_n - X_m| > \varepsilon) \leq \delta$$

Für fast sichere Konvergenz gilt:

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 P-fast sicher (gegen eine reelle Zufallsvariable)

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

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

(iii) Für jedes $\varepsilon>0$ gilt:

$$\lim_{n\to\infty} P\left(\left\{\sup_{k\geq 1}|X_{n+k}-X_n|\geq\varepsilon\right\}\right)=1$$

Beweis:

- $(i)\Leftrightarrow(ii)$: klar (Analysis 1 und Messbarkeit des Grenzwerts)
- $(ii)\Leftrightarrow(iii)$: 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:

$$P(\{|X_n-0|>\varepsilon\})=P(\{X_n=1\})=P_n$$

- $X_n\to 0$ P-fast sicher $\Leftrightarrow \sum_{n=1}^{\infty}P_n<\infty$ Denn nach dem Lemma von Borel-Cantelli gilt:

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

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 P-fast sicher.

(ii) Sei $(\Omega,\mathcal{A},P)=\big(\mathbb{R},\mathcal{B},\lambda_{[0,1]}\big)$, wobei $\lambda_{[0,1]}$ 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 P-fast sicher und somit auch in Wahrscheinlichkeit.
Die Folge konvergiert nicht in $L^p(\Omega,\mathcal{A},P)$ für $p\geq 1$, da dann nur $X=0$ als Grenzwert in Frage käme, aber

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

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 $\frac{1}{n}\sum_{i=1}^{n}x_i$ der Beobachtung typischerweise in der Nähe des Erwartungswertes des Einzelexperiments. Denken wir uns die Beobachtungen x_i als Realisierungen von unabhängigen und identisch verteilten Zufallsvariablen $(X_i)_{i=1,\dots,n}$, so sollte also in einer geeigneten Form gelten: $\frac{1}{n}\sum_{i=1}^{n}X_i \to E[X_i]$ Hierbei heißt eine Familie von Zufallsvariablen $(X_i)_{i\in I}$ identisch verteilt, falls alle X_i diesselbe Verteilung haben, d.h. $P_{X_i}=P_{X_j}$ für alle $i,j\in I$.

4.9 Definition

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

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

*Hier ist die Grenzfunktion konstant, also ist Konvergenz in Wahrscheinlichkeit äquivalent zu Konvergenz in Verteilung)

4.10 Satz(Schwaches Gesetz der Großen Zahlen)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von paarweise unkorrelierten Zufallsvariablen in $L^2(\Omega,\mathcal{A},P)$ mit

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

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

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

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

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

Beweis (wichtig!):
Aufgrund der Tschebyscheff-Ungleichung und des Satzes von Bienaymée gilt für $\varepsilon > 0$:

$$\begin{aligned} P\left(\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i - E[X_i]\right| \geq \varepsilon\right\}\right) &= P\left(\left\{\left|\sum_{i=1}^{n}X_i - E\left[\sum_{i=1}^{n}X_i\right]\right| \geq n\varepsilon\right\}\right) \\ &\leq \frac{\operatorname{Var}(\sum_{i=1}^{n}X_i)}{n^2\varepsilon^2} = \frac{1}{\varepsilon^2}\left(\frac{1}{n^2}\operatorname{Var}(X_i)\right) \to 0 \ (n\to\infty) \text{ wegen (W)} \end{aligned}$$

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

4.11 Lemma

Sie $(x_n)_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}$.

(i) (Töplitz): Sei $b_n = \sum_{i=1}^{n} a_i$ 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 $\frac{1}{b_n}\sum_{i=1}^{n}a_ix_i \to x$

(ii) (Kronecker): Sei $(b_n)_{n\in\mathbb{N}}$ eine nichtfallende Folge von reellen Zahlen mit $b_n \nearrow +\infty$. Falls $\sum_{i=1}^{\infty}x_n$, so gilt: $\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=1}^{n}b_ix_i = 0$.

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 $\frac{1}{b_n}\sum_{i=1}^{N_0}a_ix_i \leq \frac{\varepsilon}{2}$.
Dann gilt für alle $n \geq N_1$:

$$\left|\frac{1}{b_n}\sum_{i=1}^{n}x_ia_i - x\right| \leq \frac{1}{b_n}\sum_{i=1}^{n}a_i|x_i - x| \leq \frac{1}{b_n}\sum_{i=1}^{N_0}a_i|x_i-x| + \frac{1}{b_n}\sum_{i=N_0+1}^{n}a_i\underbrace{|x_i - x|}_{\leq\frac{\varepsilon}{2}} \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\frac{b_n - b_{N_0}}{b_n} < \varepsilon$$

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

$$\sum_{j=1}^{n}b_jx_j = \sum_{i=1}^{n}b_j(S_j - S_{j-1}) = b_nS_n - b_0S_0 - \sum_{i=1}^{n}S_{j-1}(b_j - b_{j-1})$$

also:

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

da nach Teil (i) mit S_n auch $\frac{1}{b_n}\sum_{j=1}^n S_j(b_j - b_j - 1)$ gegen denselben Grenzwert konvergiert. ■

Nächstes Ziel: Herleitung des folgenden starken GGZ:

4.12 Satz: Kolmogorov

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Zufallsvariablen in $L^1(\Omega, \mathcal{A}, P)$. Dann erfüllt $(X_n)_{n\in\mathbb{N}}$ das starke GGZ, falls:

(i) $(X_n)_{n\in\mathbb{N}}$ identisch verteilt, oder:

(ii) $\sum_{i=1}^\infty \frac{\mathrm{Var}(X_i)}{n^2} < \infty$

Zur Vorbereitung:

4.13 Satz

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

(i) (Kolmogorovsche Ungleichung) Für alle $\varepsilon > 0$ gilt:

$$P\left(\left\{\max_{1\le k\le n} |S_k| \ge \varepsilon\right\}\right) \le \frac{E\left[|S_n|^2\right]}{\varepsilon^2}$$

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

Beweis:

(i) Seien

$$A := \left\{\max_{1\le k\le n} |S_k| \ge \varepsilon\right\} \text{ und } A_k := \{\,\forall\, i = 1, \dots, k-1 |S_i| < \varepsilon \text{ und } |S_k| \ge \varepsilon\}, \quad k = 1, \dots, n$$

Dann ist A disjunkte Vereinigung der A_k. Also:

$$\varepsilon^2 P(A) = \varepsilon^2 \sum_{k=1}^n P(A_k) \le \sum_{k=1}^n E\left[\mathbb{1}_{A_k}|S_k|^2\right]$$

Aufgrund der Unabhängigkeit der (X_i) und da $E[X_i] = 0$, folgt:

$$E\left[\mathbb{1}_{A_k} S_k \left(\sum_{i=1}^n X_i\right)\right] = E[\mathbb{1}_{A_k} S_k]\, \underbrace{E\left[\sum_{i=k+1}^n X_i\right]}_{=0} = 0;$$

Also:

$$\begin{aligned} E\left[\mathbb{1}_{A_k} S_k^2\right] &\le E\left[\mathbb{1}_{A_k} S_k^2\right] + 2E\left[\mathbb{1}_{A_k} S_k \left(\sum_{i=k+1}^n X_i\right)\right] + E\left[\mathbb{1}_{A_k}\left(\sum_{i=k+1}^n X_i\right)^2\right] \\ &= E\left[\mathbb{1}_{A_k} S_n^2\right] \end{aligned}$$

Zusammen folgt:

$$\varepsilon^2 \le \sum_{k=1}^n E\left[\mathbb{1}_{A_k} S_n^2\right] = E\left[\mathbb{1}_A S_n^2\right] \le E\left[S_n^2\right]$$

(ii) Aufgrund von Satz 4.7 reicht es zu zeigen, dass für jedes $\varepsilon < 0$:

$$\lim_{n\to\infty} P\left(\left\{\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\right\}\right) = 0$$

Dies folgt aus (i):

$$P\left(\left\{\sup_{k\geq 1} |S_{n+k} - S_n| \geq \varepsilon\right\}\right) = \lim_{N\to\infty} P\left(\left\{\max_{1\leq k\leq n} |S_n + k - S_n| \geq \varepsilon\right\}\right)$$
$$\leq \lim_{N\to\infty} \frac{1}{\varepsilon^2} E\left[\left(\sum_{k=n}^{n+N} X_k\right)^2\right] \underset{\text{Bienaymée}}{=} \lim_{N\to\infty} \frac{1}{\varepsilon^2} \sum_{k=n}^{n+N} E\left[X_k^2\right] = \frac{1}{\varepsilon^2} \sum_{k=n}^{\infty} E\left[X_k^2\right] \to 0 \ (n \to \infty)$$

,da $\sum_{k=1}^{\infty} E\left[X_k^2\right] < \infty$ ■

Beweis Satz 4.12:

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

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

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

$$P(\{|X_n| \geq n\}) = \sum_{n=1}^{\infty} P(\{|X_1| \geq n\}) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(\{k \leq |X_1| < k+1\})$$
$$= \sum_{k=1}^{\infty} kP(\{k \leq |X_1| < k+1\}) \leq \sum_{k=1}^{\infty} E\left[\mathbb{1}_{\{k\leq |X_1| < k+1\}} |X_1|\right] = E[|X_1|] < \infty$$

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 zeiben: Das starke GGZ gilt für $(X_n)_{n\in\mathbb{N}} \Leftrightarrow$ Das starke GGZ gilt für $\left(\tilde{X}_n\right)_{n\in\mathbb{N}}$ Aus Schritt 1 folgt:

$$\frac{1}{n} \sum_{i=1}^{n} X_i - \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \to 0 \ (n \to \infty) \ P\text{-fast sicher}$$

Ferner:

$$\frac{1}{n} \sum_{i=1}^{n} E[X_i] = \frac{1}{n} \sum_{i=1}^{n} E\left[X_i \mathbb{1}_{\{|X_i| \leq i\}}\right] = \frac{1}{n} \sum_{i=1}^{n} E\left[X_1 \mathbb{1}_{\{|X_1| \leq i\}}\right]$$

Da $E\left[X_1 \mathbb{1}_{\{|X_1| \leq n\}}\right] \to E[X_1] = 0$ nach dem Satz über dominierte konvergenz, folgt aus lemma 4.11 (i), dass

$$\frac{1}{n} \sum_{i=1}^{n} e\left[\tilde{X}\right] \to 0$$

<u>Schritt 3</u>: Das starke GGZ gilt für $\left(\tilde{X}_n\right)_{n\in\mathbb{N}}$. Die Folge $\left(\tilde{X}_n\right)_{n\in\mathbb{N}}$ ist unabhängig. Ferner gilt:

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

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

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

Also ist Teil (i) anwendbar. ■

4.14 Beispiel(Ein vorteilhaftes Spiel?)

Sei $(\xi_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen Bernoulli-$\left(\frac{1}{2}\right)$-Experimenten ("Münzwürfe")
Bei einem Anfangskapital von $X_0 = 1$ wird das Kapital X_n nach der n-ten Runde in Abhängigkeit von $(n+1)$-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 $n \geq 1$:

$$X_n = Y_n X_{n-1} = \prod_{i=1}^{n} Y_i$$

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

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

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

Sei nun $\mu := E[\log(Y_1)] = \frac{1}{2}\left(\log\left(\frac{1}{2}\right) + \log\left(\frac{5}{3}\right)\right) < \frac{1}{2}\left(\log\left(\frac{1}{2}\right)\log(2)\right) = 0$

Nach dem Gesetz der großen Zahlen (Satz 4.12) folgt:

$$\frac{1}{n}\sum_{i=1}^{n} \log(Y_i) \to \mu\ P--\text{fast sicher,}$$

$$\text{also: } \sum_{i=1}^{n} \log(Y_i) \to -\infty\ P--\text{fast sicher}$$

Damit gilt aber auhc:

$$X_n = e^{\sum_{i=1}^{n} \log(Y_i)} \to 0\ P--\text{fast sicher.}$$

Hier strebt also 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 Approximationssatz besagt:
zu f existiert eine Folge von Polynomen f_n mit $\sup_{y\in[0,1]} |f_n(y) - f(y)| \to 0 \quad (n \to \infty)$. 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 GGZ konvergiert $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ gegen $E[X_1] = y$ in Wahrscheinlichkeit und also auch in Verteilung (Satz 4.3). Also gilt:

$$E\left[f(\bar{X}_n)\right] \to f(y) \quad (n \to \infty)$$

Aber:

$$f_n(y) := E\left[f(\bar{X}_n)\right] = E\left[f\left(\frac{1}{n}\underbrace{\sum_{i=1}^{n} X_i}_{\text{binomial}(n,y)}\right)\right] = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) y^k (1-y)^{n-k}$$

$f_n(y)$ 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 im Beweis von Satz 4.3 (iv) folgt:

$$\forall\, \varepsilon > 0 \;\exists\, \delta > 0 : \left|E\left[f\left(\bar{X}_n\right)\right] - f(y)\right| \leq \frac{\varepsilon}{2} + 2 \sup_{x\in[0,1]} |f(x)| \cdot P\left(\left\{\left|\bar{X}_n - y\right| > \delta\right\}\right)$$

Da für alle $y \in [0,1]$:

$$P\left(\left\{\left|\bar{X}_n - y\right| > \delta\right\}\right) \leq \frac{\operatorname{Var}\left(\bar{X}_n\right)}{\delta^2} = \frac{ny(1-y)}{\delta^2 n^2} \leq \frac{\frac{1}{4}}{\delta^2 n}$$

(Tschebyscheff-Ungleichung und Varianz der Binomialverteilung)
Also:

$$\forall\, \varepsilon > 0 \;\exists\, \delta > 0 : \sup_{y\in[0,1]} |f_n(y) - f(y)| \leq \frac{\varepsilon}{2} + \frac{1}{2} \sup_{x\in[0,1]} |f(x)| \frac{1}{\delta^2 n}$$

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 $\xi_n = \xi_n(\omega) \in \{0,\ldots,g-1\}$ für alle $n \in \mathbb{N}$ und $\xi \in \{0,\ldots,g-2\}$ für unendlich viele $n \in N$.
Die Zahl ω heißt einfach g-normal, falls für alle $\varepsilon = 0,\ldots,g-1$:

$$\lim_{n\to\infty} \frac{1}{n} S_n^{\varepsilon,g} = \frac{1}{g};\; S_n^{\varepsilon,g} = \#\{i \leq n | \xi_i(\omega) = \varepsilon\}$$

Ist ω einfach g-normal für alle $g \geq 2$, so wird ω absolut normal genannt.
Sei $\Omega = [0,1)$, $\mathcal{A}$ die Spur-σ-Algebra von $\mathcal{B}$ auf $[0,1)$ und P die Einschränkung des Lebesgue-Maßes auf $[0,1)$
Dann gilt für $\varepsilon = 0,\ldots,g-1$ (g beliebig aber fest):

$$\{\xi_i = \varepsilon\} = \bigcup_{k=0}^{g^{i-1}-1} \left[\frac{kg+\varepsilon}{g^i}, \frac{kg+\varepsilon+1}{g^i}\right) =: A_i^\varepsilon$$

Dann:

$$\omega - g^i = g \cdot \left(\sum_{n=1}^{i-1} \xi_n(\omega) g^{i-1-n}\right) + \xi_i(\omega) + \sum_{n=i+1}^{\infty} \xi_n(\omega) g^{i-n}$$

Die Folge von Ereignissen $(A_i^\varepsilon)_{i\in\mathbb{N}}$ ist für festes $\varepsilon = 0,\ldots,g-1$ unabhängig.
(Für $g = 2, \varepsilon = 0$ wurde dies in den Übungen gezeigt. Den allgemeinen Fall überprüft man analog.)
Ferner gilt:

$$P(A_i^\varepsilon) = \sum_{k=0}^{g^{i-1}-1} \frac{1}{g^i} = \frac{1}{g}$$

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 GGZ. Also:

$$\lim_{n\to\infty} S_n^{\varepsilon,g}(\omega) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^\varepsilon(\omega) = E[X_1^\varepsilon] = P(A_1^\varepsilon) = \frac{1}{g}$$

für P-fast sicher $\omega \in [0,1)$.
Dies zeigt, dass Lebesgue-fast alle $\omega \in [0,1)$ absolut normal sind.

4.3. Der zentrale Grenzwertsatz von Lindberg-Levy

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängig identisch verteilten Zufallsvariablen im $L^2(\Omega, \mathcal{A}, P)$. Seien $\mu = E[X_1]$ und $\sigma^2 = \mathrm{Var}(X_i)$. Sei $\sigma^2 > 0$. Dann gilt für alle $\alpha \in \left[0, \frac{1}{2}\right)$:

$$\sum_{n=1}^{\infty} E\left[\left(\frac{X_n - E[X_n]}{n^{1-\alpha}}\right)^2\right] = \sum_{i=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^{2-2\alpha}} = \sigma^2 \sum_{n=1}^{\infty} \frac{1}{n^{2-2\alpha}} < \infty$$

Nach dem Reihenkriterium aus Satz 4.13 (ii) folgt, dass

$$\sum_{n=1}^{\infty} \frac{X_n - E[X_n]}{n^{1-\alpha}}$$

P-fast sicher konvergiert. Also impliziert das Lemma von Kronecker (Lemma 4.11 (ii)), dass

$$\frac{1}{n^{1-\alpha}} \sum_{i=1}^{n} X_i - E[X_i] \to 0 \ P\text{-fast sicher}$$

Frage: Was passiert für $\alpha = \frac{1}{2}$?

4.17 Satz (Zentraler Grenzwertsatz von Lindeberg-Levy)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen in $L^2(\Omega, \mathcal{A}, P)$ mit $\sigma^2 = \mathrm{Var}(X_1) > 0$. Sei $\mu = E[X_1]$. Dann gilt für alle $y \in \mathbb{R}$:

$$P\left(\left\{\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \le y\right\}\right) \to \Phi(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{x^2}{2}}\,\mathrm{d}x \quad (n \to \infty)$$

Diskussion:

(a) Da Φ die Verteilungsfunktion einer Standardnormalverteilung ist, sehen wir, dass die Verteilungsfunktion von $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$ 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) Normierung $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$ ist so gewählt, dass

$$E\left[\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)\right] = 0 \text{ und}$$

$$\mathrm{Var}\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)\right) \underset{\text{Bienaymée}}{=} \frac{1}{\sigma^2 n} \sum_{i=1}^{n} \underbrace{\mathrm{Var}(X_i)}_{=\sigma^2} = 1$$

Für $S_n = \sum_{i=1}^{n} X_i$ gilt also:

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) = \frac{S_n - E[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$$

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 $X_1, \dots, X_{100}$ sind dann Bernoulli-$\frac{1}{2}$-verteilt und unabhängig. Somit ist $S_{100} = \sum_{i=1}^{100} X_i$ dann bernoulli-$\left(100, \frac{1}{2}\right)$-verteilt. Also:

$$P(\{S_{100} \ge 60\}) = \sum_{k=60}^{100} B_{100,\frac{1}{2}}(k) = \sum_{k=60}^{100} \binom{100}{k} \left(\frac{1}{2}\right)^{100} \approx 2{,}84\%$$

Frage: Wie gut ist die Approximation durch den zentralen Grenzwertsatz?

$$P(\{S_{100} \geq 60\}) = 1 - P(\{S_{100} \leq 59\}) = 1 - P\left(\left\{\frac{\sum_{i=1}^{100}(X_i - \frac{1}{2})}{\sqrt{100 \cdot \frac{1}{4}}} \leq \frac{59-50}{\sqrt{100 \cdot \frac{1}{4}}}\right\}\right)$$
$$\approx 1 - \int_{-\infty}^{\frac{9}{5}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \,\mathrm{d}x \approx 3,59\%$$

Da $\sum_{i=1}^n X_i$ für unabhängige Bernoulli(p)-verteilte Zufallsvariablen nur ganzzahlige Werte annehmen kann, hätten wir bei der obigen Rechnung auch jeden Wert im Intervall $[59, 60)$ statt 59 verwenden können. Dem tragen folgende in der Praxis oft verwendete Korrekturformel (Rechnung):
Für $\alpha \in \mathbb{N}$:

$$P\left(\left\{\sum_{i=1}^n X_i \geq \alpha\right\}\right) = P\left(\left\{\sum_{i=1}^n X_i > \alpha - \frac{1}{2}\right\}\right) = P\left(\left\{\frac{\sum_{i=1}^n (X_i - p)}{\sqrt{np(1-p)}} > \frac{\alpha - np - \frac{1}{2}}{\sqrt{np(1-p}}\right\}\right)$$
$$= 1 - \Phi\left(\frac{\alpha - np - \frac{1}{2}}{\sqrt{np(1-p)}}\right)$$

und

$$P\left(\left\{\sum_{i=1}^n X_i \leq \alpha\right\}\right) \approx \Phi\left(\alpha - np + \frac{1}{2}\right) \quad (n \text{ groß})$$

Im konkreten Beispiel ergibt sich:

$$P(\{S_{100} \geq 60\}) \approx 1 - \Phi\left(\frac{60 - 50 - \frac{1}{2}}{\sqrt{np(1-p)}}\right) \approx 2,87\%$$

Zum Vergleich: Die Abschätzung mittels der Tschebyscheff-Ungleichung, auf der schwaches Gesetz der Großen Zahlen basiert, liefert hier:

$$P(\{S_{100} \geq 60\}) \leq P\left(\left\{\left|\frac{1}{100}\sum_{i=1}^{100} X_i - \frac{1}{2}\right| \geq 0,1\right\}\right) \leq \frac{\left(\frac{1}{2}\right)^2}{100 \cdot (0,1)^2} = \frac{1}{4}$$

4.19 Beispiel(Monte-Carlo-Integration)

Sei $g : [0,1]^D \to \mathbb{R}$ stetig. Wir wollen

$$\int_{[0,1]^D} g(x) \lambda^{\otimes D}(\mathrm{d}x) =: \mu$$

numerisch approximieren. Sei $(U_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen auf $[0,1]^D$ gleichverteilten Zufallsvariablen, also mit Dichte $f(x_1, \ldots, x_d) = \prod_{d=1}^D \mathbb{1}_{[0,1]}(x_d)$. Dann:

$$E[g(U_1)] = \int_{\mathbb{R}^D} g(x) f(x) \lambda^{\otimes D}(\mathrm{d}x) = \int_{[0,1]^D} g(x) \lambda^{\otimes D}(\mathrm{d}x)$$

Aus dem starken Gesetz der großen Zahlen folgt:

$$\frac{1}{n}\sum_{i=1}^n g(U_i) \to E[g(U_1)] = \int_{[0,1]^D} g(x) \lambda^{\otimes D}(\mathrm{d}x) \quad P - \text{fast-sicher für } n \to \infty$$

Diese Methode nennt man ein Monte-Carlo-Verfahren.
Frage: Wie schnell konvergiert diese Monte-Carlo-Methode?
Hier hilft der zentrale Grenzwertsatz. Sei

$$\sigma^2 := \mathrm{Var}(b(U_1)) = \int_{[0,1]^{\otimes D}} g^2(x) \lambda^{\otimes} - \mu^2 < \infty$$

Wir nehmen an, dass $\sigma^2 > 0$
Dann gilt für beliebiges $c > 0$:

$$P\left(\left\{\mu \in \left(\frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{c\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{c\sigma}{\sqrt{n}}\right]\right\}\right)$$
$$= P\left(\left\{-c < \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (g(U_i) - \mu) \leq c\right\}\right) \underset{n\to\infty}{\to} \Phi(c) - \Phi(-c) = 2\Phi(c) - 1$$

Ist etwa $c = 1,96$, so ist $\Phi(c) \approx 0,975$. Also:

$$P\left(\left\{\mu \in \left(\frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{1,96\sigma}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^n g(U_i) - \frac{1,96\sigma}{\sqrt{n}}\right]\right\}\right) \to 0,95 \quad (n \to \infty)$$

Der gesuchte Integralwert μ liegt dann also für großes n mit Wahrscheinlichkeit von ca. 95% in dem angegebenen zufälligen Intervall der Länge $\frac{2\cdot 1,96\sigma}{\sqrt{n}}$. Mit wachsender Anzahl n der zu generierenden Zufallsvariablen $U_1, \ldots, U_n$ schrumpft dieses Intervall mit der Ordnung $\frac{1}{\sqrt{n}}$.
Für kleine Dimensionen D sind deterministische Verfahren zur Integralapproximation effizienter. Der Vorteil des Monte-Carlo-Verfahrens liegt darin, dass die Konvergenz nicht von der Dimension abhänngt und also auch für sehr hohe Dimensionen einsetzbar ist.
Zum Beweis des Zentralen Grenzwertsatzes benötigen wir einige Vorbereitungen über Konverrgenz von Wahrscheinlichkeitsmaßen.

4.20 Definition

Sei $(P_n)_{n\in\mathbb{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbb{R}, \mathcal{B})$ und P ein W-Maß auf $(\mathbb{R}, \mathcal{B})$. Dann sagt man:

(i) $(P_n)_{n\in\mathbb{N}}$ konvergiert schwach gegen P, falls für alle stetigen und beschränkten $f : \mathbb{R} \to \mathbb{R}$ gilt:

$$\lim_{n\to\infty} \int f \, \mathrm{d}P_n = \int f \, \mathrm{d}P$$

(Das heißt schwache-$*$-Konvergenz in der Funktionalanalysis)

(ii) $(P_n)_{n\in\mathbb{N}}$ konvergiert im Wesentlichen, falls für alle $A \in \mathcal{B}$ mit $P(\partial A) = 0$

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

Hierbei ist ∂A der Rand von A, also $\partial A = \bar{A} \cap \bar{A}^C$, wobei $\bar{A}$ den Abschluss von A bezeichnet.

(iii) Seien $F_n(x) := P_n((-\infty, x])$ und $F(x) := P((-\infty, x])$ die zugehörigen Verteilungsfunktionen. Dann sagt man, $(F_n)_{n\in\mathbb{N}}$ konvergiert im Wesentlichen gegen F, falls

$$\lim_{n\to\infty} = F(x) \text{ für alle Stetigkeitsstellen } x \text{ von } F.$$

Bemerkung:
Die Aussage des zentralen Grenzwertsatzes von Lindeberg-Levy besagt, dass die Verteilungsfunktion der standartisierten Summe $\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^n (X_i - \mu)$ im Wesentlichen gegen die Verteilungsfunktion Φ der Standardnormalverteilung konvergiert.
Dies ist äquivalent zu den anderen Konvergenzarten aus Definition 4.20:

4.21 Satz (Portmanteau-Theorem)

Seien $(P_n)_{n\in\mathbb{N}}$ und P W-Maße auf $(\mathbb{R}, \mathcal{B})$. $(F_n)_{n\in\mathbb{N}}$ und F seien die zugehörigen Verteilungsfunktionen. Dann sind äquivalent:

(a) $P_n \to P$ schwach

(b) Für alle abgeschlossenen Mangen $A \subset \mathbb{R}$: $\limsup_{n\to\infty} P_n(A) \leq P(A)$

(c) Für alle offenen Mangen $A \subset \mathbb{R}$: $\liminf_{n\to\infty} P_n(A) \geq P(A)$a

(d) $P_n \to P$ im Wesentlichen

(e) $F_n \to F$ im Wesentlichen

Beweis:

- a) $\Rightarrow$ b):
  Sei $\rho(x, A) := \inf\{|x-y|, y \in A\}$ 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 $A_k := \{x \in \mathbb{R} | \rho(x, A) \leq k\}$. Definiere eine stetige und beschränkte Funktion f_k durch $f_k(x) = g(k, \rho(x, A)$. Aus a) folgt:

$$\limsup_{n\to\infty} P_n(A) = \limsup_{n\to\infty} \int \mathbb{1}_A \, \mathrm{d}P_n \leq \limsup_{n\to\infty} \int f_k \, \mathrm{d}P_n \underset{\text{a)}}{=} \int f_k \, \mathrm{d}P \leq \int \mathbb{1}_{A_k} \, \mathrm{d}P = P(A_k)$$

  Die Stetigkeit von P von oben impliziert:

$$\lim_{n\to\infty} P(A_k) = P\left(\bigcap_{k\in\mathbb{N}} A_k\right) = P(A) \text{ , da A abgeschlossen}$$

  Also mit $k \uparrow +\infty$:

$$\limsup_{n\to\infty} P_n(A) \leq P(A).$$

- b) $\Rightarrow$ c):
  Sei A offen. Dann ist A^C abgeschlossen. Also folgt aus b):

$$\liminf_{n\to\infty} P_n(A) = 1 - \limsup_{n\to\infty} P_n\big(A^C\big) \geq 1 - P\big(A^C\big) = P(A)$$

- c) $\Rightarrow$ b) analog

- c) $\Rightarrow$ d):
  Mit c) gilt auch b). Da $A \in \bar{A}$ gilt für beliebiges $A \in \mathcal{B}$:

$$\limsup_{n\to\infty} P_n(A) \leq \limsup_{n\to\infty} P_n\big(\bar{A}\big) \text{ mit b)}$$

  Da Å $\subset A$ (wobei Å das Innere von A ist) folgt aus c):

$$\liminf_{n\to\infty P_n(A) \geq \liminf_{n\to\infty} P_n(\text{Å})}$$

  Sei nun $P(\partial A) = 0$. Da $\bar{A} = \text{Å} \cup \partial A$, erhalten wir:

$$P\big(\text{Å}\big) \leq P(A) \leq P\big(\bar{A}\big) = P\big(\text{Å} P(\partial A) = P\big(\text{Å}\big)\big).$$

  Somit:

$$\liminf_{n\to\infty} P_n(A) \geq P\big(\text{Å}\big) = P(A) = P\big(\bar{A}\big) \geq \limsup_{n\to\infty} P_n(A)$$

  Also:

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

- d) $\Rightarrow$ 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:
$$D = \{t \in \mathbb{R} | P(\{f = t\}) \neq 0\}$$
höchstens abzählbar (Die Mengen $f^{-1}(\{t\}), t \in D$ sind disjunkt und P ist endliches Maß).
Sei Œ$-M, M \in D$. Fixiere $\varepsilon > 0$ und wähle eine Partition $\pi = (t_0, \dots, t_k))$ von $[-M, M]$ (also: $-M = t_0 < t_1 < \dots < t_k = +M$) mit $|pi| := \max_{i=0,\dots,k-1}(t_{i+1} - t_i) \leq \varepsilon$. Diese Partition sei so gewählt, dass $t_i \notin D$ für alle $i = 0, \dots, k-1$.
Seien die Mengen B_i definiert durch:
$$B_i = \{t_i \leq f \leq t_{i+1}\}, i = 0, \dots, k-2, \quad B_{k-1} = \{t_{k-1} \leq f \leq t_k\}$$
Da f stetig ist, ist $f^{-1}((t_i, t_{i+1}))$ offen für alle $i = 0, \dots, k-1$. Somit:
$$P(\partial B_i) \leq P\big(f^{-1}(\{t_i\}) \cup f^{-1}(\{t_{i+1}\})\big) = P(\{f = t_i\}) + P(f = t_{i+1}) = 0$$
Daher liefert d):
$$\sum_{i=0}^{k-1} t_i P_n(B - i) \to \sum_{i=0}^{k-1} t_i P(B_i) \quad (n \to \infty) \quad (*)$$
Somit:
$$\begin{aligned} &\left| \int f \,\mathrm{d}P - \int f \,\mathrm{d}P_n \right| \\ \leq &\left| \sum_{i=0}^{k-1} \int \mathbb{1}_{B_i}(f - t_i) \,\mathrm{d}P_n \right| + \left| \sum_{i=0}^{k-1} t_i P_n(B_i) - \sum_{i=1}^{n} t_i P(B_i) \right| + \left| \sum_{i=0}^{n} \int \mathbb{1}_{B_i}(t_i - f) \,\mathrm{d}P \right| \\ &\leq 2\varepsilon + \left| \sum_{i=0}^{k-1} t_i P_n(B_i) - \sum_{i=0}^{k-1} t_i P(B_i) \right| \leq 3\varepsilon \end{aligned}$$
wegen $(*)$, falls $n \geq N(\varepsilon)$ hinreichend groß.

- d) $\Rightarrow$ e):
Da für $x \in \mathbb{R}$ gilt: $\partial(-\infty, x] = \{x\}$, folgt:
$$P(\partial(-\infty, x]) = P(\{x\}) = 0 (F) \text{ stetig in } x (\text{da } P((-\infty, x]) = F(x-))$$
Somit gilt an allen Stetigkeitsstellen x von F nach d):
$$F_n(x) = P_n((-\infty, x]) \to P((-\infty, x]) = F(x)$$

- e) $\Rightarrow$ c):
Sei A offen. Dann existiert eine abzählbare Menge von paarweise disjunkten Intervallen der Form $I_k = (a_k, b_k), a_k \leq b_k \in \mathbb{R}$, 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:
$$P(I_k) \leq P(I'_k) + \varepsilon 2^{-k}$$
Das Lemma von Fatou mit $\left(\mathbb{N}, 2^{\mathbb{N}}, \mu\right)$ wobei $\mu(A) = \#A$ das Zählmaß ist, liefert:
$$\liminf_{n\to\infty} P_n(A) \underset{\sigma\text{-additiv}}{=} \liminf_{n\to\infty} \sum_{k=1}^{\infty} P_n(I_k) \geq \sum_{k=1}^{\infty} \liminf_{n\to\infty} P_n(I_k) \geq \sum_{i=1}^{\infty} \liminf_{n\to\infty} P_n(I'_k)$$
Da $P_n(I'_k) = F_n(b'_k) - F_n(a'_k) \underset{\text{e)}}{\to} F(b'_k) - F(a'_k) = P(I'_k)$
Also:
$$\liminf_{n\to\infty} P_n(A) \geq \sum_{k=1}^{\infty} P(I'_k) \geq \sum_{k=1}^{\infty} P(I_k) - \varepsilon \sum_{k=1}^{\infty} 2^{-k} \underset{\sigma\text{-additiv}}{=} P(A) - \varepsilon$$
Mit $\varepsilon \downarrow 0$ folgt c) ■

Zum Beweis des zentralen Grenzwertsatzes verwenden wir die Methode der charakteristischen Funktionen.

4.22 Definition

(i) Sei P ein W-Maß auf $\big(\mathbb{R}^D, \mathcal{B}_D\big)$. Dann heißt die Funktion

$$\phi(t) := \int_{\mathbb{R}^D} e^{i<t,x>} P \,\mathrm{d}x \quad , t \in \mathbb{R}^D$$

die charakteristische Funktion von P. Hier ist $< \cdot, \cdot >$ das euklidische Skalarprodukt in $\mathbb{R}^D$, $i = \sqrt{-1}$ die imaginäre Einheit und

$$\begin{aligned}\int_{\mathbb{R}^D} e^{i<t,x>} P(\mathrm{d}x) &= \int \mathrm{Re}\big\{e^{i<t,x>} P(\mathrm{d}x)\big\} + i \int \mathrm{Im}\big\{e^{i<t,x>} P(\mathrm{d}x)\big\} \\ &= \int \cos(<t,x>) P(\mathrm{d}x) + i \int \sin <t,x> P(\mathrm{d}x)\end{aligned}$$

(ii) Für eine $\mathbb{R}^D$-wertige Zufallsvariable X ist

$$\Phi_X(t) = E\big[e^{i<t,X>}\big] = \int_{\mathbb{R}^D} e^{i<t,x>} P_X(\mathrm{d}x), \quad t \in \mathbb{R}^D$$

4.23 Beispiel

(i) Ist X Bernoulli(p)-verteilt,so

$$\phi_X(t) = e^{it} p + (1-p), \quad t \in \mathbb{R}$$

(ii) Sei X standardnormalverteilt. Aus der Übung ist bekannt, dass für alle $n \in \mathbb{N}$:

$$\begin{aligned}E\big[X^{2n}\big] &= \frac{1}{2^n}\frac{(2n)!}{n!} \\ E\big[X^{2n-1}\big] &= 0\end{aligned}$$

also:

$$\begin{aligned}\phi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} \,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(itx)^k}{k!} e^{-}\frac{x^2}{2} \,\mathrm{d}x \\ = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} &= \sum_{k=0}^{\infty} \frac{(it)^{2k}}{(2k)!} \frac{1}{2^k} \frac{(2k)!}{k!} = \sum_{k=0}^{\infty} \left(\frac{(it)^2}{2}\right)^k \frac{1}{k!} = e^{-\frac{t^2}{2}}\end{aligned}$$

Ist X $\mathcal{N}\big(\mu, \sigma^2\big)$-verteilt, so ist $Y = \frac{1}{\sigma}(X - \mu)$ $\mathcal{N}(0,1)$-verteilt. Also:

$$\phi_X = E\Big[e^{it(\sigma Y + \mu)}\Big] = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

(iii) Sei X Poisson(λ)-verteilt. Dann:

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\big(\lambda e^{it}\big)^k}{k!} = \exp\big\{\lambda\big(e^{it} - 1\big)\big\}$$

4.24 Lemma

Sei $X \in L^2(P)$. Dann:

$$\phi_X(t) = 1 + itE[X] - \frac{1}{2}t^2 E\big[X^2\big] - \frac{1}{2}t^2 \varepsilon(t)$$

mit $|\varepsilon(t)| \le 3E\big[X^2\big]$ und $\lim_{t\to 0} \varepsilon(t)$
Beweis:

$$\phi_X(t) = E\big[e^{itX}\big] = E[\cos tx] + iE[\sin tx]$$

existieren (mit Taylorentwicklung von cos bzw. sin um 0) $\theta_1(\omega)$ und $\theta_2(\omega)$ so, dass $|\theta_1(\omega)| \leq 1,\ |\theta_2(\omega)| \leq 1$ und:

$$e^{itX(\omega)} = 1 + itX(\omega) - \frac{1}{2}(\cos\theta_1(\omega)tX(\omega) + i\sin\theta_2(\omega)tX(\omega))X^2(\omega)$$

Sei

$$\varepsilon(t) = E\big[X^2\cos\theta_1 tX + i\sin\theta_2 tX - 1\big]$$

Dann:

$$\phi_X(t) = 1 + itE[X] - \frac{1}{2}t^2E\big[X^2\big] - \frac{1}{2}t^2\varepsilon(t)$$

Offensichtlich gilt: $|\varepsilon(t)| \leq 3E\big[X^2\big]$. Schließlich zeigt der Satz über dominierte Konvergenz, dass $\varepsilon(t) \to 0$ $(t \to 0)$, da

$$\cos\theta_1 tX + i\sin\theta_2 tX - 1 \to 0\ P\text{-fast sicher für } t \to 0$$

4.25 Satz

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen in $L^2(P)$ mit Erwartungswert μ und Varianz $\sigma^2 > 0$. Sei

$$T_n = \frac{1}{\sqrt{n,\sigma^2}}\sum_{k=1}^{n}(X_k - \mu)$$

Dann gilt für alle $t \in \mathbb{R}$

$$\phi_{T_n}(t) \to e^{-\frac{t^2}{2}}\ (n \to \infty)$$

Beweis:
Es gilt:

$$\begin{aligned}\phi_{T_n}(t) &= E\left(e^{i\frac{t}{\sqrt{n}}}\sum_{k=1}^{n}\frac{X_k - \mu}{\sigma}\right) = E\left[\prod_{k=1}^{n} e^{i\frac{t}{\sqrt{n}}\frac{X_k-\mu}{\sigma}}\right] \underset{\text{Unabh.}}{=} \prod_{k=1}^{n} E\left[e^{i\frac{t}{\sqrt{n}}\frac{X_k-\mu}{\sigma}}\right] \\ &= E\left[e^{i\frac{t}{\sqrt{n}}\frac{X_k-\mu}{\sigma}}\right] = \left(\phi_{\frac{X_1-\mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right)\right)^n\end{aligned}$$

Aus Lemma 4.24 folgt mit $X = \frac{X_k-\mu}{\sigma}$ folgt:

$$\phi_{\frac{X_1-\mu}{\sigma}}\left(\frac{t}{\sqrt{n}}\right) = 1 + i\frac{t}{\sqrt{n}}\underbrace{E\left[\frac{X_1-\mu}{\sigma}\right]}_{=0} - \frac{1}{2}\frac{t^2}{n}\underbrace{E\left[\left(\frac{X_1-\mu}{\sigma}\right)^2\right]}_{=1} - \frac{1}{2}\frac{t^2}{n}\varepsilon\left(\frac{t}{\sqrt{n}}\right)$$

FEHLT
Somit folgt der zentrale Grenzwertsatz von Lindeberg-Levy aus folgendem wichtigen Zusammenhang zwischen schwacher Konvergenz und charakteristischen Funktionen

4.26 Satz(Stetigkeitssatz)

Sei $(P_n)_{n\in\mathbb{N}}$ eine Folge von W-Maßen auf $(\mathbb{R}, \mathcal{B})$ und seien $\phi_n, n \in \mathbb{N}$ die zugehörigen charakteristischen Funktionen. Dann:

(i) Konvergiert $(P_n)_{n\in\mathbb{N}}$ schwach gegen ein W-Maß P, so gilt für alle $t \in \mathbb{R}$:

$$\phi_n(t) \to \phi(t)\ (n \to \infty$$

, wobei *phi* die charakteristische Funktion von P ist.

(ii) Konvergiert $(\phi_n(t))_{n\in\mathbb{N}}$ für alle $t \in \mathbb{R}$ gegen $\phi(t)$ und ist ϕ stetig in 0, so ist ϕ charakteristische Funktion eines W-Maßes P und $(P_n)_{n\in\mathbb{N}}$ konvergiert gegen P.

Beweis von Satz 4.17 (über Benutzung von Satz 4.26:
Sei $(P_n)_{n\in\mathbb{N}}$ die Folge von Verteilungen von $\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^n (X_k - \mu)$ und $(\phi_n)_{n\in\mathbb{N}}$ die zugehörige Folge von charakteristischen Funktionen. Nach Satz 4.25 und Beispiel 4.23 konvergiert (ϕ_n) punktweise gegen die charakteristische Funktion der Standardnormalverteilung P. Satz 4.26 (ii) impliziert, dass $P_n \to P$ schwach, was mach dem Portmanteau-Theorem die wesentliche Konvergenz der Verteilungsfunktion nach sich zieht. Dies war zu zeigen.
■
Beweis von Satz 4.26 (i):
$(P_n)_{n\in\mathbb{N}}$ konvergiere schwach gegen P. Da $\mathrm{Re}\{e^{itx}\} = \cos tx$ und $\mathrm{Im}\{e^{itx}\} = \sin tx$ für festes $t \in \mathbb{R}$ als Fnktion in x stetig und beschränkt sind, folgt:

$$\phi_n(t) = \int \cos tx P_n(\mathrm{d}x) + \int \sin tx P_n(\mathrm{d}x) \underset{n\to\infty}{\to} \int \cos tx P(\mathrm{d}x) + \int \sin tx P(\mathrm{d}x) = \phi(t)$$

Beweisstrategie von Satz 4.26:

(1) Sei $(P_n)_{n\in\mathbb{N}}$ eine Folge von W-Maßen auf $(\mathbb{R},\mathcal{B})$ mit zugehörigen Verteilungsfunktionen $(F_n)_{n\in\mathbb{N}}$. Dann liefert der Satz von Helly(Satz **??**):

- $\exists\ (n_k) Teilfolge, G : \mathbb{R} \to \mathbb{R} : F_{n_k}(x) \to G(x)\ (k \to \infty)$für alle Stetigkeitsstellen x von G.
- G ist rechtsstetig und nichtfallend, aber i.A. gilt nur $G(-\infty) \geq 0$ und $G(+\infty) \leq 1$

(2) Die Konvergenz der charakteeristischen Funktion stellt sicher, dass die Folge $(P_n)_{n\in\mathbb{N}}$ "straff" ist (Lemma **??**), so dass G tatsächlich eine Verteilungsfunktion ist (Satz von Prokkorov, Satz **??**)

(3) Aus einem Eindeutigkeitssatz für charakteristische Funktionen (Satz **??**) folgt dann, dass die Konvergenz nicht nur entlang der Teilfolge (F_{n_k}) sondern entlang der Folge (F_n) gilt.

(4) Das Pormanteau-Theorem erlaubt es dann, von der wesentlichen Konvergenz der Verteilungsfunktion auf die schwache Konvergenz des W-Maßes zu schließen.

4.4: Mehr zu charakteristischen Funktionen

4.27 Satz(Umkehrformel)

Sei P W-Maß auf $(\mathbb{R},\mathcal{B})$ mit Verteilungsfunktion F und charakteristischer Funktion ϕ. Dann:

(i) Für alle Stetigkeitsstellen $a < b$ von F:

$$F(b) - F(a) = \lim_{t\to\infty} \frac{1}{2\pi} \int_{[-\sigma,\sigma]} \frac{e^{-ita} - e^{-itb}}{it} \phi(t)\lambda(\mathrm{d}t)$$

(ii) Falls $\int_{\mathbb{R}} |f(t)|\lambda(\mathrm{d}t) < \infty$, so hat F eine Dichte f bzgl λ und

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(t)\lambda(\mathrm{d}t)$$

Beweis:

(i) Seien $a < b$ Stetigkeitsstellen von F. Mit dem Satz von Fubini folgt:

$$\Phi_C := \frac{1}{2\pi}\int_{-C}^{+C} \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)\lambda(\mathrm{d}t)$$
$$= \int_{\mathbb{R}} \left(\underbrace{\int_{-C}^{+C} \frac{1}{2\pi}\frac{e^{-ita} - e^{-itb}}{it} e^{itx}\lambda(\mathrm{d}t)}_{=\Psi_C(x)} \right) P(\mathrm{d}x)$$

FEHLT
Also:

$$\lim_{C\uparrow+\infty} \Psi_C(x) = \left\{\begin{array}{ll} 0 & , x < a \vee x > b \\ 1 & , x \in (a,b) \\ \frac{1}{2} & , x \in \{a,b\} \end{array}\right\} =: \Psi(x)$$

Da fehlt gaaaanz viel

4.28 Satz

Ein Vektor $X = (X_1, \dots, X_D)$ in reellwertigen Zufallsvariablen ist genau dann unabhängig, falls für alle $t = (t_1, \dots, t_D) \in \mathbb{R}^D$ gilt:

$$\varphi_X(t) = \prod_{d=1}^{D} \varphi_{X_d}(t_d) \quad (*)$$

Beweis:
Es gelte $(*)$. Sei P_X die Verteilung von X und P_{X_D} die Verteilung von X_d $(d = 1, \dots, D)$
Dann gilt für alle $t \in \mathbb{R}^D$:

$$\int_{\mathbb{R}^D} e^{i<t,x>} \varphi_X(\mathrm{d}x) = P_X(t) \underset{(*)}{=} \prod_{d=1}^{D} \varphi_{X_d}(t_d) = \prod_{d=1}^{D} \int_{\mathbb{R}} e^{it_d x_d} P_{X_d}(\mathrm{d}x_d) \underset{\text{Fubini}}{=} \int_{\mathbb{R}^D} e^{i<t,x>} (P_{X_1} \otimes \cdots \otimes P_{X_D})(\mathrm{d}x)$$

Da fehlt wieder gaaaanz viel
Das Folgende ist Teil vom Beweis von Satz 4.26 Teil (ii).
Somit:

$$P_n\left(\mathbb{R} \backslash \left(-\frac{1}{a}, \frac{1}{a}\right)\right) \le \frac{7}{a} \int_0^a (1 - \mathrm{Re}(\varphi_n(t)))\lambda(\mathrm{d}t) \underset{n\to\infty}{\to} \frac{7}{a} \int_0^a (1 - \mathrm{Re}(\varphi(t)))\lambda(\mathrm{d}t)$$

Da φ stetig in 0 ist, folgt:

$$\lim_{a\downarrow 0} \frac{7}{a} \int_0^a (1 - \mathrm{Re}(\varphi(t)))\lambda(\mathrm{d}t) = 7(1 - \mathrm{Re}(\varphi(0))) = 0$$

da $\varphi_n(0) = \int 1 \,\mathrm{d}P_n = 1$ für alle $n \in \mathbb{N}$, also $\varphi(0) = 1$
Für ε existiert ein $a > 0$ mit:

$$\sup_{n\in\mathbb{N}} P_n\left(\mathbb{R} \backslash \left[-\frac{1}{a}, \frac{1}{a}\right]\right) \le \sup_{n\in\mathbb{N}} P_n\left(\mathbb{R} \backslash \left(-\frac{1}{a}, \frac{1}{a}\right)\right) \le \varepsilon$$

Also ist $\{P_n | n \in \mathbb{N}\}$ straff.
Aufgrund des Satzes von Prokorov existiert also eine gegen ein W-Maß schwach konvergente Teilfolge von $(P_n)_{n\in\mathbb{N}}$. Sei (P_{n_k}) eine gegen ein W-Maß P schwach konvergente Teilfolge.
Aufgrund von Satz 4.26(i) gilt für alle $t \in \mathbb{R}$

$$\lim_{k\to\infty} \varphi_{n_k} = \varphi_P(t)$$

wobei φ_P die charakteristische Funktion von P ist. Insbesondere folgt, dass $\varphi(t) = \varphi_P(t)$ für alle $t \in \mathbb{R}$, so dass die charakteristische Funktion aller schwachen Grenzwerte P von Teilfolgen von $(P_n)_{n\in\mathbb{N}}$ gleich sind. Der Eindeutigkeitssatz **??** impliziert nun, dass alle schwach konvergenten Teilfolgen von $(P_n)_{n\in\mathbb{N}}$ gegen dasselbe W-Maß P konvergieren, dessen charakteristische Funktion φ ist.
Nach Satz **??** konvergiert dann die Folge $(P_n)_{n\in\mathbb{N}}$ selbst schwach gegen P. ■

4.6. Der zentrale Grenzwertsatz von Lindeberg-Feller

4.29 Definition

Eine Folge von unabhängigen Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ in $L^2(P)$ mit $\mathrm{Var}(X_n) > 0$ genügt dem zentralen Grenzwertsatz, falls die Folge $(P_{S_n})_{n\in\mathbb{N}}$ der Verteilungen von

$$S_n := \frac{\sum_{k=1}^n X_k - E[X_k]}{\sqrt{\mathrm{Var}(X_1 + \cdots + X_n)}}$$

schwach gegen eine $\mathcal{N}(0,1)$-Verteilung konvergiert. Notation:

$$\begin{aligned} \sigma_n^2 &:= \mathrm{Var}(X_n) \\ s_n^2 := \mathrm{Var}(X_1 + \cdots + X_n) = \sum_{k=1}^n \mathrm{Var}(X_k) &= \sum_{k=1}^n \sigma_k^2 \\ \mu_n &:= E(X_n) \end{aligned}$$

Feller-Bedingung:

$$(F) \ \lim_{n\to\infty} \max_{j=1,\dots,n} \frac{\sigma_j}{s_n} = 0$$

Lindeberg-Bedingung:

$$(L) \ \underbrace{\frac{1}{s_n^2} \sum_{j=1}^n \int_{\{|x-\mu| \geq \varepsilon s_n\}} (x-\mu_j)^2 P_{X_j}(\mathrm{d}x)}_{L_n(\varepsilon)} = 0$$

4.30 Satz(Lindeberg/Feller)

Sei $(X_n)_{n\in\mathbb{N}}$ eine Folge in $L^2(P)$ mit $\mathrm{Var}(X_n) > 0$. 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)

4.31 Beispiel

(i) Sind die $(X_n)_{n\in\mathbb{N}}$ unter den Voraussetzungen von Satz 4.30 identisch verteilt, so folgt:

$$L_n(\varepsilon) = \frac{1}{\sigma^2} \int_{\{|x-\mu| \geq \varepsilon\sigma\sqrt{n}\}} (x-\mu)^2 P_{X_1}(\mathrm{d}x) \to 0 \ (n \to \infty)$$

mit $\mu = \mu_1$ und $\sigma^2 = \sigma_1^2$ Der Satz von Lindeberg-Feller beinhaltet also den zentralen Grenzwertsatz von Lindeberg-Levy als Spezialfall.

(ii) Existiert eine Konstante $K < \infty$ mit $|X_n| \leq K$ P-fast sicher für alle $n \in \mathbb{N}$ und falls $s_n \to \infty$ 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:

$$(\Lambda) \ \exists\, \delta > 0 : \ \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{j=1}^n E\Big[|X_j - \mu_j|^{2+\delta}\Big] = 0$$

Denn:

$$\frac{1}{s_n^{2+\delta}} \sum_{j=1}^n E\Big[|X_j - \mu_j|^{2+\delta}\Big] \leq \frac{(2K)^\delta}{s_n^{2+\delta}} \underbrace{\sum_{j=1}^n E[X_j - \mu_j]^2}_{s_n^2} = \frac{(2K)^\delta}{s_n^\delta} \to 0 \ (n \to \infty)$$

Die Lyapunoff-Bedingung impliziert aber die Lindeberg-Bedingung. Denn:

$$|x-\mu_j| \geq \varepsilon s_n \Rightarrow |x-\mu_j|^{2+\delta} \geq |x-\mu_j|^2 (\varepsilon s_n)^\delta$$

also:

$$L_n(\varepsilon) \leq \frac{1}{s_n^{2+\delta}\varepsilon^\delta} \sum_{j=1}^n \int_{\mathbb{R}} |x-\mu_j|^{2+\delta} P_{X_j}(\mathrm{d}\,\acute{}\,) = \frac{1}{\varepsilon^\delta} \frac{1}{s_n^{2+\delta}} E\Big[|X_j-\mu_j|^{2+\delta}\Big] \underset{(\Lambda)}{\to} 0 \; (n \to \infty)$$

Beweis:
nur (i)⇒(ii), für (ii)⇒(i) siehe z.B. Bauer. Œ Sei $E[X_n]=0$ für alle $n \in \mathbb{N}$. Sei $Y_{n,j} := \frac{X_j}{s_n}$. Dann: $\sum_{j=1}^n Y_{n,j} = \frac{\sum_{j=1}^n X_j}{\sqrt{\operatorname{Var}(X_1+\cdots+X_n)}}$. 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:

$$\varphi_{\sum_{j=1}^n Y_{n,j}}(t) \underset{\text{unabh.}}{=} \prod_{j=1}^n \varphi_{Y_{n,j}}(t) = \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{s_n}\right)$$

Sei Z $\mathcal{N}(0,1)$-verteilt und Z_j $\mathcal{N}\big(0,\sigma_j^2\big)$-verteilt. Dann:

$$\varphi_Z(t) = e^{-\frac{t^2}{2}} = e^{-\frac{1}{s_n^2}\sum_{j=1}^n \sigma_j^2 \frac{t^2}{2}} = \prod_{j=1}^n e^{-\sigma_j^2 \frac{\left(\frac{t}{s_n}\right)^2}{2}} = \prod_{j=1}^n \varphi_{j=1}^n \varphi_{Z_j}\left(\frac{t}{s_n}\right)$$

Seien $a_1,\dots,a_n$ und $b_1,\dots,b_n$ komplexe Zahlen vom Betrag kleiner gleich 1. Dann:

$$\left|\prod_{\nu=1}^n a_j - \prod_{\nu=1}^n b_j\right| = |(a_1-b_1)a_2 \cdot \dots \cdot a_n + b_1(a_2-b_2)a_3 \cdot \dots \cdot a_n + \dots + b_1 \cdot \dots \cdot b_{n-1}(a_n-b_n)| \leq \sum_{\nu=1}^n |a_\nu - b_\nu|$$

Also:

$$\left|\varphi_Z(t) - \varphi_{\sum_{j=1}^n Y_{n,j}}(t)\right| \leq \left|\prod_{j=1}^n \varphi_{Z_j}\left(\frac{t}{s_n}\right) - \prod_{j=1}^n \varphi_{X_j}\left(\frac{t}{s_n}\right)\right| \leq \sum_{j=1}^n \left|\phi_{Z_j}\left(\frac{t}{s_n}\right) - \phi_{X_j}\left(\frac{t}{s_n}\right)\right| := (*)$$

Da $E[X_j] = E[Z_j] = 0$ und $E\big[X_j^2\big] = E\big[Z_j^2\big] = \sigma_j^2$, liefert Folgerung **??**:

$$(*) \leq \frac{t^2}{2}\frac{1}{s_n^2}\sum_{j=1}^n \left(\sup_{0\leq\vartheta\leq 1}\left|\varphi''_{X_j}\left(\vartheta\frac{t}{s_n}\right) - \varphi''_{X_j}(0)\right| + \sup_{0\leq\vartheta\leq 1}\left|\varphi''_{Z_j}\left(\vartheta\frac{t}{s_n}\right) - \varphi''_{Z_j}(0)\right|\right) := (**)$$

Nach Satz **??** gilt:

$$\left|\varphi''_{X_j}\left(\vartheta\frac{t}{s_n}\right) - \varphi''_{X_j}(0)\right| \leq E\Big[X_j^2|e^{-i\vartheta\frac{t}{s_n}X_j} - 1\Big] := (***)$$

Für festes $t \in \mathbb{R}$ sei $\delta > 0$. Dann existiert ein $\varepsilon > 0$ mit $\left|e^{i\vartheta t x} - 1\right| \leq \delta$ für alle $|x| \leq \varepsilon$ und $\vartheta \in [0,1]$. Also:

$$(***) \leq \delta E\big[X_j^2\big] + \int_{\{|x|\geq s_n\varepsilon\}} x^2\left|e^{i\vartheta\frac{t}{s_n}x} - 1\right| P_{X_j}(\mathrm{d}x) \leq \delta\sigma_j^2 + 2\int_{\{|x|\geq s_n\varepsilon\}} x^2 P_{X_j}(\mathrm{d}x)$$

Die gleiche Abschätzung gilt für Z_j statt X_j. Also folgt mit (**) und (***):

$$\left|\varphi_Z(t) - \varphi_{\sum_{j=1}^n Y_{n,j}}\right| \leq t^2\left(\delta + \underbrace{\frac{1}{s_n^2}\sum_{j=1}^n \int_{\{|x|\geq s_n\varepsilon\}} x^2 P_{X_j}(\mathrm{d}x)}_{I_n} + \underbrace{\frac{1}{s_n^2}\sum_{j=1}^n \int_{\{|x|\geq s_n\varepsilon\}} x^2 P_{Z_j}(\mathrm{d}x)}_{II_n}\right)$$