

# Maß- und Wahrscheinlichkeitstheorie Skript Felsenstein W2022

Ida Hönigmann

January 26, 2023

## 1 Produkträume und Maße

Auf dem kartesischen Produkt von Grundmengen

$$\Omega = \bigtimes_{i=1}^n \Omega_i$$

wird eine Produkt-(Sigma)Algebra konstruiert, wobei  $(\Omega_i, \mathcal{A}_i)$  Messräume sind. Es soll  $(\Omega_i, \mathcal{A}_i)$  in  $(\Omega, \mathcal{A})$  eingebettet werden.

**Def** (Projektion).

$$\pi_i : \Omega \mapsto \Omega_i \text{ mit } \pi_i(\omega_1, \dots, \omega_n) = \omega_i$$

Die Produktalgebra wird von den Projektionen erzeugt.

**Definition 1** (Produktalgebra).

$$\mathcal{A} := \bigotimes_{i=1}^n \mathcal{A}_i := \sigma(\pi_1, \dots, \pi_n)$$

$$\text{d.h. } \pi_1^{-1}(A_1) \cap \pi_2^{-1}(A_2) \cap \dots \cap \pi_n^{-1}(A_n) \text{ für } A_i \in \mathcal{A}_i$$

erzeugen diese Algebra (bzw.  $A_1 \times A_2 \times \dots \times A_n$ ).

Wenn  $\mathcal{A}_i = \sigma(\mathcal{E}_i)$  erzeugt wird von  $\mathcal{E}_i$ , wird  $\bigotimes \mathcal{A}_i$  von  $\bigtimes \mathcal{E}_i$  erzeugt.

**Satz 1** (1.PR).  $\mathcal{A}_i = \sigma(\mathcal{E}_i)$  mit  $\mathcal{E}_i \subset 2^{\Omega_i}$ .  $\Omega_i$  sei aus  $\mathcal{E}_i$  monoton erreichbar  $(E_{i,k} \nearrow \Omega_i)^a$ .

$$\mathcal{E} = \left\{ \bigtimes_{i=1}^n E_i, E_i \in \mathcal{E}_i \right\}$$

Dann gilt

$$\bigotimes_{i=1}^n \mathcal{A}_i = \sigma(\mathcal{E}).$$

---

<sup>a</sup>Man kann fordern, dass  $\Omega_i \in \mathcal{E}_i$ , dann ist es erfüllt.

*Beweis Satz 1.PR.*  $\sigma(\mathcal{E})$  ist die kleinste Sigma-Algebra, sodass die Projektionen messbar sind:  $\tilde{\mathcal{A}}$  sei Sigma-Algebra auf  $\Omega = \bigtimes_i \Omega_i$ .

$\pi_i$  ist  $\tilde{\mathcal{A}} - \mathcal{A}_i$  messbar  $\iff \sigma(\mathcal{E}) \subseteq \tilde{\mathcal{A}}$ .

$\implies$  : Alle  $\pi_i$  seien  $\tilde{\mathcal{A}} - \mathcal{A}_i$  messbar, d.h.  $\pi_i^{-1}(\mathcal{E}_i) \subseteq \tilde{\mathcal{A}}$ , wobei für  $E_i \in \mathcal{E}_i$

$$\pi_i^{-1}(E_i) = \Omega_1 \times \Omega_2 \times \dots \times E_i \times \Omega_{i+1} \times \dots \times \Omega_n$$

Da

$$\underbrace{\bigtimes_{i=1}^n E_i}_{\in \mathcal{E}} = \bigcap_{i=1}^n \pi_i^{-1}(E_i) \subseteq \tilde{\mathcal{A}} \implies \sigma(\mathcal{E}) \subseteq \tilde{\mathcal{A}}.$$

$\Leftarrow$  : Es gelte  $\sigma(\mathcal{E}) \subseteq \tilde{\mathcal{A}}$ . Jedes  $\Omega_i$  ist aus  $\mathcal{E}_i$  monoton erreichbar mit  $E_{i,k} \nearrow \Omega_i, k \rightarrow \infty$ .

$$F_k = E_{1,k} \times E_{2,k} \times \dots \times E_i \times \dots \times E_{n,k} \nearrow \pi_i^{-1}(E_i)$$

$$\lim F_k = \bigcup_{k=1}^{\infty} F_k = \pi_i^{-1}(E_i) \in \sigma(\mathcal{E}) \subset \tilde{\mathcal{A}}$$

Urbild vom Erzeuger in  $\tilde{\mathcal{A}}$ ,  $\pi_i^{-1}(\mathcal{E}_i) \subseteq \tilde{\mathcal{A}} \forall i$ , alle  $\pi_i$  sind  $\tilde{\mathcal{A}} - \mathcal{A}_i$  messbar.

Alle Projektionen  $\sigma(\mathcal{E})$  messbar, also  $\sigma(\pi_1, \dots, \pi_n) \subset \sigma(\mathcal{E})$ . Nach obigem setze  $\tilde{\mathcal{A}} = \sigma(\pi_1, \dots, \pi_n) \implies \sigma(\mathcal{E}) \subseteq \sigma(\pi_1, \dots, \pi_n)$  also  $\sigma(\mathcal{E}) = \sigma(\pi_1, \dots, \pi_n)$ .  $\square$

**Beberkung.**  $\{A_1 \times \dots \times A_n | A_i \in \mathcal{A}_i\}$  ist keine Sigma-Algebra, da nicht Vereinigungs-stabil.

Die komponentenweise Behandlung im Produktraum ist anwendbar auf n-dim. Funktionen.

**Folgerung.**

$$f_i : \Omega_0 \mapsto \Omega_i, f = (f_1, \dots, f_n) =: \otimes_i f_i$$

$$f : \Omega_0 \mapsto \bigtimes_i \Omega_i = \Omega$$

Dann gilt:

$$f \text{ ist } (\Omega_0, \mathcal{A}_0) \mapsto \underbrace{\bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i)}_{\text{Produktraum}} \text{ messbar} \iff f_i \mathcal{A}_0 - \mathcal{A}_i \text{ messbar}$$

*Beweis Folgerung.*  $\implies$  : Da  $f_i = \pi_i \circ f$  und  $\pi_i \otimes_j \mathcal{A}_j - \mathcal{A}_i$  messbar ist  $f_i$  als Verkettung messbar.

$\Leftarrow$  :  $A \in \otimes \mathcal{A}_i$  Menge aus der Produktalgebra mit  $A = \times_{i=1}^n A_i \leftarrow$  erzeugen  $\otimes_{i=1}^n \mathcal{A}_i$ .

$$f^{-1}(A) = \bigcap_{i=1}^n f_i^{-1}(A_i) \in \mathcal{A}_0$$

Diese Rechtecke erzeugen  $\sigma(f) = \sigma(f_1, \dots, f_n) \subseteq \mathcal{A}_0$  also ist  $f \mathcal{A}_0 - \otimes \mathcal{A}_i$  messbar.  $\square$

Anwendung auf Borel-Algebra:  $\mathcal{B}^k = \otimes_{i=1}^k \mathcal{B}$  erzeugt von den Rechtecken  $\times_i(a_i, b_i]$ . Die Lebesgue-Mengen werden nicht von den Produkten erzeugt:  $\otimes_{i=1}^k \mathcal{L} \subsetneq \mathcal{L}_k$ .

Nicht alle Nullmengen von  $\mathcal{B}^k$  sind durch die Produkte mit allen Nullmengen erzeugbar.

**Def (Schnitt).** *Besondere Mengen (für Integralberechnungen) sind die Schnitte:*

$$A \in \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$$

$$A_{x_1} := \{x_2 \in \Omega_2 | (x_1, x_2) \in A\}$$

$$A_{x_2} := \{x_1 \in \Omega_1 | (x_1, x_2) \in A\}$$

Die Schnitte sind messbar.

**Satz 2** (2.PR).  $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$  messbar bez. Produktalgebra, dann ist  $A_{x_1} \in \mathcal{A}_2$  und  $A_{x_2} \in \mathcal{A}_1$  für alle  $x_1$  bzw.  $x_2$ .

*Proof.*  $x_1$  sei fest. Betrachte  $\mathcal{M} = \{A \subseteq \Omega_1 \times \Omega_2 | A_{x_1} \in \mathcal{A}_2\}$ .  $\mathcal{M}$  ist Sigma-Algebra:  $\Omega \in \mathcal{M}$ ,  $(A_{x_1})^c = (A^c)_{x_1}$  und  $\cup A_{x_i,1} = (\cup A_i)_{x_1}$ .

Für die Erzeuger Mengen  $A_1 \times A_2 \in \mathcal{E}$  mit  $A_i \in \mathcal{A}_i$

$$(A_1 \times A_2)_{x_1} = \begin{cases} A_2, & x_1 \in A_1 \\ \emptyset, & x_1 \notin A_1 \end{cases}$$

also  $(A_1 \times A_2)_{x_1} \in \mathcal{A}_2$ ,  $\mathcal{E} \subseteq \mathcal{M}$  und  $\sigma(\mathcal{E}) = \mathcal{A}_1 \otimes \mathcal{A}_2 \subseteq \mathcal{M}$ , alle Mengen, alle  $x_1$  erfüllen die Messbarkeit-Bedingungen.  $\square$

Die Abbildungen  $x_1 \mapsto \mu_2(A_{x_1})$  sind messbare Abbildungen.

**Satz 3** (3.PR).  $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ , auf  $\mathcal{A}_2$  sei  $\mu_2$  ein sigma-endliches Maß auf  $(\Omega_2, \mathcal{A}_2, \mu_2)$ . Dann ist  $x_1 \mapsto \mu_2(A_{x_1})$  eine  $\mathcal{A}_1$  messbare Abbildung (entsprechendes gilt auch für  $x_2 \mapsto \mu_1(A_{x_2})$ ).

*Proof.* Für ein  $A \in \mathcal{A}$  sei  $f_A(x_1) = \mu_2(Ax_1)$

1.  $\mu_2$  sei endlich. Betrachte  $\mathcal{D} = \{E \in \mathcal{A} \mid f_E \text{ ist } \mathcal{A}_1 \text{ Borel-messbar}(\mathbb{R}^+, \mathcal{B}^+)\}$

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

- $\Omega \in \mathcal{D}$ , da  $\Omega_{x_1} = \Omega_2 \in \mathcal{A}_2 \forall x_1 : f_\Omega \equiv \mu_2(\Omega_2)$  konstant
- $A \subseteq B$  und  $f_A, f_B$  messbar  
 $A_{x_1} \subseteq B_{x_1}$  und  $\mu_2(B_{x_1} \setminus A_{x_1}) = \mu_2(B_{x_1}) - \mu_2(A_{x_1}) = f_B(x_1) - f_A(x_1)$  ist messbar als Differenz messbarer Funktionen.
- $A_i \in \mathcal{A}$  und disjunkt,  $B = \bigcup_i A_i$   
 Wenn  $A_i \in \mathcal{D}$ ,  $B_{x_1} = \bigcup A_{i,x_1}$

$$\mu_2(B_{x_1}) = f_B(x_1) = \sum_i f_{A_i}(x_1) = \mu_2\left(\bigcup A_{x_1}\right)$$

Eine abzählbare Summe messbarer Funktionen ist messbar (Darstellung jeder messbarer Funkt  $f = \sum_j c_j 1_{c_j}$ ).

Jede Menge aus  $\mathcal{E} = \{A_1 \times A_2 \mid A_i \in \mathcal{A}_i\}$  (dem Erzeuger von  $\mathcal{A}$ ) ist auch in  $\mathcal{D}$ , weil  $f_{A_1 \times A_2}(x_1) = \mu_2((A_1 \times A_2)_{x_1}) = \mu_2(A_2) 1_{A_1}(x_1)$ .  $c_1$  ist messbar.

$\mathcal{E}$  ist ein durchschnitt-stabiler Erzeuger und  $\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{A} = \sigma(\mathcal{E})$  also  $\mathcal{A} = \mathcal{D}$ , alle solchen Funktionen sind messbar bez.  $\mathcal{A}_\infty$ .

2. Ist  $\mu_2$  sigma-endlich, es gibt eine Folge  $A_{2,i} \nearrow \Omega_2$  mit endlichem Maß, daher auch eine disjunkte Folge  $(D_n)$ , die eine Zerlegung von  $\Omega_2$  sind und  $\mu_2(B) = \mu_2(\bigcup_n (D_n \cap B)) = \sum_n \mu_2(B \cap D_n)$ .

$$f_B(x_1) = \mu_2(B_{x_1}) = \sum_n \underbrace{\mu_2(B_{x_1} \cap D_n)}_{< \infty} = \sum_n f_{(B_{x_1} \cap D_n)}(x_1)$$

also Summe messbarer Funktionen.

□

Mit diesen Funktionen wird das Produktmaß erklärt.

**Def** (Produktmaß).

$$\mu\left(\bigtimes_{i=1}^k A_i\right) = \prod_{i=1}^k \mu_i(A_i)$$

Wenn  $\Omega = \Omega_1^k$  mit  $\mu_i = \mu$  gilt

$$\mu\left(\bigtimes_{i=1}^k A_i\right) = \prod_{i=1}^k \mu(A_i).$$

Dieses Maß ist eindeutig definiert, wenn  $\mu_1, \mu_2$  sigma-endlich sind, dann sind die Funktionen  $f_B(x_1)$  bzw.  $f_B(x_2)$  die "Dichten" bezüglich den Randmaßen  $\mu_1$  bzw.  $\mu_2$ .  $\mu_1, \mu_2$  werden auch als marginale Maße bezeichnet.

**Satz 4** (4.PR). 1.  $\mu_2$  sei sigma-endlich. Dann definiert

$$\mu(A) = \int_{\Omega_1} \mu_2(A_{x_1}) d\mu_1(x_1)$$

das Produktmaß auf  $(\Omega, \mathcal{A})$ .

2. Sind beide Maße  $\mu_1, \mu_2$  sigma-endlich, dann ist  $\mu$  eindeutig und

$$\mu(A) = \int_{\Omega_1} \mu_2(A_{x_1}) d\mu_1(x_1) = \int_{\Omega_2} \mu_1(A_{x_2}) d\mu_2(x_2)$$

*Proof.*  $f_A(x_1) = \mu_2(A_{x_1})$  ist eine  $\mathcal{A}_1$ -messbare Funktion  $\forall A$  und durch die Additivität  $f_{\cup A_i} = \sum f_{A_i}$  ist  $\mu$  ein Maß auf  $(\Omega, \mathcal{A})$ .

Für  $A = A_1 \times A_2$  gilt

$$\mu(A_1 \times A_2) = \int_{\Omega_1} \underbrace{\mu_2((A_1 \times A_2)_{x_1})}_{\mu_2(A_2)1_{A_1}(x_1)} d\mu_1 = \mu_2(A_2) \underbrace{\int_{A_1} d\mu_1}_{\mu_1(A_1)}$$

ist Produktmaß.

$\mu$  ist auf dem (durchschnitts-stabilen) Erzeuger definiert, da die Maße sigma-endlich sind, ist  $\mu$  eindeutig (Eindeutigkeitssatz). Vice versa gelten alle Gleichungen analog bez.  $\mu_2$ .  $\square$

**Beispiel.** Insbesondere bei endlichem Maß anwendbar.  $X, Y$  stochastische Größen, dann wird eindeutig eine zweidim. Verteilung auf  $\mathbb{R}^2$  durch  $P[(X, Y) \in A \times B] := P[X \in A]P[Y \in B]$  Produktverteilung.

Verallgemeinerung der Maße von Schnitten ist die Schnittfunktion.

**Definition 2** (Schnittfunktion).

$$\begin{aligned} f : \Omega_1 \times \Omega_2 &\mapsto \Omega' \text{ Dann heißt} \\ x_2 &\mapsto f_{x_1}(x_2) = f(x_1, x_2) \text{ } x_1\text{-Schnitt} \\ x_1 &\mapsto f_{x_2}(x_1) = f(x_1, x_2) \text{ } x_2\text{-Schnitt} \end{aligned}$$

von  $f$ . (messbare Funktion  $\Omega_1 \times \Omega_2 \rightarrow (\Omega', \mathcal{A}')$ ).

**Satz 5** (5.PR).  $f$  sei messbar  $(\Omega, \mathcal{A}) \mapsto (\Omega', \mathcal{A}')$ . Die Schnittfunktionen sind messbar:

$$\begin{aligned} f_{x_1} &\text{ ist messbar } (\Omega_2, \mathcal{A}_2) \mapsto (\Omega', \mathcal{A}') \\ f_{x_2} &\text{ ist messbar } (\Omega_1, \mathcal{A}_1) \mapsto (\Omega', \mathcal{A}') \end{aligned}$$

*Proof.*  $A' \in \mathcal{A}'$

$$f_{x_1}^{-1}(A') = \{x_2 \in \Omega_2 | f(x_1, x_2) \in A'\} = \{x_2 \in \Omega_2 | (x_1, x_2) \in f^{-1}(A')\} = \underbrace{(f^{-1}(A'))_{x_1}}_{\text{mb } \in \mathcal{A}}$$

jede Schnittmenge ist messbar.

Analog für  $f_{x_2}^{-1}(A') \in \mathcal{A}_1$ .  $\square$

Mit den Funktionenschnitten lässt sich auch ein mehrdim. Integral "zerteilen".

**Satz 6** (6.PR Satz von Fubini (-Tonelli)). Produktraum  $(\Omega, \mathcal{A}, \mu) = \otimes_i (\Omega_i, \mathcal{A}_i, \mu_i)$ ,  $f$  messbar  $\Omega \mapsto \mathbb{R}$ ,  $\mu_i$  sigma-endlich. Das zweidimensionale Integral von  $f$  ist aufspaltbar

$$\int f d\mu = \int_{\Omega_1} \left( \int_{\Omega_2} f_{x_1}(x_2) d\mu_2 \right) d\mu_1 = \int_{\Omega_2} \left( \int_{\Omega_1} f_{x_2}(x_1) d\mu_1 \right) d\mu_2$$

wenn eine der folgenden Bedingungen gilt:

1.  $f \geq 0$ : Dann ist  $x_2 \mapsto \phi_2(x_2) = \int_{\Omega_1} f_{x_2} d\mu_1$  messbar  $\mathcal{A}_2$  und  $x_1 \mapsto \phi_1(x_1) = \int_{\Omega_2} f_{x_1} d\mu_2$  messbar  $\mathcal{A}_1$
2.  $f$  ist integrierbar,  $\int f d\mu < \infty$ : Dann sind  $f_{x_1}$   $\mu_2$ -integrierbar  $[\mu_1]$  f.ü. und  $f_{x_2}$   $\mu_1$ -integrierbar  $[\mu_2]$  f.ü.
3.  $\int_{\Omega_1} \int_{\Omega_2} |f_{x_1}| d\mu_2 d\mu_1 < \infty$  oder  $\int_{\Omega_2} \int_{\Omega_1} |f_{x_2}| d\mu_1 d\mu_2 < \infty$ : Daraus folgt  $f$  integrierbar.

*Proof.* 1.  $f \geq 0$ : 4 Schritte des Integralaufbaus

$$f = 1_A, A \in \mathcal{A}$$

$$x_2 \mapsto \phi_2(x_2) = \int_{\Omega_1} \underbrace{(1_A)_{x_2}}_{1_{Ax_2}} d\mu_1 = \mu_1(A_{x_2})$$

$\phi_2$  ist messbar (laut Satz PR5) und nach Satz PR7 gilt

$$\int f d\mu = \mu(A) = \int_{\Omega_2} \mu_1(A_{x_2}) d\mu_2(x_2) = \int_{\Omega_2} \phi_2 d\mu_2$$

Für einfache Funktionen  $f = \sum_{i=1}^n \alpha_i 1_{A_i}$  ergibt sich das aus der Linearität:

$$f_{x_2} = \sum \alpha_i 1_{(A_i)_{x_2}}(x_1)$$

Schritt 3:  $f_2 \nearrow f, f_n \dots$  einfache Funktionen, dann gilt  $(f_n)_{x_2} \nearrow f_{x_2}$  (wieder aus der Darstellung  $f = \sum_{i=1}^{\infty} \alpha_i 1_{A_i} +$  monotone Konvergenz ablesbar)

2.  $f$  integrierbar, betrachte Positiv- und Negativteil

$$\begin{aligned} \max(0, f)_{x_2} &= \max(0, f_{x_2}) = (f^+)_{x_2} \\ \text{also } (f^+)_{x_2} &= (f_{x_2})^+ \text{ und } (f^-)_{x_2} = (f^-)_{x_2} \text{ mit } |f|_{x_2} = |f_{x_2}| \end{aligned}$$

Wegen  $\int |f| d\mu < \infty$  gilt wegen oben für  $|f|$ :

$$\int |f| d\mu = \int_{\Omega_1} \int_{\Omega_2} |f|_{x_1} d\mu_2 d\mu_1 = \int_{\Omega_1} \underbrace{\int_{\Omega_2} |f_{x_1}| d\mu_2}_{\phi_1} d\mu_1 < \infty$$

Daher muss  $\phi_1$  integrierbar sein (wie auch  $\phi_2$ ). Integrierbarkeit erfordert auch  $(f^+), (f^-)$  integrierbar sind, aufgespalten  $f = f^+ - f^-$  und  $f_{x_1} = f_{x_1}^+ - f_{x_1}^-$ . Nach 1. für  $f^+, f^-$  getrennt ergibt

$$\int f d\mu = \int f^+ - \int f^- = \int_{\Omega_1} \int_{\Omega_2} f_{x_1}^+ d\mu_2 d\mu_1 - \int \int f^- d\mu_2 d\mu_1$$

3. impliziert  $\int_{\Omega_2} f_{x_1}^+ d\mu_2 < \infty, \int_{\Omega_2} f_{x_1}^- d\mu_2 < \infty$  und somit  $f$  integrierbar und 2. □

Durch Iteration gilt die Fubini-Schnitt Konstruktion auch für mehrdimensionale  $k \in \mathbb{N}$  Integrale:  
Wenn  $f : \Omega \mapsto \mathbb{R}$  messbar mit  $\Omega = \Omega_1 \times \dots \times \Omega_n$

$$\int f d\mu = \int_{\Omega_1} \left( \int_{\Omega_2 \times \dots \times \Omega_n} f_{x_1} d\mu_2 \otimes \dots \otimes \mu_k \right) d\mu_1(x_1)$$

Viele Folgerungen: Doppelreihen-Satz  $\sum_i \sum_j a_{ij} = \sum_j \sum_i a_{ij}$  Kriterien für Konvergenz

**Folgerung** (Maße mit Dichten bezüglich dem Produktmaß).  $\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \mu = \mu_1 \otimes \mu_2$   
 $\nu$  sei absolut stetig bzgl.  $\mu$  ( $\nu \ll \mu$ ), es existiert ein  $f = \frac{d\nu}{d\mu} \geq 0$  mit  $\nu(A) = \int_A f d\mu$  ( $f$  integrierbar)  
 $\mu$  sei sigma-endlich,  $\nu$  erzeugt ein  $\nu_1 \ll \mu_1$  auf  $(\Omega_1, \mathcal{A}_1)$  und ein  $\nu_2 \ll \mu_2$  auf  $(\Omega_2, \mathcal{A}_2)$  mit den Dichten  $\phi_1, \phi_2$ .

$$A \in \mathcal{A}_1 : \nu_1(A_1) = \int_{A_1} \phi_1(x_1) d\mu_1 = \int_{A_1} \int_{\Omega_2} f_{x_1}(x_2) d\mu_2 d\mu_1$$

**Beispiel.** 2 dim SG mit Gleichverteilung auf dem Einheitskreis  $P[(X, Y) \in K] = 1$ . Dichte  $f$  bezgl.  $\lambda^2 = \lambda_1 \otimes \lambda_1$ .  $f(x, y) = \frac{1}{\pi} 1_K$ .

$$\phi_1(x) = \int f_X(y) d\lambda = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} d\lambda(y) = \frac{2\sqrt{1-x^2}}{\pi}$$

(Symmetrie:  $\phi_2(y) = \frac{2\sqrt{1-y^2}}{\pi}$ )

Randverteilung  $\nu_1$   $P[X \in A] = \int_A \frac{2}{\pi} \sqrt{1-y^2} d\lambda(y)$

$\nu$  ist nicht das Produktmaß  $\nu \neq \nu_1 \otimes \nu_2$  außer  $f(x, y) = f_1(x)f_2(y)$ .

**Def** (Ordnatenmenge, Graph). Die Punkte unter einer positiven Funktion  $f \geq 0$  heißt Ordinatensmenge  $O_f = \{(x, y) | 0 \leq y \leq f(x)\}$  und der Graph ist  $\Gamma_f = \{(x, f(x)) | x \in \Omega\}$ .

Für sigma-endliches Maß  $\mu$  und  $f$  messbar, dann ist  $O_f$  eine bezüglich  $\mathcal{A} \otimes \mathcal{B}$  messbare Menge und  $\int f d\mu = (\mu \otimes \lambda)(O_f)$ . Der Graph ist eine Nullmenge,  $(\mu \otimes \lambda)(\Gamma_f) = 0$

## 1.1 $\infty$ -dim. Produkträume

Die Konstruktion der  $\infty$ -dim. Produkt-Messräume ist der endl. dim. Konstruktion entsprechend.  $I$  sei Index-Menge,

$$\otimes_{i \in I} \mathcal{A}_i := \sigma(\pi_i, i \in I) \quad \pi_i \dots \text{Projektion auf } \Omega_i$$

beispielsweise  $\otimes_{i \in I} \mathcal{B}$  auf  $\Omega = \mathbb{R}^I$  (Funktionenraum).

**Def** (Zylinder, Pfeiler).  $Z \subseteq \Omega^I$  heißt Zylinder, wenn  $Z = \pi_J^{-1}(C) = C \times \Omega^{J^c}$  wobei  $J \subseteq I$  eine endliche Teilindexmenge ist und  $C$  die endlich dim. Basis des Zylinders ist; und Pfeiler, wenn  $C$  ein Rechteck  $\times_{i \in J} A_i$  ist.

Auf den Pfeilern lässt sich das  $n$ -dim. Produktmaß erklären.  $P^n(C) = \pi_{j \in J} P(A_j)$   
Im abzählbar unendlichen Fall ist die Vorgangsweise ähnlich, wie im endl. dim. Fall:

**Satz 7** (7.PR).  $X_i$  sei eine Folge von SGn auf  $(\Omega_i, \mathcal{A}_i)$  mit gegebener gemeinsamer Verteilung

$$P_i(A) = P[X_j \in A_j, j \leq i, X_i \in A]$$

für  $A_j \in \mathcal{A}_j$  und  $A \in \mathcal{A}_i$ . D.h. die gemeinsame Verteilung von  $(X_1, \dots, X_n)$  sind auf  $(\times_i \Omega_i, \otimes_i \mathcal{A}_i)$  mit

$$P[(X_1, \dots, X_n) \in A_n, X_i \in \mathbb{R}, i > n] = P[(X_1, \dots, X_n) \in A_n]$$

also für Zylinder  $Z = \pi_{1, \dots, n}^{-1}(C_n)$ ,  $C_n \in \otimes_i \mathcal{A}_i$  gilt  $P[Z] = P_n(C_n)$ .

Wenn diese Wahrscheinlichkeitsverteilung verträglich sind, d.h. die Randverteilungen (bzw. alle Teilmen-gen endl.) eindeutig sind, lässt sich obiges Prinzip verallgemeinern.

**Satz 8** (8.PR Satz von Kolmogoroff).  $(\Omega_i, \mathcal{A}_i), i \in I$  sei eine Familie von Messräumen und  $P_j$  sind endl. dim. Verteilungen auf  $(\mathbb{R}^J, \mathcal{B}_{|J|}), J \subset I, J$  endlich. Diese Verteilungen seien verträglich ( $P_j = P_k \pi_{k,j}^{-1}, J \subset K, K$  ebenfalls endlich).

Dann existiert ein eindeutiges Maß  $P$  auf  $(\mathbb{R}^I, \mathcal{B}_I = \otimes_{i \in I} \mathcal{B})$  mit genau diesen Randverteilungen ( $P(A) = P_J(\pi_J^{-1}(A)), A \in \otimes_{i \in J} \mathcal{A}_i$ ).

Das Produktmaß  $\mu_1 \otimes \mu_2$  auf  $\mathcal{A}_1 \otimes \mathcal{A}_2$  muss nicht vollständig sein, wenn  $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1, 2$  vollständige Maßräume sind.

**Beispiel.**  $(\mathbb{R}^2, \mathcal{L}_2, \lambda_2)$  ist ein vollständiger Maßraum.  $\lambda_2(\mathbb{R} \times \{1\}) = 0$  und für beliebiges  $A \subseteq \mathbb{R} \lambda_2(A \times \{1\}) = 0$ . Wenn  $A \notin \mathcal{L}$ , dann sollte aber trotzdem der Schnitt von  $A \times \{1\}, (A \times \{1\})_1 = A \in \mathcal{L}$  messbar sein, das ist ein Widerspruch. Es folgt also  $\mathcal{L} \otimes \mathcal{L} \neq \mathcal{L}_2$ .

**Satz 9** (9.PR).  $(\Omega_i, \mathcal{A}_i, \mu_i), i = 1, 2$  seien sigma-endliche Maßräume. Wenn die messbaren Funktionen  $f_i : (\Omega_i, \mathcal{A} : i) \rightarrow (\mathbb{R}, \mathcal{B})$  entweder

- $f_i \geq 0$  oder
- $f_i$  ist integrierbar,  $i = 1, 2$

ist, dann gilt

$$\int f_1 \cdot f_2 d\mu_1 \otimes d\mu_2 = \int f_1 d\mu_1 \cdot \int f_2 d\mu_2$$

Im 2. Fall ist  $f_1 \cdot f_2$  auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$  integrierbar.

*Proof.* Eigentlich klar, da  $(f_1 \cdot f_2)_{x_1} = f_1(x_1) \cdot f_2(\cdot)$   $\mathcal{A}_2$  messbar ist und nach Fubini

$$\int f_1 \cdot f_2 d\mu_1 \otimes d\mu_2 = \int_{\Omega_1} \int_{\Omega_2} f_1(x_1) f_2(x_2) d\mu_2(x_2) d\mu_1(x_1) = \int_{\Omega_1} \left( f_1(x_1) \int_{\Omega_2} f_2(x_2) d\mu_2(x_2) \right) d\mu_1(x_1) = \int f_1 d\mu_1 \cdot \int f_2 d\mu_2$$

Genauso gilt  $\int |f_1 f_2| d\mu_1 \otimes d\mu_2 = \int |f_1| d\mu_1 \int |f_2| d\mu_2$  und wenn die rechte Seite endlich ist, gilt  $f_1 \cdot f_2 \in \mathcal{L}_1$  bezüglich dem Produktraum.  $\square$

Die Betrachtung unabhängiger SGn  $X_i, i = 1, \dots, k$  erfolgt bequemer auf dem Produktraum. Auch wenn alle  $\Omega_i = \Omega$ , also alle SGn auf dem selben Raum definiert sind, übersiedelt man für die Erklärung der gemeinsamen Verteilung auf den Produktraum.

Wenn  $X = (X_1, \dots, X_k)$  ein Vektor unabhängiger SGn  $X_i$  ist, gilt

$$PX^{-1} = PX_1^{-1} \otimes \dots \otimes PX_k^{-1}.$$

Wenn  $PX^{-1}$  ein Wahrscheinlichkeitsmaß mit Dichte ist, dann gilt

$$PX^{-1}(A) = \int_A f_X d\lambda_k = \int_A f(x_1, \dots, x_k) d\lambda(x_1) \dots d\lambda(x_k)$$

und die Randverteilung von  $X_i$  ist mit Dichte

$$PX_i^{-1}(A_i) = \int_{A_i \times \mathbb{R}^{k-1}} f_X d\lambda_k = \int_{A_i} \left( \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_X d\lambda \dots d\lambda \right) d\lambda$$

und wieder erhält man die Randdichte

$$f_i(x_i) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1, \dots, x_k) d\lambda(x_1) \dots d\lambda(x_{i-1}) d\lambda(x_{i+1}) \dots d\lambda(x_k)$$

(Das ist nicht neu, aber jetzt wird kein Umweg über das Riemann-Integral benötigt.)

Wenn die  $X_i$  unabhängig sind ist die gemeinsame Dichte

$$f_X(x_1, \dots, x_k) = \prod_{i=1}^k f_i(x_i).$$

Nach dem letzten Satz gilt für integrierbare  $f$  und  $g$  und unabhängige  $X$  und  $Y$

$$\mathbb{E}f(X)g(Y) = \mathbb{E}f(X) \cdot \mathbb{E}g(Y)$$

d.h. auch  $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$  und die SGn  $X$  und  $Y$  sind unkorreliert.

Die Umkehrung gilt natürlich nicht, da Unkorreliertheit nur bedeutet, dass es keinen linearen Zusammenhang gibt.

Wenn  $Y = c$  f.s., dann sind  $X, Y$  unabhängig, wenn  $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$ , dass dann automatisch gilt. Auch wenn  $X \sim A_{p_1}$  (Alternativ-verteilt) und  $Y \sim A_{p_2}$  und

$$\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y = 0(1 - p_1p_2) + 1p_1p_2 = p_1p_2$$

d.h.  $P(X=1)P(Y=1) = p_1p_2$  und  $X$  und  $Y$  sind unabhängig.

Ansonsten ist nur bei Normalverteilung kein Unterschied zwischen Unkorreliertheit und Unabhängigkeit.

**Beispiel.**  $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$  o.B.d.A  $\mu_x = \mu_y = 0$

Die gemeinsame Dichte zerfällt (siehe EI24)

$$f(x, y) = \underbrace{\frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{1}{2}\left(\frac{x}{\sigma_x}\right)^2\right)}_{f_1(x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}\left(\frac{\sigma_x^2(y-m)^2}{(1-\rho^2)\sigma_y^2}\right)^2\right)}_{f_2(x,y)}$$

mit  $m = \rho x \frac{\sigma_y}{\sigma_x}$  d.h.

$$\mathbb{E}XY = \int \int xy f(x, y) d\lambda_2 = \int \underbrace{\int y f_2(x, y) dy}_{m} x f_1(x) dx = \int m x f_1(x) dx = \rho \int x^2 f_1(x) dx = \rho \frac{\sigma_y}{\sigma_x} \sigma_x^2 = \rho \sigma_x \sigma_y$$

$X, Y$  sind unkorreliert  $\iff \rho = 0$ , dann ist  $f(x, y) = f_1(x) \cdot \Phi\left(\frac{y}{\sigma_y}\right)$  ( $\Phi$  Dichte der  $\mathcal{N}(0, 1)$ ) und sind  $X, Y$  unabhängig.

Zwei Normalverteilungen können nur linear abhängen. Ansonsten kann sogar eine vollständige Abhängigkeit (nicht linear) bei unkorrelierten SGn vorliegen.

**Beispiel.**  $X \sim U_{-1,1}, Y = X^2$ . Dann gilt  $\mathbb{E}X = 0$  und  $\mathbb{E}XY = \mathbb{E}X^3 = 0$  und  $X$  und  $Y$  sind unkorreliert.

Nur wenn für alle integrierbaren  $f, g$   $f(X)$  und  $g(Y)$  unkorreliert sind, dann sind  $X$  und  $Y$  unabhängig.

Der Satz von Fubini ist ein wichtiges Werkzeug auch um bekannte Sätze der Integrationstheorie zu verallgemeinern, wie beispielsweise die partielle Integration.

**Satz 10** (10.PR).  $\mu_F$  und  $\mu_G$  seien Lebesgue-Stieltjes Maße mit  $F$  bzw.  $G$  als Verteilungsfunktionen.  $G_-(x) = \lim_{\tilde{x} \nearrow x} G(\tilde{x})$  ist der linksseitige Grenzwert von  $G$ . Dann gilt

$$\int_{(a,b]} F d\mu_G + \int_{(a,b]} G_- d\mu_F = F(b)G(b) - F(a)G(a)$$

*Proof.* in der Übung.

Wenn  $F$  und  $G$  stetig differenzierbar sind, dann gilt  $d\mu_G = G'd\lambda$  und  $d\mu_F = F'd\lambda$  und

$$\int_{(a,b]} F \cdot G' d\lambda + \int_{(a,b]} G \cdot F' d\lambda = F(b)G(b) - F(a)G(a)$$

und in der üblichen Schreibweise für Stammfunktionen  $\int FG' = FG - \int F'G$  □

Mit dem Hauptsatz der Diff- u. Integrationstheorie lässt sich auch die bedingte Verteilung auf stetige Verteilungen erweitern.

Für diskrete SGN  $X, Y$  (beispielsweise auf  $\mathbb{N}$  verteilt) ist

$$P[X = k | Y = l] = \frac{P[X = k, Y = l]}{P[Y = l]}, k, l \in \mathbb{N}$$

die Punktwahrscheinlichkeit  $p_k$  der bedingten Verteilung  $X|Y = l$ .

Besitzt  $X$  eine stetig differenzierbare VF  $F$  mit  $F' = f_X$  als Dichte, gilt

$$\begin{aligned} \lim_{\Delta \rightarrow 0} \frac{F(x + \Delta) - F(x)}{\Delta} &= f_X(x) \text{ oder} \\ \frac{P[X \in [x - \Delta, x]]}{\Delta} &\rightarrow f_X(x) \text{ für } \Delta \rightarrow 0 \text{ oder} \\ P[X \in [x - \Delta, x]] &\sim f_X(x)\Delta \end{aligned}$$

Mit dieser "infinitesimalen Wahrscheinlichkeit" als Dichte erhält man, wenn auch  $Y$  die Dichte  $f_Y$  hat,

$$\begin{aligned} P[X \in [x - \Delta, x] | Y \in [y - \Delta, y]] &= \frac{P[X \in [x - \Delta, x], Y \in [y - \Delta, y]]}{P[Y \in [y - \Delta, y]]} = \\ \frac{\int_{x-\Delta}^x \int_{y-\Delta}^y f_{X,Y}(s, t) ds dt}{\int_{y-\Delta}^y f_Y(t) d\lambda(t)} &\sim \frac{f_{X,Y}(x, y) \Delta^2}{f_Y(y) \Delta} = \frac{f_{X,Y}(x, y) \Delta}{f_Y(y)} \end{aligned}$$

**Definition 3** (bedingte Dichte).  $X : \Omega \rightarrow \mathbb{R}$  und  $Y : \Omega \rightarrow \mathbb{R}$  sind SGN mit Dichte  $f(x, y)$ .  $f_X, f_Y$  sind die Randdichten von  $X$  und  $Y$ . Dann heißt für  $y$  mit  $f_Y(y) > 0$

$$f(x|y) := \frac{f(x, y)}{f_Y(y)}$$

die bedingte Dichte von  $X$  bedingt durch  $Y = y$ .  $f(x|y)$  ist  $PY_-^{-1}$  f.s. definiert.

Da für festes  $y$

$$\int_{\mathbb{R}} f(x|y) d\lambda(x) = \frac{\int_{\mathbb{R}} f(x, y) d\lambda(x)}{\int_{\mathbb{R}} f(x, y) d\lambda(x)} = 1$$

und  $f(x|y) \geq 0$  ist  $f(\cdot|y)$  eine Wahrscheinlichkeitsdichte und tatsächlich eine Verteilung festgelegt.

Wenn  $X|Y = y$   $PY_-^{-1}$ -f.s. einen endlichen Erwartungswert besitzt, dann heißt die Funktion

$$y \mapsto \mathbb{E}[X|Y = y] = \int x f(x|y) d\lambda(x)$$

bedingter Erwartungswert. Dann ist  $h(Y) := \mathbb{E}[X|Y]$  eine  $PY_-^{-1}$ -f.s. messbare Funktion.  $h(\cdot)$  heißt auch Regressionsfunktion.  $h(Y)$  ist als Prognose von  $X$  nach der Beobachtung von  $Y$  zu verstehen.



**Beispiel.**  $(X, Y)$  sei bivariat normalverteilt  $(X, Y) \sim \mathcal{N}(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ . Die Darstellung der bivariaten Normalverteilungsdichte (PR18) führt auf die bedingte Dichte

$$f(x|y) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{1}{2} \left(\frac{x-m}{v}\right)^2\right)$$

mit  $m = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)$  und  $v^2 = \sigma_x^2(1 - \rho^2)$

d.h.  $X|Y \sim \mathcal{N}(m, v^2)$ .

Die Regressionsfunktion ist linear

$$\mathbb{E}[X|Y] = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)$$

Auch diese Eigenschaft charakterisiert die Normalverteilung.

Für die Regressionsfunktion  $H(Y) := \mathbb{E}[X|Y]$  gilt bei Normalverteilung  $\mathbb{E}H(Y) = \mu_x$ . Das ist generell der Fall:

$$\begin{aligned} \mathbb{E}H(Y) &= \int \mathbb{E}[X|Y=y] dPY^{-1} = \int \int x f(x|y) dx f_Y(y) dy \\ &= \int \int x \underbrace{f(x|y) f_Y(y)}_{f(x,y)} dy dx = \int \int x f(x,y) dy dx = \int x f_X(x) dx = \mathbb{E}X. \end{aligned}$$

Die mittlere Prognose entspricht dem unbedingten Erwartungswert.

## 1.2 Faltung von Maßen

**Definition 4** (Faltungsmaß).  $\mu_1, \mu_2$  sind sigma-endliche Maße auf  $(\mathbb{R}, \mathcal{B})$ . Das Faltungsmaß ist durch

$$\mu_1 * \mu_2(A) = \int \mu_1(A - y) \mu_2(dy)$$

definiert.

Das  $\mu_1 * \mu_2$  tatsächlich ein Maß ist, ergibt sich daraus, dass es das von der Summe  $S = X_1 + X_2$  erzeugte Maß ist. D.h.  $\mu_1 * \mu_2 = (\mu_1 \otimes \mu_2)S^{-1}$

*Proof.*  $S^{-1}(A) = \{(x, y) \in \mathbb{R}^2 : x + y \in A\}$ , wobei  $A \in \mathcal{B}$ . Der Schnitt von  $S^{-1}(A)$  ist  $S^{-1}(A)_y = \{x | x \in A - y\} = A - y$

Die Darstellung des Produktmaßes mittels Schnitten wie zuvor,

$$\mu_1 \otimes \mu_2(B) = \int \mu_1(B_y) d\mu_2(y), B \in \mathcal{A}_1 \otimes \mathcal{A}_2$$

ergibt

$$\begin{aligned} \mu_1 * \mu_2(S^{-1}(A)) &= \int \mu_1(A - y) d\mu_2(y) \text{ und} \\ &(\mu_1 \otimes \mu_2)S^{-1} = \mu_1 * \mu_2. \end{aligned}$$

Es ist auch in die andere "Richtung" darstellbar:  $\mu_1 * \mu_2 = \int \mu_2(A - x) d\mu_1(dx)$  □

Neben der Kommutativität  $\mu_1 * \mu_2 = \mu_2 * \mu_1$  besitzt die Faltung noch folgende Eigenschaften:

- $\mu_i, i = 1, 2, 3$  sind sigma-endliche Maße auf  $(\mathbb{R}, \mathcal{B})$

$$(\mu_1 * \mu_2) * \mu_3 = \mu_1 * (\mu_2 * \mu_3)$$

*Proof.*  $A \in \mathcal{B}_2$

$$\begin{aligned} (\mu_1 * \mu_2) * \mu_3(A) &= \int \mu_1 * \mu_2(A - z) d\mu_3(z) = \int \left( \int \mu_2(A - z - x) d\mu_1(x) \right) d\mu_3 = \\ &= \int \int \mu_2(A - z - x) d\mu_3(z) d\mu_1 = \int \mu_2 * \mu_3(A - x) d\mu_1(x) = (\mu_2 * \mu_3) * \mu_1(A) = \mu_1 * (\mu_2 * \mu_3)(A) \end{aligned}$$

□

- $\mu_1(\mathbb{R}) = \mu_2(\mathbb{R}) = 1 \implies \mu_1 * \mu_2(\mathbb{R}) = 1$

*Proof.* Da  $\mathbb{R} - y = \mathbb{R} \forall y \in \mathbb{R}$  gilt

$$\mu_1 * \mu_2(\mathbb{R}) = \int \mu_2(\mathbb{R} - x) d\mu_1(x) = \int \mu_2(\mathbb{R}) d\mu_1(x) = \mu_2(\mathbb{R}) \cdot \mu_1(\mathbb{R}) = 1$$

Die Faltung von Wahrscheinlichkeitsmaßen ist auch ein Wahrscheinlichkeitsmaß. □

Sind  $\mu_1 = PX_1^{-1}$  und  $\mu_2 = PX_2^{-1}$  von  $\text{SGn } X_1$  bzw.  $X_2$  induziert, dann ist  $\mu_1 * \mu_2 = PX_1^{-1} * PX_2^{-1} = P(X_1 + X_2)^{-1}$  das von  $X_1, X_2$  induzierte Maß, vorausgesetzt  $X_1, X_2$  sind unabhängig, also für  $X = (X_1, X_2), PX = PX_1^{-1} \otimes PX_2^{-1}$ .

- Es existiert auch ein neutrales Element der Faltung  $\mu_0$  mit  $\mu_0 * \mu = \mu = \mu * \mu_0$ .

*Proof.* Für  $\mu_0 = \delta_0$ , d.h.  $\delta_0(A) = 1_A(0)$  gilt

$$\mu_0 * \mu(A) = \int \mu_0(A - x) d\mu(x) = \int \underbrace{1_{A-x}(0)}_{1_A(x)} d\mu(x) = \int 1_A(x) d\mu(x) = \mu(A).$$

□

Die Faltung bildet eine kommutative Halbgruppe.

Die bereits vorher erklärte Faltung von messbaren Funktionen hängt erwartungsgemäß mit der Faltung von Maßen zusammen. Bei Maßen mit Dichten ist die Dichte der Faltung genau die Faltung der Dichten.

**Satz 11** (11.PR).  $\mu_1, \mu_2$  seien Maße mit Dichten bezüglich  $\lambda$ :  $\mu_1 = \int f_1 d\lambda$  und  $\mu_2 = \int f_2 d\lambda$  und  $f_1, f_2$  sind reellwertig. Dann ist

$$\mu_1 * \mu_2(A) = \int_A \left( \int_{\mathbb{R}} f_1(s - y) f_2(y) d\lambda(y) \right) d\lambda(s)$$

also die Dichte von  $\mu_1 * \mu_2$  ist die Faltung der Dichten  $f_1 * f_2$ .

*Proof.*

$$\mu_1 * \mu_2(A) = \int_{\mathbb{R}} \mu_1(A - y) d\mu_2(y) = \int_{\mathbb{R}} \int_{A-y} f_1(x) d\lambda(x) f_2(y) d\lambda(y)$$

Die Transformation  $T_y(x) = x - y$  ist als lineare Transformation für  $\lambda$  translationsinvariant. Für jedes  $y$  gilt

$$\lambda T_y^{-1}(\cdot) = \lambda(T_y^{-1}(\cdot)) = \lambda(\cdot)$$

Das innere Integral ist nach dem Transformationssatz

$$\int_{A-y} f_1(x) d\lambda = \int_{T_y(A)} f_1 d\lambda = \int_A f_1 \circ T_y(s) d\lambda(s)$$

und mit dem Satz von Fubini ergibt sich

$$\mu_1 * \mu_2(A) = \int_{\mathbb{R}} \left( \int_A f_1(s - y) d\lambda(s) \right) f_2(y) d\lambda(y) = \int_A \underbrace{\int_{\mathbb{R}} f_1(s - y) f_2(y) d\lambda(y)}_{f_1 * f_2} d\lambda(s)$$

Analog gilt natürlich auch

$$\mu_1 * \mu_2(A) = \int_A \left( \int f_2(s-x) f_1(x) d\lambda(x) \right) d\lambda(s).$$

□

Damit wurde die Faltung im stetigen Fall auf die Faltungsdichte zurückgeführt. Analog werden diskrete Maße gefaltet.

**Satz 12** (12.PR).  $\mu_i, i = 1, 2$  sind diskrete Maße auf  $(\mathbb{R}, \mathcal{B})$  mit Trägermenge  $D_i$  (abzählbar). Dann ist  $\mu_1 * \mu_2$  diskret verteilt mit der Trägermenge  $D = D_1 + D_2 = \{s | s = x + y, x \in D_1, y \in D_2\}$  und für  $s \in \mathbb{R}$  gilt

$$\mu_1 * \mu_2(s) = \sum_{y \in D_2} \mu_1(\{s - y\}) \mu_2(\{y\})$$

*Proof.*

$$\mu_1 * \mu_2(A) = \int_{D_2} \mu_1(A - y) d\mu_2(y) = \sum_{y \in D_2} \mu_1(A - y) \mu_2(\{y\})$$

Die Trägermenge von  $\mu_1 * \mu_2$  ist  $D$ , da  $D^c - y \subseteq D_1^c$  für  $y \in D_2$  und  $\mu_1(D_1^c) = 0$  gilt

$$\begin{aligned} \mu_1 * \mu_2(D^c) &= \sum_{y \in D_2} \underbrace{\mu_1(D^c - y)}_{=0} \mu_2(\{y\}) \\ &\implies \mu_1 * \mu_2(D^c) = 0 \end{aligned}$$

$\mu_1 * \mu_2(A) = \mu_1 * \mu_2(A \cap D)$  und für  $s \in D$  gilt

$$\mu_1 * \mu_2(\{s\}) = \sum_{y \in D_2} \mu_1(\{s - y\}) \mu_2(\{y\})$$

oder mit vertauschten  $\mu_i, i = 1, 2$

$$\mu_1 * \mu_2(\{s\}) = \sum_{x \in D_1} \mu_2(\{s - x\}) \mu_1(\{x\})$$

□

Für stochastischen Größen und die entsprechenden induzierten Wahrscheinlichkeitsmaße wurde diese diskrete Faltung bereits erklärt, o.B.d.A sei  $X$  auf  $\mathbb{N}$  diskret verteilt, genauso wie  $Y$ , dann ist

$$P(X + Y = k) = \sum_{m \geq 0} P(X = k - m) P(Y = m)$$

wenn  $X, Y$  unabhängig sind, bzw.

$$P(X + Y = k) = \sum_{m \geq 0} P(X = k - m | Y = m) P(Y = m)$$

wenn  $X, Y$  nicht unabhängig sind.

Mit der bedingten Dichte gilt auch im stetigen Fall für die Dichte von  $X + Y =: S$  bei abhängigen  $X$  und  $Y$

$$f_S(s) = \int f(s - t | Y = t) f_Y(t) d\lambda(t)$$

wobei  $f(x|y)$  die Dichte der bedingten Verteilung von  $X|Y$  bezeichnet. Auch hier können die marginalen Dichten getauscht werden, d.h. mit  $g(y|x)$  als bedingte Dichte  $Y|X$

$$f_S(s) = \int g(s - x | X = x) f_X(x) d\lambda(x).$$

Für die Binomialverteilung wurde die Faltung bereits durchgeführt.

**Beispiel** (Faltung negativer Binomialverteilung).  $X$  ist  $NegB_{k,p}$ , wenn  $X$  die Anzahl der Fehlversuche ( $\geq 0$ ) bis zum  $k$ -ten Erfolg ( $\leq 1$ ) bei unabhängigen 0-1 Versuchen mit Erfolgswahrscheinlichkeit  $p$ .

$$P(X = m) = \binom{m+k-1}{k-1} p^k (1-p)^m$$

Die geometrische Verteilung  $G_p$  ist ein Spezialfall,  $G_p = NegB_{1,p}$  (Variante 2 der geometrischen Verteilung ist hier gemeint).

$$X_i \sim G_p : P(X_i = m) = p(1-p)^m, m \in \mathbb{N}$$

Die Faltung zweier  $G_p$  heißt

$$P(X_1 + X_2 = m) = \sum_{l=0}^m p(1-p)^{m-l} p(1-p)^l = (m+1)p^2(1-p)^m = \binom{m+2-1}{2-1} p^2(1-p)^m$$

daher  $X_1 + X_2 \sim NegB_{2,p}$ .

Genauso ergibt die Faltung  $NegB_{k,p} * G_p =: \mu$

$$\mu(\{m\}) = \sum_{i=0}^m \binom{i+k-1}{k-1} p^k (1-p)^i p(1-p)^{m-i} = p^{k+1} (1-p)^m \sum_{i=0}^m \binom{i+k-1}{k-1} = p^{k+1} (1-p)^m \binom{k+m}{k}$$

Mit dem Pascalschen Dreieck  $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$  und Induktion kann gezeigt werden, dass  $\sum_{i=0}^m \binom{i+k-1}{k-1} = \binom{k+m}{k}$ .

Somit ist  $NegB_{k,p} * G_p = NegB_{k+1,p}$ . Mit Induktion gilt daher sofort  $NegB_{k,p} * NegB_{\bar{k},p} = NegB_{k+\bar{k},p}$ .

Die Summe unabhängiger  $NegB$ -Verteilungen (mit gleichem  $p$ ) ist wieder negativ Binomialverteilt.

Bei verschiedenen  $p$  ist die Summe nicht nach einer  $NegB$  verteilt.

Die Faltung eines (absolut) stetigen Maßes und eines diskreten Maßes ergibt wieder ein Maß mit Dichte.

**Satz 13** (13.PR).  $\mu$  sei ein Maß mit Dichte  $f$  bez.  $\lambda$ .  $\mu$  daher  $\mu \ll \lambda$  und  $\nu$  sei ein diskretes Maß mit Trägermenge  $D$  (abzählbar) auf  $(\mathbb{R}, \mathcal{B})$ . Dann ist  $\mu * \nu \ll \lambda$  und

$$\mu * \nu(A) = \int_A \underbrace{\sum_{k \in D} f(x-k) \nu(\{k\})}_{\text{Dichte von } \mu * \nu} d\lambda(x).$$

*Proof.* Da  $\mu \ll \lambda$  gilt für eine Nullmenge  $N$ ,  $\lambda(N) = 0$  auch  $\mu(N) = 0$ . Wegen der Translationsinvarianz  $\lambda(N-s) = 0$  für alle  $s \in \mathbb{R}$ , daher auch

$$\begin{aligned} \mu(N-s) &= 0 \text{ und } \mu * \nu(N) = \int \mu(N-s) \nu(ds) = 0 \\ &\implies \mu * \nu \ll \lambda. \end{aligned}$$

Sei  $A = (-\infty, t]$ ,  $t \in \mathbb{R}$ , dann gilt

$$\begin{aligned} \mu * \nu(A) &= \int \sum_{k \leq t-s, k \in D} \nu(\{k\}) f(s) d\lambda(s) = \int \sum_{k \in D} \nu(\{k\}) 1_{(-\infty, t]}(k+s) f(s) d\lambda(s) = \\ &= \sum_{k \in D} \nu(k) \int 1_{(-\infty, t]}(x) f(x-k) d\lambda(x) = \int_{(-\infty, t]} \sum_{k \in D} f(x-k) \nu(k) d\lambda(x) \end{aligned}$$

Die Aussage gilt daher für Intervalle. Wegen des Fortsetzungssatzes gilt sie auch für jedes  $A \in \mathcal{B}$ .  $\square$

**Beberkung.** Bei mehr als 2 Maßen genügt ein Maß  $\mu_i \ll \lambda$ .

**Beispiel.**  $\mu$  sei eine Exponentialverteilung  $E_{x_1}$  mit Dichte  $f(s) = e^{-s} 1_{(0, \infty)}(s)$ .  $\nu$  entspreche einer Geometrischen Verteilung  $G_{\frac{1}{2}}$  (Version 1).  $\nu(\{k\}) = \frac{1}{2^k}$ ,  $k \geq 1$ .

Die Lebesgue-Dichte des Faltungsmaßes ist

$$h(x) = \sum_{k \geq 1} \frac{1}{2^k} e^{-(x-k)} 1_{(0, \infty)}(x-k) = e^{-(x-1)} \frac{\left(\frac{e}{2}\right)^{\lfloor x \rfloor} - 1}{e-2}$$

was einer stückweisen (auf  $[n, n+1)$ ) gewichteten Exponentialfunktion (Verteilung) entspricht.

Die Faltung kann auch mit der Verteilungsfunktion durchgeführt werden:

$X$  besitzt  $F$  als Verteilungsfunktion  $Y$  hat  $G$  als VF,  $X$  und  $Y$  unabhängig. Die Verteilungsfunktion von  $X + Y$  ist

$$H(t) = \int F(t-s)dG(s) = \int G(t-s)dF(s)$$

(siehe Übung).

Wie im letzten Beispiel ist  $H(\cdot)$  bis auf eine Nullmenge differenzierbar, wenn  $F$  und  $G$  differenzierbar mit Dichten  $F' = f$  und  $G' = g$  sind. Die (stückweise) differenzierbare VF  $H$  ergibt die Dichte aus Satz 11.PR.

$$\frac{dH}{dx}(s) = \int \frac{dF}{dx}(s-t) \frac{dG}{dx}(t) d\lambda(t).$$

Natürlich wird auch die Verteilung der Differenz  $X - Y$  unabhängiger Sgn  $X, Y$  mit der Faltung von  $X$  und  $-Y$  bestimmt.

Wenn Dichten existieren ist die Dichte von  $X - Y$   $h(t) = \int f_X(s)f_Y(s-t)d\lambda(s)$ . Die Wahrscheinlichkeit für  $X \leq Y$  kann auch mit der FaltungsVF für  $X - Y$  berechnet werden.

**Lemma 1.**  $X, Y$  Sgn mit VF  $F_X, F_Y$  und Dichten  $f_X, f_Y$ . Wenn  $X$  u.a.  $Y$ , dann ist

$$P[X \leq Y] = \int_{\mathbb{R}} F_X(t)f_Y(t)d\lambda(t).$$

*Proof.*

$$P[X \leq Y] = \int_{\{x \leq y\}} f(x, y) d\lambda_2 = \int_{\mathbb{R}} \int_{(-\infty, Y]} f_X(x)f_Y(y) d\lambda_2(x, y) = \int_{\mathbb{R}} F_X(y)f_Y(y) d\lambda(y)$$

□

**Beispiel.**  $X, Y$  unabhängige Exponentialverteilungen mit Raten  $\lambda_1, \lambda_2$ , d.h.  $X \sim E_{\lambda_1}, Y \sim E_{\lambda_2}$ .

$$P(X \leq Y) = \int_{\mathbb{R}^+} (1 - e^{-\lambda_1 y}) \lambda_2 e^{-\lambda_2 y} d\lambda(y) = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

Anteil der Raten von  $X$ .

## 2 Signierte Maße und Zerlegungen

Die Summe von Maßen ist wieder ein Maß, die Differenz von Maßen ist eine zumindest sigma-additive Mengenfunktion.

**Definition 5** (signiertes Maß). Die Mengenfunktion  $\nu$  ist ein signiertes Maß, wenn auf  $(\Omega, \mathcal{A})$  gilt  $\nu : \mathcal{A} \rightarrow (-\infty, \infty]$  oder  $[-\infty, \infty)$  mit  $\nu(\emptyset) = 0$  und für disjunkte  $A_i \in \mathcal{A}$  ist  $\nu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$ .

Viele Eigenschaften von Maßen lassen sich auf signierte Maße übertragen, allerdings ein Integral über ein signiertes Maß macht Schwierigkeiten. Es stellt sich aber heraus, dass signierte Maße sich als Differenz zweier Maße darstellen lässt. Das und vieles folgende wäre sofort klar, wenn es für  $\nu$  eine Dichte  $f : \Omega \rightarrow \mathbb{R}$  bezüglich eines Maßes  $\mu$  gibt  $\nu(A) = \int_A f d\mu$ , was auch  $\nu \ll \mu$  zur Folge hat.

$$\nu(A) = \int_A f^+ - f^- d\mu = \underbrace{\int_A f^+ d\mu}_{\mu_1} - \underbrace{\int_A f^- d\mu}_{\mu_2}$$

Für  $A \subseteq [f^+ > 0]$  ist eine Menge mit  $\nu(A) \geq 0$  und auch jede Teilmenge  $B \subseteq A$  erfüllt  $\nu(B) \geq 0$ .

**Definition 6** (positive Menge, negative Menge).  $(\Omega, \mathcal{A}, \nu)$  sei signierter Maßraum.  $A^+$  heißt positive Menge bez.  $\nu$ , wenn  $\forall B \subseteq A^+ : \nu(B) \geq 0$  erfüllt,  $A^-$  ist  $\nu$ -negative Menge, wenn  $\nu(B) \leq 0$  für  $B \subseteq A^-$ .

Negative Mengen eines signierten Maßraumes bilden einen sigma-Ring (Bezeichnung  $\mathcal{A}^-$ ).  $\mathcal{A}^-$  ist  $\cap$ -stabil und  $\Delta$ -stabil. Die Monotonie folgt aus der Stetigkeit von signierten Maßen.

Sie Maße sind signierte Maße stetig von unten. Die monoton wachsende Folge  $A_n \nearrow A$  wird wieder mit disjunkten Teilmengen dargestellt,  $A = \bigcup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \dots$ .

Bei der Stetigkeit von oben wird noch  $|\nu(A_{n_0})| < \infty$  für ein  $n_0$  (damit für  $n \geq n_0$ ) verlangt.

**Definition 7** (Hahn-Zerlegung). Eine Hahn-Zerlegung von  $\Omega$  besteht aus  $\{P, P^c\}$ , wobei  $P$  eine positive und  $P^c$  eine negative Menge ist.

**Satz 14** (1.SZ). Jeder signierte Maßraum  $(\Omega, \mathcal{A}, \nu)$  besitzt eine Hahn Zerlegung von  $\Omega$ .

*Proof.* Sei  $\nu : \mathcal{A} \rightarrow (-\infty, \infty]$ .  $\mathcal{A}^-$  das System der negativen Mengen. Sei  $c := \inf_{B \in \mathcal{A}^-} \nu(B)$  daher gibt es eine Folge  $B_n$  mit  $\nu(B_n) < c + \frac{1}{n}$  und  $B_n \in \mathcal{A}^-$ .

$N := \bigcup_n B_n$  ist eine negative Menge. Sei  $P := N^c$ .

Angenommen  $P$  enthält eine negative Menge  $B$  mit  $\nu(B) < 0$ , dann wäre  $\nu(B \cup N) = \underbrace{\nu(B)}_{<0} + \nu(N) < c$

also  $c$  nicht das Infimum.

Demnach ist für jedes  $B \subseteq P$  und  $\nu(B) < 0$ .

$$\mathcal{E}_1 := \sup\{\nu(M) | M \subseteq B\} > 0$$

Es gibt ein  $M_1 \subseteq B$  mit  $\frac{\mathcal{E}_1}{2} \leq \nu(M_1) \neq \mathcal{E}_1$ .

Daher  $\nu(B \setminus M_1) = \nu(B) - \nu(M_1) < \nu(B) < 0$ . Auch  $B \setminus M_1 \notin \mathcal{A}^-$ , daher gibt es ein  $M_2$  mit  $\nu(B \setminus M_1 \setminus M_2) < 0$ .

Mit Induktion gibt es also eine Folge disjunkter Mengen  $M_n \subseteq B$  mit  $\nu(B \setminus \bigcup_{i=1}^n M_i) < 0$ .

Betrachte  $D = B \setminus \bigcup_{i=1}^{\infty} M_i \implies B = \bigcup M_i \cup D$ .

Für  $M_i$  gilt  $\nu(M_i) > 0$  und  $\nu(B) = \sum_i \nu(M_i) + \nu(D)$ , daher  $0 \leq \sum \nu(M_i) < \infty$  und  $\nu(D) < 0$ .

Somit muss  $\nu(M_i) \rightarrow 0$  und für  $C \subseteq D \subseteq B \setminus \bigcup_{i=1}^n M_i$  jede Teilmenge  $C$  erfüllt  $\nu(C) \leq \mathcal{E}_n$  mit  $\mathcal{E}_n \rightarrow 0 \implies \nu(C) \leq 0$ .

$D$  ist eine negative Menge mit  $\nu(D) < 0$ . Da  $D \subseteq P = N^c$  wäre  $N \cup D$  eine negative Menge mit  $\nu(N \cup D) < \nu(N) = 0$ . Damit kann  $c = \inf_{B \in \mathcal{A}^-} \nu(B)$  nicht stimmen, die Annahme  $\nu(B) < 0$  hält nicht,  $P$  ist positive Menge.  $\square$

Hahn-Zerlegungen sind nicht eindeutig, für  $\nu(\cdot) = \int f d\lambda$  kann der Trennungspunkt gewählt werden. Aber die symmetrische Differenz positiver Mengen ist eine Nullmenge:

Sei  $\{P_1, P_1^c\}$  und  $\{P_2, P_2^c\}$  jeweils eine Hahn-Zerlegung von  $\Omega$ . Dann gilt für  $A \subseteq P_1 \setminus P_2$ , dass  $\nu(A) \geq 0$  ( $A \subseteq P_1$ ) und  $\nu(A) \leq 0$  ( $A \subseteq P_2^c$ ), also  $\nu(A) = 0$ .

**Definition 8** (Jordan-Zerlegung). Eine Jordan-Zerlegung des signierten Maßes  $\nu$  bilden Maße  $\mu_1, \mu_2$  mit

$$\nu = \mu_1 - \mu_2,$$

wobei  $\mu_1, \mu_2$  singular zueinander sind,

$$\mu_1 \perp \mu_2 \text{ d.h. } \exists M \text{ mit } \mu_1(M) = \mu_2(M^c) = 0.$$

Die Singularität führt auf folgende Minimaleigenschaft der Jordan-Zerlegung.

Ist  $\nu^+, \nu^-$  eine Jordanzerlegung von  $\nu = \nu^+ - \nu^-$  und  $\nu = \mu_1 - \mu_2$  für beliebige Maße  $\mu_1, \mu_2$ , dann gilt  $\nu^+(A) \leq \mu_1(A)$  und  $\nu^-(A) \leq \mu_2(A)$  für alle  $A \in \mathcal{A}$ .

Mit obiger Trennungsmenge  $M$  gilt

$$\begin{aligned} \nu^+(A) &= \underbrace{\nu^+(A \cap M)}_{=0} + \nu^+(A \cap M^c) = \nu^+(A \cap M^c) = \nu(A \cap M^c) \\ \nu^-(A) &= \nu^-(A \cap M) + \underbrace{\nu^-(A \cap M^c)}_{=0} = \nu^-(A \cap M) = -\nu(A \cap M) \\ \implies \nu^+(A) &= \mu_1(A \cap M^c) - \mu_2(A \cap M^c) \leq \mu_1(A) \\ \nu^-(A) &= \mu_2(A \cap M) - \mu_1(A \cap M) \leq \mu_2(A) \end{aligned}$$

Mit der Hahn-Zerlegung findet man auch eine Jordan-Zerlegung.

**Satz 15** (2.SZ Jordan'scher Zerlegungssatz). Jedes signierte Maß  $\nu$  besitzt eine eindeutige Jordan-Zerlegung  $\nu = \mu_1 - \mu_2$ .

*Proof.* Aus der Hahn-Zerlegung  $\Omega = P \cup P^c$  erhält man mit

$$\mu_1(A) = \nu(A \cap P) \text{ und } \mu_2(A) = -\nu(A \cap P^c)$$

zwei singuläre Maße. Die Eindeutigkeit folgt aus der Minimaleigenschaft.

Sei  $\nu = \nu^+ - \nu^-$  mit singulären Maßen  $\nu^+, \nu^-$ , dann gilt  $\nu^+ \leq \mu_1 \wedge \nu^- \leq \mu_2$  und  $\mu_1 \leq \nu^+ \wedge \mu_2 \leq \nu^-$ .  $\square$

**Def** (Totalvariation). *Durch  $\mu_1 + \mu_2$  aus der Jordan-Zerlegung entsteht ein Maß, der Variation bzw. Totalvariation mit der Bezeichnung  $|\nu| = \mu_1 + \mu_2$  genannt wird. Die Rechtfertigung der Bezeichnung gilt.*

**Lemma 2.** *Für das signierte Maß  $\nu$  gilt*

$$\forall A \in \mathcal{A} : |\nu(A)| \leq |\nu|(A).$$

*Proof.* Die Aussage folgt sofort aus

$$|\nu(A)| = |\mu_1(A) - \mu_2(A)| \leq |\mu_1(A)| + |\mu_2(A)| = |\nu|(A).$$

$\square$

Die Zerlegungssätze für signierte Maße werden für sigma-endliche Maße konkretisiert.

**Definition 9** (Lebesgue-Zerlegung).  *$\nu$  und  $\mu$  seien sigma-endliche Maße auf  $(\Omega, \mathcal{A})$ . Eine Lebesgue-Zerlegung von  $\nu$  besteht aus Maßen  $\nu_c, \nu_s$ , sodass*

$$\nu = \nu_c + \nu_s \quad \text{und} \quad \nu_c \ll \mu \quad \text{und} \quad \nu_s \perp \mu$$

*also  $\nu_c$  absolut stetig bezüglich  $\mu$  und  $\nu_s$  singulär zu  $\mu$  ist.*

$\nu$  ist in Bezug auf  $\mu$  in einen stetigen und einen singulären Anteil zerlegt.

Die zentrale Aussage ist, dass eine Lebesgue-Zerlegung für sigma-endliche Maße immer existiert und eindeutig ist. Für den Beweis können verschiedene Wege genommen werden. Eine Möglichkeit ist die Verwendung des Satzes von Riesz.

**Satz 16** (Darstellungssatz von Riesz-Frechet).  *$(V, \langle \cdot, \cdot \rangle)$  sei ein Hilbertraum und  $H : V \rightarrow \mathbb{R}$  eine reellwertige Abbildung. Dann sind folgende Aussagen äquivalent:*

- *$H$  ist stetig und linear*
- *$\exists a \in V$  mit  $H(x) = \langle x, a \rangle$  und  $a$  ist eindeutig bestimmt.*

Der Hilbertraum hier ist  $\mathcal{L}_2(\mu)$  (also die quadratisch integrierbaren Funktionen) mit  $\langle f, g \rangle = \int f g d\mu$ .

Die Abbildung  $H : \mathcal{L}_2(\mu) \rightarrow \mathbb{R}$  ist also genau dann stetig und linear, wenn es ein  $f \in \mathcal{L}_2(\mu)$  gibt mit  $H(g) = \int g f d\mu \forall g \in \mathcal{L}_2$ .

**Satz 17** (3.SZ Zerlegungssatz von Lebesgue). *Ein sigma-endliches Maß  $\nu$  auf dem sigma-endlichen Maßraum  $(\Omega, \mathcal{A}, \mu)$  hat eine eindeutige Lebesgue Zerlegung.*

**Satz 18** (4.SZ Satz von Radon-Nikodym). *Für sigma-endliche Maße  $\mu, \nu$  gilt  $\nu \ll \mu \iff \nu$  ist Maß mit Dichte, d.h.*

$$\nu(A) = \int_A f d\mu$$

*mit f.ü. eindeutiger Dichte  $f \geq 0$ ;  $f = \frac{d\nu}{d\mu}$ .*

*Satz 3 und 4.* Da beide Maße als sigma-endlich vorausgesetzt werden, kann (wie üblich) sich auf den endlichen Fall beschränken. Sei für das Maß  $\gamma := \mu + \nu$ .

$\gamma(\Omega) < \infty$ . Für  $f, g \in \mathcal{L}^2(\gamma)$  gilt

$$\int |f - g| d\nu \leq \int |f - g| d\mu \leq (1 + \gamma(\Omega)) \|f - g\|_2 \text{ (Norm auf } \mathcal{L}^2(\gamma))$$

Das bedeutet  $h \rightarrow \int h d\nu$  ist eine stetige Abbildung  $\mathcal{L}_2(\gamma) \mapsto \mathbb{R}$  und natürlich auch eine lineare Abbildung.

Nach dem Satz von Riesz-Frechet existiert ein  $g \in \mathcal{L}_2(\gamma)$  mit

$$(1) \quad \int h d\nu = \int h g d\gamma \quad \forall h \in \mathcal{L}_2(\gamma) \text{ oder}$$

$$(2) \quad \int \tilde{h}(1-g) d\gamma = \int \tilde{h} d\mu \quad \forall \tilde{h} \in \mathcal{L}_2(\gamma).$$

Betrachtet man  $h = 1_{[g < 0]}$  so gilt

$$0 \leq \underbrace{\nu([g < 0])}_{=0} = \int h d\nu = \int 1_{[g < 0]} g d\gamma \leq 0$$

$$\text{da, } \underbrace{\int 1_{g < 0} g \nu}_{=0} + \underbrace{\int 1_{g < 0} g \mu}_{=0}$$

Das erste Integral ist über eine Nullmenge und das Zweite weil: Wäre  $\mu[g < 0] > 0$ , dann wäre das Integral  $< 0$ .

Betrachtet man  $\tilde{h} = 1_{[g > 1]}$  ist genauso

$$\mu([g > 1]) = \int 1_{[g > 1]}(1-g) d\gamma \leq 0$$

und  $g \leq 1$   $\gamma$ -f.ü.

Es gilt  $\gamma$ -f.ü.  $0 \leq g \leq 1$  und  $\tilde{\gamma} := (1-g)\gamma$  ist ein Maß mit Dichte  $(1-g)$ ,  $\tilde{\gamma}(A) = \int_A (1-g) d\gamma$ .

Für jedes  $h \geq 0$  existiert eine Folge  $h_n \in \mathcal{L}_2(\tilde{\gamma})$  mit  $h_n \nearrow h$  (beispielsweise sind Treppenfunktionen aus  $\mathcal{L}^2$ ). Nach dem Satz von der monotonen Konvergenz gilt

$$\int h d\tilde{\gamma} = \int \lim h_n d\tilde{\gamma} = \lim \int h_n d\tilde{\gamma} \stackrel{\text{wegen (2)}}{=} \lim \int h_n d\mu = \int h d\mu.$$

Also gilt für alle  $h \geq 0$  (1) und (2),

$$\int h(1-g) d\gamma = \int h d\mu$$

Sei  $E = [g = 1]$  und  $h = 1_E$ , dann folgt  $\int h d\mu = 0 \implies \mu([g = 1]) = 0$ .

Seien die (Spur-)Maße

$$\nu_c(A) = \nu(A \setminus E) \quad \text{und} \quad \nu_s(A) = \nu(A \cap E)$$

Offensichtlich ist  $\nu = \nu_c + \nu_s$  und die Maße sind singulär aufeinander, und  $\nu_s \perp \mu$  da  $\mu(E) = 0$  und  $\nu_s(E^c) = 0$ .

Sei  $A$  eine  $\mu$ -Nullmenge  $\mu(A) = 0$ , dann genügt  $A \cap E = \emptyset$  anzunehmen da  $\mu(e) = 0$ , dh.  $\int_A (1-g) d(\mu + \nu) = 0$  (nach 2), da  $1-g > 0$  auf  $A$  ist  $\mu(A) + \nu(A) = 0 \implies \nu(A) = 0$  und auch  $\nu_c(A) = 0$ .

Daher ist  $\nu_c \ll \mu$ .

Es bleibt noch die Dichte zu finden, sei

$$f := \frac{g}{1-g} 1_{E^c}.$$

$$\int_A f d\mu \stackrel{\text{wegen (2)}}{=} \int_A 1_{E^c} g d(\mu + \nu) \stackrel{\text{wegen (1)}}{=} \int 1_A 1_{E^c} d\nu = \nu(A \setminus E) = \nu_c(A)$$

und die Dichte ist  $f$ .

□

Der Satz von Radon-Nikodym ist damit auch gezeigt, da wenn  $\nu \ll \mu$  wird  $\nu_s \equiv 0$  in der Lebesgue-Zerlegung gesetzt.



**Beispiel.**  $\mu_i, i = 1, 2$  seien Poissonverteilungen mit Raten  $\lambda_i > 0$

$$\mu_1(\{k\}) = \frac{\lambda_1^k}{k!} e^{-\lambda_1}, k \in \mathbb{N}$$

$\mu_1 \ll \mu_2$  und  $\mu_2 \ll \mu_1$ , da  $\mu_i(\{k\}) > 0 \forall k \in \mathbb{N}$ .

$$\mu_2(\{k\}) = \left(\frac{\lambda_2}{\lambda_1}\right)^k e^{\lambda_1 - \lambda_2} \mu_1(\{k\})$$

$$\text{also } \frac{d\mu_2}{d\mu_1} = \left(\frac{\lambda_2}{\lambda_1}\right)^k e^{\lambda_1 - \lambda_2} \text{ und}$$

$$\mu_2(A) = \sum_{k \in A} \left(\frac{\lambda_2}{\lambda_1}\right)^k e^{\lambda_1 - \lambda_2} \mu_1(\{k\})$$

**Definition 10** (Äquivalenz). Die Maße  $\mu_1, \mu_2$  heißen äquivalent ( $\mu_1 \approx \mu_2$ ), wenn  $\mu_1 \ll \mu_2$  und  $\mu_2 \ll \mu_1$ .

**Lemma 3.** Die sigma-endlichen Maße  $\mu_1$  und  $\mu_2$  seien äquivalent  $\mu_1 \approx \mu_2$ . Dann gilt für die RN-Dichten

$$\frac{d\mu_1}{d\mu_2} = \left(\frac{d\mu_2}{d\mu_1}\right)^{-1} \mu_1\text{-f.ü.}$$

*Proof.* Da  $\mu_1 \ll \mu_2 \ll \mu_1$  ergibt die Anwendung der Kettenregel mit  $\mu_2 = f_2 \cdot \mu_1$

$$\mu_1(A) = \int_A f_1 d\mu_2 = \int_A \underbrace{f_1}_{\frac{d\mu_1}{d\mu_2}} \underbrace{f_2}_{\frac{d\mu_2}{d\mu_1}} d\mu_1$$

Da  $\mu_1(A) = \int_A 1 d\mu_1$  und die RN-Dichte f.ü. eindeutig ist, gilt  $\mu_1\text{-f.ü.}$

$$\frac{d\mu_1}{d\mu_2} \frac{d\mu_2}{d\mu_1} = 1 \implies \frac{d\mu_1}{d\mu_2} = \left(\frac{d\mu_2}{d\mu_1}\right)^{-1}$$

(„Kehrwert und Ableitung sind vertauschbar.“) □

Ein einfaches Beispiel zeigt, dass die (sigma-)Endlichkeit der Maße unverzichtbar ist.

**Beispiel.**  $\mathcal{A} = \{\Omega, \emptyset\}$ ,  $\mu_1(\emptyset) = 0 = \mu_2(\emptyset)$  und  $\mu_1(\Omega) = 1, \mu_2(\Omega) = \infty$ .  $\mu_1 \approx \mu_2$

$$\mu_1(\Omega) = \int_{\Omega} f d\mu_2$$

ist für keine messbare Funktion  $f \geq 0$  möglich.

Im obigem Beispiel (Poissonverteilung) waren  $\mu_i \ll \zeta$  (Zählmaß) und die Dichten wurden über den „Umweg“ über  $\zeta$  bestimmt.

### 3 Dichten und absolute Stetigkeit

Lebesgue-Stieltjes Maße auf  $(\mathbb{R}, \mathcal{B})$  besitzen eine Verteilungsfunktion  $F$  von  $\mu$ . Angenommen  $F$  ist überall stetig differenzierbar, dann gilt

$$\begin{aligned} \frac{dF}{dx} &= \frac{d}{dx} \mu((-\infty, x]) = f \text{ und} \\ \mu((-\infty, x]) &= \int_{-\infty}^x f(t) dt. \end{aligned}$$

Nach dem Hauptsatz der Differenzial- und Integralrechnung gilt

$$f = \frac{dF}{dx}$$

und die Dichte entspricht der Ableitung der Verteilungsfunktion.

Im Gegensatz zu den vorigen (Existenz-)Sätzen zu Dichten, besteht jetzt eine einfache Möglichkeit eine Dichte (bezüglich  $\lambda$ ) zu finden.

Für eine Verallgemeinerung des obigen Prinzips, wird die stetige Differenzierbarkeit gegen schwächere Eigenschaften der Dichten gewechselt.

**Definition 11** (Beschränkte Variation).  $f : [a, b] \rightarrow \mathbb{R}$  ist von beschränkter Variation mit Schranke  $M$ , wenn für jede endliche Partition des Intervalls,  $a = x_0 < x_1 < \dots < x_n = b$

$$\sum_{i=1}^n |f(x_i) - f(x_{i-1})| \leq M$$

**Definition 12** (Totalvariation). Die Totalvariation von  $f$  ist

$$V_a^b f := \sup \left\{ \sum_{i=1}^n |f(x_i) - f(x_{i-1})|, n \in \mathbb{N} \right\}$$

wobei das Supremum über alle solche Partitionen gebildet wird. Die Menge  $BV(a, b)$  beherbergt alle  $f$  von beschränkter Variation auf  $[a, b]$ .

**Beispiel.**  $f(x) = x \sin(1/x)$  mit  $f(0) = 0$  ist auf  $[0, 1]$  stetig, aber nicht von beschränkter Variation. Sei die Partition definiert durch

$$x_i = \left( \frac{\pi}{2} (1 + 2i) \right)^{-1}, i = 1, 2, \dots$$

womit

$$|f(x_i) - f(x_{i-1})| \geq \frac{1}{\pi(i + \frac{3}{2})}.$$

Die rechte Seite ist nicht summierbar, für jedes  $M$  existiert eine Anzahl  $n$  mit  $\sum^n |f(x_i) - f(x_{i-1})| > M$ .

**Lemma 4.** Wenn  $f \in BV(a, b)$  und  $a < c < b$  gilt

$$V_a^c f + V_c^b f = V_a^b f$$

*Proof.* Für eine beliebige Partition mit  $n$  Parametern des Intervalls  $(a, b)$  sei  $s(a, b) := \sum |f(x_i) - f(x_{i-1})|$  mit  $x_0 = a, x_n = b$ .

Ist  $c$  ein Punkt von  $a = x_0 < \dots < x_n = b$  gilt

$$s(a, c) + s(c, b) \leq V_a^b f$$

da die Partitionen von  $(a, c)$  und  $(c, b)$  beliebig waren, gilt auch

$$V_a^c f + V_c^b f \leq V_a^b f$$

Ist  $c$  kein Punkt und  $x_i < c < x_{i+1}$

$$s(a, b) \leq s(a, x_j) + |f(x_{j+1}) - f(c)| + |f(c) - f(x_j)| + s(x_{j+1}, b).$$

Auch das gilt für beliebige Partitionen, somit  $V_a^b f \leq V_a^c f + V_c^b f$  □

**Lemma 5.**  $f \in BV(a, b)$  dann sind auf  $(a, b)$

$$x \rightarrow V_a^x f \text{ und } x \rightarrow V_a^x f - f(x)$$

monoton wachsende Funktionen.

*Proof.*  $V(x) := V_a^x f$ . Für  $y > x$  gilt nach vorigem Lemma  $V_a^y f = V_a^x f + V_x^y f$ , d.h.  $V(\cdot)$  ist wachsend.

$$\begin{aligned} f(y) - f(x) &\leq |f(y) - f(x)| \leq V_x^y f = V_a^y f - V_a^x f \\ &\implies V_a^x f - f(x) \leq V_a^y f - f(y) \end{aligned}$$

□

Die Darstellung von Funktionen mit beschränkter Variation durch monotone Funktionen ist auch eine Charakterisierung.

**Satz 19** (1.DS).  $f \in BV(a, b)$  genau dann, wenn  $f = v - w$  mit monoton wachsenden Funktionen  $v, w$ .

*Proof.* Es ist nur mehr  $\Leftarrow$  zu zeigen.

Sei  $f = v - w$

$$\sum |f(x_i) - f(x_{i-1})| = \sum |v(x_i) - v(x_{i-1}) + w(x_{i-1}) - w(x_i)| \leq \sum |v(x_i) - v(x_{i-1})| + \sum |w(x_i) - w(x_{i-1})|.$$

Da jede monotone Funktion von beschränkter Variation ist, gilt

$$\sum |f(x_i) - f(x_{i-1})| \leq V_a^b v + V_a^b w < \infty.$$

□

Eine monotone Funktion hat nur höchstens abzählbar viele Unstetigkeitsstellen.  $f : [a, b] \mapsto \mathbb{R}$  monoton steigend, d.h.  $f_-(x) \leq f(x) \leq f_+(x)$  (also jeweils der linksseitige bzw. rechtsseitige Grenzwert).

Jede Unstetigkeitsstelle  $x_0$  erzeugt ein Intervall  $(f_-(x_0), f_+(x_0))$ , diese Intervalle sind eindeutig und disjunkt, wovon es nur höchstens abzählbar viele geben kann.

Eine Funktion beschränkter Variation,  $f \in BV(a, b)$  ist stetig  $\lambda$ -f.ü.

Wenn  $f \in BV(a, b)$  mit  $f = F - G$  und  $F, G$  monoton wachsend, dann kann  $F, G$  an der Unstetigkeitsstellen durch den rechtsseitigen Grenzwert ersetzt werden.

$F_+, G_+$  sind dann Verteilungsfunktionen von Lebesgue-Stieltjes Maßen.

$$f_+ := F_+ - G_+ = f \lambda\text{-f.ü.}$$

$f_+, f$  ist als Verteilungsfunktionen eines signierten LS-Maßes zu verstehen.

Eine Verteilungsfunktion  $F$  ist zerlegbar in  $F = F_s + F_d$ , wobei  $F_s$  stetig ist und  $F_d$  eine Verteilungsfunktion eines diskreten Maßes ist. (Abzählbare Trägermenge  $\{x_i\}_{i \in I}$  mit  $\mu(\{x_i\}) > 0$ ).

Der stetige Anteil ( $F_s$ ) lässt sich weiter zerlegen. Dazu betrachten wir absolut stetige Funktionen.

**Definition 13** (absolut stetig). Die Funktion  $f : [a, b] \rightarrow \mathbb{R}$  heißt absolut stetig, wenn zu beliebigem  $\epsilon > 0$  ein  $\delta > 0$  existiert, sodass aus  $\sum_{i=1}^n (b_i - a_i) < \delta$  folgt  $\sum_{i=1}^n |f(b_i) - f(a_i)| < \epsilon$ . Dabei sind  $(a_i, b_i)$  disjunkte Teilintervalle von  $[a, b]$ , die Anzahl der Intervalle ist  $n$  (beliebig).

Solche absolut stetigen Funktionen liegen "zwischen" stetig differenzierbaren und gleichmäßig stetigen Funktionen.

Wenn  $f \in C^1[a, b]$ , dann ist  $f$  absolut stetig, da

$$|f(b_i) - f(a_i)| \leq \sum_{x \in (a, b)} |f'(x)| (b_i - a_i), a_i < b_i$$

**Lemma 6.**  $f$  sei auf  $[a, b]$  absolut stetig. Dann ist  $f$  gleichmäßig stetig und  $f \in BV[a, b]$ .

*Proof.* Nach Voraussetzung existiert zu  $\epsilon > 0$  ein  $\delta > 0$  sodass aus  $(b - a) < \delta \implies |f(b) - f(a)| < \epsilon$  egal was  $a, b$  ist. Damit ist  $f$  gleichmäßig stetig.

Für eine Partition  $a = x_0 < x_1 < \dots < x_n = b$  mit  $(x_i - x_{i-1}) < \delta \forall i$  gilt für jede endliche Partition  $(a_i, b_i)$  des Intervalls  $(x_{i-1}, x_i)$

$$\sum_{i=1}^m |f(b_i) - f(a_i)| \leq V_{x_{i-1}}^{x_i} f < \epsilon$$

nach Voraussetzung der absoluten Stetigkeit.

$V_a^b f$  ist additiv, also

$$V_a^b f = \sum_{i=1}^n V_{x_{i-1}}^{x_i} f < n\epsilon$$

daher ist  $V_a^b f$  endlich.

□

Absolute Stetigkeit ist stärker als gleichmäßige Stetigkeit, aber glm. Stetigkeit ist nicht hinreichend. Lipschitz Stetigkeit wiederum ist hinreichend.

Wenn  $f \in C^1[a, b]$ , also stetig differenzierbar ist, dann ist  $f$  absolut stetig.

Für  $a_i, b_i \in [a, b]$  und  $\sup_{x \in [a, b]} |f'(x)| < \infty$  ist

$$|f(b_i) - f(a_i)| \leq \sup_{x \in [a, b]} |f'(x)| |b_i - a_i|$$

und daher ist  $f$  absolut stetig.

Im folgenden soll geklärt werden, wie weit die komfortable Situation einer stetig differenzierbaren Verteilungsfunktion mit daraus folgender Maßdichte auf absolut stetige Verteilungsfunktionen übertragen werden kann. Damit könnte dann auch die Dichte eines Maßes  $\mu$  mit  $\mu \ll \lambda$  bestimmt werden.

Für eine absolut stetige Funktion diene eine Verteilungsfunktion eines signierten Maßes als Vorlage. Das rechtfertigt folgender Satz.

**Satz 20** (2.DS).  $f : [a, b] \rightarrow \mathbb{R}$  ist absolut stetig. Dann ist  $f = F - G$ , wobei beide Funktionen  $F, G$  monoton wachsend und absolut stetig sind.

*Proof.* Zunächst wird gezeigt, dass  $x \rightarrow V_a^x f$  eine absolut stetige Funktion ist.

Für  $\epsilon > 0$  existiert ein  $\delta$ , sodass eine Partition  $P = \{(a_i, b_i) | i = 1, \dots, n\}$  disjunkter Intervalle mit  $\sum (b_i - a_i) < \delta$  existiert, sodass  $\sum_{i=1}^n |f(b_i) - f(a_i)| < \epsilon$ .

Wird jedes einzelne Intervall  $(a_i, b_i)$  wieder zerlegt,  $a_i \leq a_{i_0} < \dots < a_{i_{m_i}} = b_i$ , ist

$$\sum_{i=1}^n \sum_{j=1}^{m_i} (a_{i_j} - a_{i_{j-1}}) = \sum_{i=1}^n b_i - a_i < \delta$$

und nach Voraussetzung

$$\sum_{i=1}^n \sum_{j=1}^{m_i} \underbrace{|f(a_{i_j}) - f(a_{i_{j-1}})|}_{=V_{a_{i-1}}^{a_{i_i}} f} < \epsilon$$

Da die Zerlegung von  $(a_i, b_i)$  beliebig war, ist auch

$$\sum_{i=1}^n \sup \left\{ \sum_{j=1}^{m_i} |f(a_{i_j}) - f(a_{i_{j-1}})| \mid a_i = a_{i_0} < \dots < a_{i_{m_i}} = b_i \right\} = \sum_{i=1}^n V_{a_i}^{b_i} f = \sum_{i=1}^n \underbrace{|V_{a_i}^{b_i} f - V_{a_i}^{a_i} f|}_{=V_{a_i}^{b_i} f} < \epsilon$$

also  $v(x) = V_a^x f$  ist absolut stetig. Dann ist auch  $v(x) - f(x)$  absolut stetig (als Summe von absolut stetigen Funktionen) Nach vorigem Lemma sind  $v(\cdot)$  und  $v(\cdot) - f(\cdot)$  beide monoton wachsend.  $\square$

Die Bezeichnung "absolut stetig" für Maße  $\mu \ll \lambda$  kommt von einer der gleichmäßigen Stetigkeit entsprechenden Eigenschaft für Funktionen.

**Satz 21** (3.DS).  $\mu$  sei ein LS-Maß auf  $(\mathbb{R}, \mathcal{B})$  mit Verteilungsfunktion  $F$ .  $[a, b] \subseteq \mathbb{R}$ .  $\mu \ll \lambda$  ist äquivalent mit:

1.  $\epsilon, \delta$ -Kriterium: Zu  $\epsilon > 0$  existiert ein  $\delta > 0$ , sodass aus  $\lambda(A) < \delta \implies \mu(A) < \epsilon$ , wenn  $\mu$  endlich ist,  $A \in \mathcal{B}$ .
2.  $F|_{[a, b]}$  ist Verteilungsfunktion von  $\mu$  auf  $([a, b], \mathcal{B}|_{[a, b]})$  ist absolut stetig.

*Proof.* 1. Wenn das  $\epsilon, \delta$  Kriterium gilt und  $\lambda(A) = 0$  aber  $\mu(A) > 0$ , dann ist für  $\epsilon = \frac{\mu(A)}{2}$ , zwar  $\lambda(A) < \delta$  für alle  $\delta > 0$  aber es gilt nicht  $\mu(A) < \epsilon$ .

Es sei  $\mu \ll \lambda$ . Da  $\mu$  endlich ist (und daher natürlich sigma-endlich) existiert eine RN-Dichte  $f$  mit  $\mu(A) = \int_A f d\lambda$ .

Da  $\mu$  endlich  $\lambda([f = \infty]) = 0$   $\lambda([f > m]) \rightarrow 0$  für  $m \rightarrow \infty$ . Mit  $\Omega = [a, b]$  folgt

$$\int_{\Omega} f 1_{[f > m]} d\lambda \rightarrow 0$$

wegen der Stetigkeit von oben des Maßes  $\mu$ . Für  $\epsilon > 0$  wähle  $m$  mit

$$\int_{\Omega} f 1_{[f>m]} d\lambda < \frac{\epsilon}{2}$$

und  $\delta < \frac{\epsilon}{2m}$ , dann gilt

$$\begin{aligned} \mu(A) &= \int_A f 1_{[f>m]} d\lambda + \int_A f 1_{[f\leq m]} d\lambda \leq \frac{\epsilon}{2} + m\lambda(A), \text{ wenn} \\ \lambda(A) &< \delta \implies \mu(A) < \epsilon. \end{aligned}$$

Bemerkung: Teil 1) gilt für beliebige (endliche) Maße  $\mu, \nu$  mit  $\mu \ll \nu$ .

2. Wenn  $\mu \ll \lambda$  gilt nach dem  $\epsilon, \delta$  Kriterium für  $A = \bigcup_{i=1}^n (a_i, b_i]$  mit  $\lambda(A) = \sum_{i=1}^n (b_i - a_i) < \delta$

$$\mu(A) = \sum_{i=1}^n F(b_i) - F(a_i) = \sum |F(b_i) - F(a_i)| < \epsilon$$

und  $F$  ist absolut stetig.

Sei  $F$  absolut stetig und  $A \in \mathcal{B} \cap [a, b]$ .

$\lambda$  ist (auch) das äußere Maß, somit gilt

$$\lambda(A) = \inf \left\{ \sum_{i=1}^{\infty} (b_i - a_i) \mid A \subseteq \bigcup_{i=1}^{\infty} (a_i, b_i] \right\}$$

Sei  $A = N$  eine  $\lambda$ -Nullmenge,  $\lambda(N) = 0$ , dann existieren disjunkte Intervalle  $(a_i, b_i]$  mit  $N \subseteq \bigcup_i (a_i, b_i]$  und  $\sum_i (b_i - a_i) < \delta$

Für jede endliche Auswahl gilt  $\sum_{i=1}^n (b_i - a_i) < \delta$  und nach Voraussetzung

$$\sum_{i=1}^n F(b_i) - F(a_i) < \epsilon$$

für jedes  $\epsilon > 0$ . Das bedeutet

$$\mu(N) \leq \mu \left( \bigcup_{i=1}^{\infty} (a_i, b_i] \right) \leq \sum_{i=1}^n |F(b_i) - F(a_i)| < \epsilon$$

$N$  ist auch eine  $\mu$ -Nullmenge. □

**Satz 22** (4.DS). Eine Funktion  $F : [a, b] \rightarrow \mathbb{R}$  besitzt genau dann eine Integraldarstellung

$$F(x) = \int_{(a,x]} f d\lambda + c, x \in [a, b]$$

mit  $f \in \mathcal{L}_1(\lambda|_{[a,b]})$ , wenn  $F$  absolut stetig ist.

*Proof.*  $F$  als Verteilungsfunktion eines signierten Maßes aufgefasst erklärt den Zusammenhang:

Ist  $F$  absolut stetig, so existieren monoton wachsende Funktionen  $G, H$  mit  $F = G - H$  und beide  $G, H$  sind absolut stetig.

Die zugehörigen LS-Maße seien  $\mu_G, \mu_H$  und beide  $\mu_G \ll \lambda, \mu_H \ll \lambda$  mit RN-Dichten

$$\begin{aligned} g &:= \frac{d\mu_G}{d\lambda} \text{ und } h := \frac{d\mu_H}{d\lambda}, \text{ somit ist} \\ F(x) - F(a) &= \mu_G((a, x]) - \mu_H((a, x]) - F(a) \\ \implies F(x) &= F(a) + \int_{(a,x]} g - h d\lambda \end{aligned}$$

Besitzt  $F$  die Integraldarstellung mit  $f$

$$F(x) = F(a) + \int_{(a,x]} f d\lambda$$

dann ist  $f^+$  die Dichte des Maßes  $\mu_G$ ,  $f^-$  die von  $H$  und  $\mu_G \ll \lambda, \mu_G \ll \lambda$ . Die Verteilungsfunktionen beider Maße  $G, H$  sind absolut stetig.

Für  $x \in (a, b]$  ist

$$F(x) = G(x) - G(a) - (H(x) - H(a))$$

und  $F$  ist die Differenz zweier absolut stetigen Funktionen und auch absolut stetig.  $\square$

Für die Klärung, in wie weit die Integraldarstellung mit der Ableitung  $f = F'$  gelingt wird zuerst die Lebesgue-Zerlegung von  $\mu$  bezüglich  $\lambda$  untersucht

$$\mu = \mu_c + \mu_s, \mu_s \perp \lambda, \mu_c \ll \lambda$$

Für folgenden Satz, der den Zusammenhang zwischen Nullmengen und Ableitung angibt, wird ein Hilfssatz benötigt.

**Lemma 7.**  $A \in \mathcal{B}|_{[a,b]}$  für ein endliches Intervall  $[a, b]$  und  $\mu$  sei ein endliches Maß auf  $\mathcal{B}|_{[a,b]}$ . Wenn  $\mu(A) = 0$ , dann ist für  $n \in \mathbb{N}$

$$\lambda \left( x \in A \mid \lim_{h \searrow 0} \frac{\mu((x-h, x+h])}{2h} > \frac{1}{n} \right) = 0$$

*Proof.* ohne Beweis. Der Beweis ist eine Folgerung des Überdeckungslemmas von Vitali, Details findet man in Elstrodt, "Maß- und Integrationstheorie".  $\square$

**Satz 23** (5.DS).  $\mu$  sei ein endliches Maß auf  $\mathcal{B}|_{[a,b]}$  mit Verteilungsfunktion  $F(x) = \mu([a, x])$ . Wenn  $\mu(A) = 0$ , dann ist  $F' = 0$   $\lambda$ -f.ü. auf  $A$ .

*Proof.*  $F$  ist wachsend,  $F' = 0$  bedeutet

$$\frac{F(x+h) - F(x-h)}{2h} = \frac{\mu((x-h, x+h])}{2h} \rightarrow 0$$

wenn  $h \rightarrow 0$  für  $\lambda$ -f.ü. Punkte  $x \in A$ .

Wir zeigen

$$\lim_{h \rightarrow 0} \frac{\mu(x-h, x+h]}{2h} = \lim_{k \rightarrow \infty} \frac{\mu(x - \frac{1}{k}, x + \frac{1}{k})}{\frac{2}{k}} \quad (1)$$

sonst könnte die Menge

$$A_n := \left\{ x \in A \mid \lim_{h \searrow 0} \frac{\mu(x-h, x+h]}{2h} > \frac{1}{n} \right\}$$

nicht unbedingt messbar sein,  $A_n \in \mathcal{B}$ .

(1) gilt, da für  $\frac{1}{k+1} < h \leq \frac{1}{k}$

$$\frac{\mu(x - \frac{1}{k+1}, x + \frac{1}{k+1})}{\frac{1}{k+1}} \frac{1}{k+1} \leq \frac{\mu(x-h, x+h]}{h} \leq \frac{\mu(x - \frac{1}{k}, x + \frac{1}{k})}{\frac{1}{k}} \frac{1}{k}$$

Für  $k \rightarrow \infty$  gilt mit  $\frac{k+1}{k} \rightarrow 1$  auch (1) ( $\lim$  von abzählbaren Funktionen ist messbar).

Nach vorigem Lemma ist  $\lambda(A_n) = 0$  und daher auch der Grenzwert  $n \rightarrow \infty$ , also

$$\lambda \left( \bigcup_{n=1}^{\infty} A_n \right) = 0$$

und  $F' = 0$   $\lambda$ -f.ü.  $\square$

**Satz 24** (6.DS).  $\mu$  sei ein endliches Maß auf  $([a, b], \mathcal{B}_{[a, b]})$  und singulär zu  $\lambda$ ,  $\mu \perp \lambda$ . Die Verteilungsfunktion  $F_\mu$  von  $\mu$  erfüllt  $F' = 0$   $\lambda$ -f.ü. auf  $[a, b]$ .

*Proof.* Wenn  $\mu \perp \lambda$  existiert ein  $A \in \mathcal{B}_{[a, b]}$  mit  $\mu(A) = 0$  und  $\lambda(A^c) = 0$ . Nach Satz 5.DS gilt  $F' = 0$   $\lambda$ -f.ü. auf  $A$ . Da  $\lambda(A^c) = 0$  gilt das auch auf  $[a, b]$   $\lambda$ -f.ü.  $\square$

**Beispiel.** Wir haben die Cantor-Menge

$$C = \{x \in [0, 1] \mid x = \sum_{i=1}^{\infty} \frac{x_i}{3^i}, x_i = 0 \vee 2\}$$

bereits behandelt, mit  $\lambda(C) = 0$  und die Funktion für  $x \in C$

$$F_C(x) = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$$

bzw. die Fortsetzung auf  $[0, 1]$

$$\tilde{F}_C(x) = \begin{cases} F_C(x) & x \in C \\ \sup\{F_C(y) \mid y \in C, y \leq x\} & x \notin C \end{cases}$$

diskutiert.  $\tilde{F}_C$  ist monoton und stetig. Daher gehört  $\tilde{F}_C$  als Verteilungsfunktion zu einem Maß (sogar einem Wahrscheinlichkeitsmaß).

Die Verteilung heißt Cantor-Verteilung.

$\tilde{F}_C$  ist eine stetige Verteilungsfunktion eines zu  $\lambda$  singulären Maßes (Der Vollständigkeit wegen sei  $\tilde{F}_C(x) = 0, x \leq 0$  und  $\tilde{F}_C(x) = 1, x > 1$ ).

Die bisherigen Sätze sind für Verteilungsfunktionen auf endlichen Intervallen formuliert. Durch Erweiterung auf  $\mathbb{R}$  (stückweise Verknüpfung) bleiben die Aussagen für VF bzw. LS-Maße auf  $(\mathbb{R}, \mathcal{B})$  erhalten.

So ist beispielsweise die VF  $F$  eines LS-Maßes  $\mu$  auf jeder  $\mu$ -Nullmenge differenzierbar ( $\lambda$ -f.ü.) und  $F' = 0$   $\lambda$ -f.ü.

Zur Klärung, ob  $F'$  die Dichte des LS-Maßes ist, betrachten wir die einseitigen Ableitungen.

$$\begin{aligned} \text{rechtsseitig} \quad & \begin{cases} \delta^r f(x) := \lim_{n \rightarrow \infty} \sup_{x < y < x+1/n} \frac{f(y) - f(x)}{y - x} \\ \delta_r f(x) := \lim_{n \rightarrow \infty} \inf_{x < y < x+1/n} \frac{f(y) - f(x)}{y - x} \end{cases} \\ \text{linksseitig} \quad & \begin{cases} \delta^l f(x) := \lim_{n \rightarrow \infty} \inf_{x-1/n < y < x} \frac{f(y) - f(x)}{y - x} \\ \delta_l f(x) := \lim_{n \rightarrow \infty} \sup_{x-1/n < y < x} \frac{f(y) - f(x)}{y - x} \end{cases} \end{aligned}$$

**Satz 25** (7.DS). Für das LS-Maß  $\mu$  mit  $\mu \ll \lambda$  gilt  $F' = \frac{d\mu}{d\lambda}$   $\lambda$ -f.ü., wobei  $F$  die VF von  $\mu$  ist.

*Proof.* Es sei  $\delta \bar{F}$  die "obere" Ableitung  $\delta \bar{F} = \max(\delta^r F, \delta^l F)$ .

Die Menge  $[\delta \bar{F} > \frac{d\mu}{d\lambda}]$  soll eine  $\lambda$ -Nullmenge sein, d.h. für alle  $q \in \mathbb{Q}$  soll

$$\lambda[\delta \bar{F} > q > \frac{d\mu}{d\lambda}] = 0$$

sein.

Dazu betrachten wir

$$\nu(A) := \int_{A \cap [\frac{d\mu}{d\lambda} \geq q]} \left( \frac{d\mu}{d\lambda} - q \right) d\lambda$$

$\nu$  ist ein LS-Maß und  $\nu(\frac{d\mu}{d\lambda} < q) = 0$ . Nach Satz 5.DS gilt für die VF  $F_\nu$  von  $\nu$

$$F'_\nu = 0 \text{ auf } [\frac{d\mu}{d\lambda} < q] \text{ } \lambda\text{-f.ü.}$$

Das signierte Maß  $\gamma$  erfüllt

$$\gamma(A) := \mu(A) - \lambda(A) = \int_A \left( \frac{d\mu}{d\lambda} - q \right) d\lambda \leq \int_{A \cap [\frac{d\mu}{d\lambda} \geq q]} \left( \frac{d\mu}{d\lambda} - q \right) d\lambda = \nu(A)$$

Daher folgt für  $x, y \in \mathbb{R}$

$$\begin{aligned}\gamma((\min(x, y), \max(x, y)]) &= \mu((\min, \max]) - \lambda(\min, \max)q \\ \implies \frac{F(y) - F(x)}{y - x} - q &\leq \frac{F_\nu(y) - F_\nu(x)}{y - x} \\ \implies \delta\bar{F} - q &\leq F'_\nu = 0 \text{ auf } [\frac{d\mu}{d\lambda} < q] \\ \text{d.h. } \lambda\left(\frac{d\mu}{d\lambda} < q < \delta\bar{F}\right) &= 0\end{aligned}$$

Dieselbe Vorgangsweise für die "untere" Ableitung angewandt auf  $\delta(-F)$  liefert

$$\begin{aligned}\lambda\left(\delta F < \frac{d\mu}{d\lambda}\right) &= 0 \\ \implies F' &= \frac{d\mu}{d\lambda} \lambda\text{-f.ü.}\end{aligned}$$

□

Die VF  $F$  eines beliebigen LS-Maßes  $\mu$  auf  $(\mathbb{R}, \mathcal{B})$  wurde bereits in eine stetige VF  $F_0$  und eine diskrete VF  $F_d$  zerlegt. Nach der Lebesgue-Zerlegung des Maßes  $\mu_0$  (von  $F_0$  erzeugt)  $\mu_0 = \mu_c + \mu_s$   $\mu_c \ll \lambda$  und  $\mu_s \perp \lambda$  mit VFen  $F = F_c + F_s + F_d$ .

Die vorigen Ergebnisse zusammengefasst bedeutet,  $F$  ist  $\lambda$ -f.ü. differenzierbar

$$F' = \underbrace{F'_c}_{=\frac{d\mu}{d\lambda}} + \underbrace{F'_s}_{=0} + (F'_d)\lambda\text{-f.ü.}$$

Ein LS-Maß  $\mu$  mit  $F$  für das  $F' = 0$   $\lambda$ -f.ü. gilt, ist zu  $\lambda$  singulär. Das folgt daraus, dass der absolut stetige Teil  $\mu_c$  die Dichte  $F' = \frac{d\mu_c}{d\lambda}$   $\lambda$ -f.ü. besitzt.

$$\mu_c(A) = \int_A \underbrace{\frac{d\mu_c}{d\lambda}}_{=0 \lambda\text{-f.ü.}} d\lambda = 0$$

Aus der Zerlegung ergibt sich, dass für jede monoton wachsende Funktion

$$\underbrace{F(b) - F(a)}_{\mu([a, b])} \geq \underbrace{\int_{[a, b]} F' d\lambda}_{\mu_c([a, b])}$$

gilt und  $F$  ist  $\lambda$ -f.ü. differenzierbar.

Wenn  $F$  keine VF ist (also nicht rechtsstetig), wird  $F$  an den Unstetigkeitsstellen (sind höchstens abzählbar viele) geändert, ohne dass sich bezüglich  $\lambda$  etwas ändert, ( $F'$   $\lambda$ -f.ü. oder  $\int F' d\lambda$  etc).

Damit sind aber auch Funktionen von beschränkter Variation  $\lambda$ -f.ü. differenzierbar, da sie als Differenz monoton wachsender Funktionen darstellbar ist.

Die Aussagen zusammen ergeben

**Satz 26** (8.DS (Satz von Lebesgue für Differenzierbarkeit)).  $F : [a, b] \rightarrow \mathbb{R}$  sei von beschränkter Variation. Dann ist  $F$   $\lambda$ -f.ü. differenzierbar.

Ist  $F$  monoton steigend, dann ist  $F$   $\lambda$ -f.ü. differenzierbar und

$$F(b) - F(a) \geq \int_{[a, b]} F' d\lambda$$

Die einzelnen Ergebnisse dieses Abschnitts münden in der Formulierung des Hauptsatzes der Differential- und Integralrechnung. Die gewohnte Form des Hauptsatzes für das Riemann-Integral wird die absolute Stetigkeit nicht benötigt, dagegen gilt der Hauptsatz für  $\mathcal{L}_1$ -Funktionen in folgender Form.



**Satz 27** (9.DS (Hauptsatz der DI-Rechnung für das Lebesgue-Integral)).  $f : [a, b] \rightarrow \mathbb{R}$  sei  $\in \mathcal{L}_1$ , also  $\lambda$ -integrierbar. Dann ist

$$F(x) := \int_{[a,x]} f d\lambda$$

absolut stetig und  $\lambda$ -f.ü. differenzierbar mit  $F' = f$   $\lambda$ -f.ü.  
Ist  $F : [a, b] \rightarrow \mathbb{R}$  absolut stetig, dann existiert  $\lambda$ -f.ü.  $F'$  mit

$$F(x) - F(a) = \int_{[a,x]} F' d\lambda, x \in [a, b]$$

*Proof.*  $F$  kann zerlegt werden in

$$F(x) = F_1(x) - F_2(x) = \int_{[a,x]} f^+ d\lambda - \int_{[a,x]} f^- d\lambda$$

$F_1, F_2$  sind monoton und  $\mu^+(A) = \int_A f^+ d\lambda$  ist absolut stetig bezüglich  $\lambda$  und  $F'_1 = \frac{d\mu^+}{d\lambda} = f^+$  genauso mit  $f^-$ .

Ist  $F$  absolut stetig, dann ist das (signierte) Maß absolut stetig mit  $F' = \frac{d\mu}{d\lambda}$  und  $\int_A F' d\lambda = \int_A \frac{d\mu}{d\lambda} d\lambda = \mu(A)$  mit  $A = [a, x]$  ergibt den Hauptsatz.  $\square$

### 3.1 Weitere Eigenschaften der Dichten

Im folgenden seien die behandelten Maße als sigma-endlich angenommen.

- Wenn  $\nu_i \ll \mu, i = 1, \dots, n$ , dann ist  $\tau := \sum_{i=1}^n \nu_i$  ein sigma-endliches Maß,  $\tau \ll \mu$  und die Dichte ist auch die Summe

$$\frac{d\tau}{d\mu} = \sum_{i=1}^n \frac{d\nu_i}{d\mu} \mu\text{-f.ü.}$$

- $T$  sei  $\Omega \rightarrow \Omega'$  und  $\mathcal{A}/\mathcal{A}'$  messbar. Wenn  $\nu \ll \mu$ , dann sind die induzierten Maße  $\nu^T$  und  $\mu^T$  auf  $(\Omega', \mathcal{A}')$  auch  $\nu^T \ll \mu^T$  absolut stetig.

Da für  $B \in \mathcal{A}'$  mit  $\mu^T(B) = 0$

$$\mu^T(B) = \mu(T^{-1}(B)) = 0 \implies \nu(T^{-1}(B)) = 0.$$

$\mu^T, \nu^T$  sind auch endlich, da für  $B_n \nearrow \Omega'$  und  $\mu, \nu$  endlich sind.

Es kann aber die Dichte  $\frac{d\nu^T}{d\mu^T}$  keinerlei Verwandtschaft zu  $\frac{d\nu}{d\mu}$  haben. Das hängt davon ab, in wieweit  $T$  Nullmengen in Nullmengen abbildet.

- $\mu$  auf  $(\mathbb{R}, \mathcal{B})$  sei  $\mu \ll \lambda$  und  $T$  eine messbare (sogar monotone) Abbildung. Dann ist  $\mu^T$  i.a. nicht mehr  $\mu^T \ll \lambda$  absolut stetig.

Betrachtet man beispielsweise die Cantor-Funktion  $F_C : C \rightarrow [0, 1]$  bzw.  $F_C^{-1} : [0, 1] \rightarrow C$

$$\lambda^{F_C^{-1}}(C) = \lambda([0, 1]) = 1 \text{ aber } \lambda(C) = 0.$$

- Allgemein werden Nullmengen nicht von bijektiven Transformationen erhalten.

**Beispiel.**  $\Omega = \{1, 2\}, \mathcal{A} = 2^\Omega, T(x) = 3 - x, T : \Omega \rightarrow \Omega$  bijektiv, sei  $\mu(1) = 1, \mu(2) = 0$ , dann ist  $A = \{2\}$  eine Nullmenge, aber  $\mu^T(A) = \mu(T^{-1}(\{2\})) = \mu(1) = 1$

Für bijektive Abbildungen  $T : (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')$  kann die Dichte dargestellt werden.

Genauer muss  $T$  nur eine "Links-Inverse" haben. Es existiere ein  $\tilde{T} : (\Omega', \mathcal{A}') \rightarrow (\Omega, \mathcal{A})$  mit  $\tilde{T} \circ T = id_\Omega$ , also  $id_\Omega(\omega) = \omega, \omega \in \Omega$ . Dann kann mit dem allgemeinen Transformationssatz die Dichte von  $\nu^T$  bzw.  $\mu^T$  bestimmt werden.

**Satz 28** (10.DS).  $\nu, \mu$  sind sigma-endlich und  $\nu \ll \mu$ .  $T : \Omega \rightarrow \Omega'$  besitzt ein  $\tilde{T}$  mit  $\tilde{T} \circ T = id_\Omega$  dann gilt  $\nu^T \ll \mu^T$  mit der Dichte

$$\frac{d\nu^T}{d\mu^T} = \frac{d\nu}{d\mu} \circ \tilde{T} \mu^T \text{-f.ü.}$$

*Proof.*  $\nu^T \ll \mu^T$  ist klar.  $T$  ist injektiv, wenn  $T(x) = T(y) \implies x = \tilde{T} \circ T(x) = \tilde{T} \circ T(y) = y$ .  
 $\frac{d\nu}{d\mu} = f$ , dann gilt für  $B' \in \mathcal{A}'$

$$\nu^T(B') = \nu(T^{-1}(B')) = \int_{T^{-1}(B')} f d\mu = \int 1_{T^{-1}(B')}(\omega) f(\omega) d\mu = \int 1_{B'}(T(\omega)) f \circ \tilde{T} \circ T(\omega) d\mu(\omega)$$

nach dem Transformationssatz

$$= \int 1_{B'}(\omega') f \circ \tilde{T}(\omega') d\mu^T$$

Also ist  $\frac{d\nu^T}{d\mu^T} = f \circ \tilde{T}$  □

Betrachtet man Maße  $\mu \ll \lambda$ , dann ist auch  $\mu^T \ll \lambda$ , wenn  $T$   $\lambda$ -Nullmengen immer in  $\lambda$ -Nullmengen abbildet.

Sei  $T : \mathbb{R} \rightarrow \mathbb{R}$  bijektiv und differenzierbar ( $\lambda$ -f.ü.) und auch  $T^{-1}$  ist  $\lambda$ -f.ü. differenzierbar  $T' \neq 0$   $\lambda$ -f.ü. und  $T^{-1} \neq 0$ .  $T$  sei wachsend

$$\lambda^T((a, b]) = \lambda((T^{-1}(a), T^{-1}(b)]) = \int_{[T^{-1}(a), T^{-1}(b)]} d\lambda$$

Auf  $(a, b]$  sei  $T' \neq 0, T^{-1'} \neq 0$ , dann ist

$$\lambda^T(a, b] = \int 1_{(T^{-1}(a), T^{-1}(b)]} d\lambda(s = T(\omega)) = \int 1_{(a, b]}(s) |T^{-1'}(s)| d\lambda(s)$$

d.h. die Dichte von  $\lambda^T$  ist  $|T^{-1'}(s)| = \frac{d\lambda^T}{d\lambda}$ .

Die Bestimmung der Dichte einer Transformation einer stochastischen Größe  $Y = T(X)$ , wobei die Verteilung von  $X$ ,  $P^X \ll \lambda$  mit  $F_X$  als Verteilungsfunktion, erfolgt über die Verteilungsfunktion von  $X$ .  $T$  sei wieder differenzierbar mit  $T' \neq 0$  (auf  $[a, b]$ ) und wachsend.

$$F_Y(y) = P[Y \in (-\infty, y]] = P[T(X) \in (-\infty, y]] = P[X \in (-\infty, T^{-1}(y))] = F_X(T^{-1}(y))$$

und die Dichte von  $Y$  ist  $\lambda$ -f.ü.

$$f_Y(y) = f_X(T^{-1}(y)) |T^{-1'}(y)|$$

Diese beiden Beispiele für die Bestimmung von Dichten können auf  $\mathbb{R}^k$  verallgemeinert werden.  $T : \mathbb{R}^k \rightarrow \mathbb{R}^k$  sei ein Diffeomorphismus, also  $T$  ist bijektiv und die Jakobi-Determinanten

$$\left| \det \left( \frac{\delta T_i}{\delta x_j}(x) \right) \right| \neq 0, x \in \mathbb{R}^k \text{ und} \\ \left| \det \left( \frac{\delta T'_i}{\delta y_j}(y) \right) \right| \neq 0, y \in \mathbb{R}^k$$

dann ist  $\lambda^T \ll \lambda$  (hier ist  $\lambda = \lambda_k$  das  $k$ -dimensionale Lebesgue-Maß) und die Dichte ist

$$\frac{d\lambda_k^T}{d\lambda_k} = \left| \det \left( \frac{\delta T_i^{-1}}{\delta y_j} \right) (y) \right|.$$

Wenn  $X \in \mathbb{R}^k$  eine SG mit  $P^X \ll \lambda_k$  und Dichte  $\frac{dP^X}{d\lambda_k} = f_X$ ,  $T$  sei bijektiv, diffbar und  $|\det T^{-1}| \neq 0$ , dann gilt für  $Y = T \circ X$   $P^Y \ll \lambda_k$  mit der Dichte

$$f_Y(y) = f_X(T^{-1}(y)) \left| \det \left( \frac{\delta T_i^{-1}}{\delta y_j} \right) (y) \right|.$$

### 3.1.1 Stückweise Anwendung

$\mathbb{R}^k = \bigcup_{i=1}^n B_i$  und  $\underbrace{T|_{B_i}}_{T_i}$  hat obige Eigenschaften.

$$|J_i^{-1}| := \left| \det \frac{\delta T_i^{-1}}{\delta x_j} \right| \neq 0 \text{ auf } B_i \text{ für}$$

$$T := \sum_{i=1}^n T_i 1_{B_i} \text{ ist die Dichte von } Y = T \circ X$$

$$f_Y(y) = \sum_{i=1}^n f(T_i^{-1}) |J_i^{-1}|$$

### 3.2 Illustrative Beispiele

1.

$$\mu = \xi_N, X \sim B_{N,p} = \nu, \nu(k) = P[X = k] = \binom{N}{k} p^k (1-p)^{N-k}, k \leq N$$

$$T : x \mapsto N - x, x \leq N, (Tx = x, x \geq N)$$

$$Y = T \circ X = N - X$$

$$P[Y = y] = P[T(X) = y] = P[N - X = y] = P[X = N - y] = \nu(N - y) = \underbrace{\binom{N}{N-y}}_{\binom{N}{y}} p^{N-y} (1-p)^y$$

$$Y \sim B_{N,1-p} \text{ (Anzahl der "0")}$$

2.

$$X \sim \mathcal{N}_k(\mu, \Sigma), f_X(x) = \frac{1}{(2\pi)^{k/2} \sqrt{\det \Sigma}} \exp\{-1/2(x - \mu)^T \Sigma^{-1}(x - \mu)\}$$

$$A \text{ reguläre Matrix, } Y = AX, f_Y(y) = f_X(A^{-1}y) |A^{-1}|$$

$$f_Y(y) = \frac{1}{(2\pi)^{k/2} \sqrt{\det \Sigma}} \exp\{-1/2 \underbrace{(A^{-1}y - A^{-1}A\mu)^T \Sigma^{-1}(A^{-1}y - A^{-1}A\mu)}_{(y-A\mu)^T \underbrace{(A^{-1}\Sigma^{-1}A^{-1})}_{(A\Sigma A^T)^{-1} = \tilde{\Sigma}}(y-A\mu)}\} |A^{-1}|$$

$$\text{Wegen } \frac{|A^{-1}|}{\sqrt{\det \Sigma}} = \frac{1}{\sqrt{(\det A)^2 \det \Sigma}} = \frac{1}{\sqrt{\det \tilde{\Sigma}}}$$

$$Y \sim \mathcal{N}(A\mu, A\Sigma A^T)$$

3.  $X \sim \mathcal{N}(\mu, \sigma^2)$ ,  $Y = e^X$  log-Normalverteilung,  $X = \log(Y)$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\log Y - \mu}{\sigma}\right)^2} \frac{1}{|y|}, y > 0$$

4.  $(X_1, X_2)$  u.a. exponential

$$f(x_1, x_2) = e^{-x_1 - x_2}, x_1, x_2 \geq 0$$

$$Y = \begin{pmatrix} x_1^2 + x_2^2 \\ x_1 \end{pmatrix} = T \circ X, y_1 = x_1^2 + x_2^2, y_2 = x_1, x_1 = y_2, x_2 = \sqrt{y_1 - y_2^2}, y_1 \geq y_2^2$$

$$|J| = \begin{vmatrix} 2x_1 & 2x_2 \\ 1 & 0 \end{vmatrix} = |-2x_2| = 2|x_2|$$

$$|J^{-1}| = \frac{1}{2x_2} = \frac{1}{2\sqrt{y_1 - y_2^2}}$$

$$\text{Dichte von } Y : f_Y(y) = e^{-y_2 - \sqrt{y_1 - y_2^2}} \frac{1}{2\sqrt{y_1 - y_2^2}} \text{ für } y_1 > y_2^2 \geq 0$$

5.  $X \sim \mathcal{N}(0, 1)$ ,  $Y = X^2$  stückweise invertierbar auf  $\mathbb{R}^+$

$$T_1^{-1}(y) = \sqrt{y}, y \geq 0, B_1 = (0, \infty)$$

$$T_2^{-1}(y) = -\sqrt{y}, y < 0, B_2 = (-\infty, 0]$$

$$\text{allgemein gilt dann } f_Y(y) = f_X(\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| + f_X(-\sqrt{y}) \left| -\frac{1}{2\sqrt{y}} \right|$$

bei symmetrischer Dichte von  $X$ , also  $f_X(\sqrt{y}) = f_X(-\sqrt{y})$

$$f_Y(y) = \frac{1}{\sqrt{y}} f_X(\sqrt{y})$$

$$\mathcal{N}(0, 1) \text{ für } f_X : f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \frac{1}{\sqrt{y}} \sim \gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$

$\chi^2$ -Verteilung mit Freiheitsgrad 1

## 4 $L_p$ -Räume

Der Raum der integrierbaren Funktionen bilden einen Vektorraum  $\mathcal{L}(\mu)$  (genauer  $(\mathcal{L}(\mu), +, \mathbb{R})$ ) mit  $\|f\| = \int |f| d\mu$  sogar einen normierten Raum, also einen metrischen Raum, wenn man  $f = g[\mu]$  als gleich ansieht.

**Def** (Pseudonorm, Pseudometrik).  $\|\cdot\|$  bildet eine Pseudonorm (=Halbnorm), es gilt also

$$\|\alpha f\| = |\alpha| \|f\|$$

$$\|f + g\| \leq \|f\| + \|g\| \triangle\text{-Ungleichung (damit bereits } \|f\| \geq 0 \text{ und } \|0\| = 0)$$

*Norm:*  $\|f\| = 0 \implies f = 0$  Das ist vorläufig nicht erfüllt.

$\|\cdot\|$  führt zu einer Pseudometrik, also  $d(x, y) := \|y - x\|$  mit

$$d(x, y) = d(y, x); d(x, x) = 0; \text{ und } d(x, y) \leq d(x, z) + d(z, y)$$

*Metrik:*  $d(x, y) = 0 \iff x = y$

Die Funktionenräume der p-fach integrierbaren Funktionen bilden wesentliche, abstrakte Funktionenräume deren Analyse wichtig ist (besonders  $L_2$ ).

**Definition 14** (p-fach integrierbar).  $f$  heißt p-fach integrierbar für  $1 \leq p < \infty$  wenn

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

Raum:  $\mathcal{L}^p(\mu)$

Wenn  $f = g[\mu]$  als gleich angesehen wird, bildet dieser Raum einen normierten Vektorraum.

Wenn eine Funktion beschränkt ist, dann ist auch

$$\int |f|^p d\mu < \infty \forall p \geq 1$$

Es sei daher  $\|f\|_\infty := \inf\{K \geq 0 | \mu(|f| > K) = 0\} = \text{ess sup } |f|$  essentielles Supremum.

$\|f\|_\infty$  bildet genauso eine Pseudonorm.

Eine Norm (bzw. Metrik) entsteht durch Übergang zum Quotientenraum. Es sei

$$\mathcal{N} := \{f \text{ messbar und } f = 0[\mu] \text{ f.ü.}\}$$

**Definition 15** ( $L_p$ -Raum).

$$L^p(\Omega, \mathcal{A}, \mu) := \mathcal{L}^p(\mu) / \mathcal{N} = \{\bar{f} = f + g; f \in \mathcal{L}^p, g \in \mathcal{N}\}$$

$L^p$  enthält die Repräsentanten und alle Elemente aus einer Klasse haben die gleiche Norm:  $\|f\|_p = \|g\|_p$ , wenn  $f = g[\mu]$ .

Um zu zeigen, dass  $L^p(\mu)$  wirklich ein normierter Vektorraum ist, benötigen wir noch die  $\triangle$ -Ungleichung, die gleich folgt.

Die Norm (bzw. Metrik) ergibt eine Konvergenz Definition im  $L^p$ .

**Definition 16** (Konvergenz im p-ten Mittel).  $f_i$  sei eine Folge in  $\mathcal{L}^p(\mu)$ .

$$f_n \rightarrow^{L^p} f \text{ wenn } \|f_n - f\|_p \rightarrow 0, n \rightarrow \infty.$$

Für SG  $X$  in einem W-Raum  $(\Omega, \mathcal{A}, P)$  heißt

$$\begin{aligned} \int X dP &= \mathbb{E}X \text{ Erwartung} \\ \text{und } \int X^k dP &= \mathbb{E}X^k \text{ k-tes Moment} \\ \text{sowie falls } \mathbb{E}X < \infty: &\mathbb{E}|X - \mathbb{E}X|^k \text{ k-tes zentrales Moment} \end{aligned}$$

Besondere Rolle spielt das 2. zentrale Moment:

$$\text{Varianz } \mathbb{E}(X - \mathbb{E}X)^2 \text{ und } \|X - \mathbb{E}X\|_2 \text{ Streuung}$$

**Beispiel.**  $X \sim \mathcal{N}(\mu, \sigma^2)$  Normalverteilung besitzt alle Momente, da

$$\forall p: \mathbb{E}X^p = \int_{-\infty}^{\infty} x^p \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx < \infty$$

**Def.** Die Folge von SG  $X_i$  konvergiert gegen  $X$  im  $p$ -ten Mittel, wenn  $\mathbb{E}|X_n - X|^p \rightarrow 0$ .  
Schreibweise:  $X_n \rightarrow^{L^p} X$ .

Aus der Konvergenz der Integrale folgt natürlich i.a. nicht, dass die Funktionen konvergieren  $[\mu]$ .

**Beispiel.**  $X_n \dots$  SG mit Wahrscheinlichkeitsmaß  $P[X_n = 0] = 1 - \frac{1}{n^{p+1}} = 1 - P[X = n]$  auf  $\{0, n\}$ . Es sei  $X_0 \sim \delta_0$ . Dann gilt

$$\mathbb{E}|X_n - X_0|^p = n^p \frac{1}{n^{p+1}} = \frac{1}{n} \rightarrow 0 \implies X_n \rightarrow^{L^p} X_0$$

Aber  $\limsup X_n = \infty$  also  $X_n \not\rightarrow X_0$   $P$ -f.s.  $\lim X_n$  existiert nicht.

Auch, wenn die Funktionen beschränkt sind, muss nicht  $f_n \rightarrow^{L^p} f$  folgen.

**Beispiel.** 1.  $f_n := 1_{[a_n, b_n]}$  mit  $a_n = \frac{n - \lfloor \sqrt{n} \rfloor^2}{2\lfloor \sqrt{n} \rfloor + 1}$  und  $b_n = \frac{n + 1 - \lfloor \sqrt{n} \rfloor^2}{2\lfloor \sqrt{n} \rfloor + 1}$  bildet eine Folge auf  $([0, 1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$  die nicht punktweise ( $\lambda$ -f.ü.) konvergiert.

Da  $b_n - a_n = \frac{1}{2\lfloor \sqrt{n} \rfloor + 1} \rightarrow 0$  gilt  $\|f_n - 0\|_p \rightarrow 0 \forall p \geq 1$ .

Dass die Umkehrung zwischen punktweiser und  $L_p$ -Konvergenz auch nicht uneingeschränkt gilt, scheint überraschender zu sein.

2.  $f_n := \frac{1}{n^{1/p}} 1_{[0,n]}$  konvergiert glm.  $f_n \rightarrow 0$  aber

$$\begin{aligned} \int f_n^p d\lambda &= \frac{1}{n} \int_{[0,n]} d\lambda = 1 \\ \|f_n\|_p &\not\rightarrow \|f_0\|_p, f_0 = 0 \end{aligned}$$

In dem letzten Beispiel war  $\|f_n\|_p \rightarrow \|f_0\|_p$  nicht erfüllt.

Fügt man weitere Bedingungen ein, dann kann Äquivalenz zwischen  $\|f_n - f\|_p \rightarrow 0$  und  $f_n \rightarrow f \mu$ -f.ü. erreicht werden.

In Beispiel 2) ist die Folge konvergent auch konvergent im Maß

$$\lambda(|f_n - f_0| > \epsilon) = 0$$

wenn für  $\epsilon > 0$  gewählt wird, sodass  $\frac{1}{n^{1/p}} < \epsilon \implies n > (\frac{1}{\epsilon})^p$ .

Konvergenz im Maß garantiert nicht  $L_p$ -Konvergenz. Die Umkehrung gilt.

**Satz 29** (1.LP). Für  $1 \leq p \leq \infty$  gelte  $f_n \rightarrow^{L^p} f; \|f_n - f\|_p \rightarrow 0$ , dann gilt  $\forall \epsilon > 0: \lim_{n \rightarrow \infty} \mu(|f_n - f| > \epsilon) = 0, f_n \rightarrow^\mu f$

*Proof.* Folgt unmittelbar aus der Markov-Ungleichung (Satz 1.UG) mit  $\phi(x) = |x|^p$

$$\mu(|f_n - f| > \epsilon) \leq \frac{\int |f_n - f|^p d\mu}{\epsilon^p} = \frac{\|f_n - f\|_p^p}{\epsilon^p}$$

□

**Def** (vollständig). Ein metrischer Raum  $M$  heißt vollständig, wenn jede Cauchy-Folge konvergiert, d.h. Cauchy-Folge:  $x_n \in M : \forall \epsilon > 0 \exists N_\epsilon \forall m, n \geq N_\epsilon : d(x_m, x_n) < \epsilon$  hat Grenzwert  $x_0 \in M$ .

$L^p$ -Raum für  $1 \leq p \leq \infty$  ist vollständig (folgt später).

## 4.1 Wichtige Ungleichungen für p-integrierbare Funktionen

**Lemma 8.**  $p, q \in (1, \infty)$  und  $\frac{1}{p} + \frac{1}{q} = 1$  (zueinander konjugiert) für  $x, y \geq 0$  gilt  $xy \leq \frac{x^p}{p} + \frac{y^q}{q}$ .

*Proof.*  $f(x) := \frac{x^p}{p} + \frac{y^q}{q} - xy$ ;  $f' = x^{p-1} - y$ ;  $f'' = (p-1)x^{p-2} \geq 0$  konvex mit Minimierung in  $x_0 = y^{\frac{1}{p-1}}$  da  $\frac{p}{p-1} = q$  gilt  $x_0^p = y^q$  und  $f(x_0) = 0 \implies f(x) \geq 0$  □

**Satz 30** (Allgemeinerer Satz von Young (Young'sche Ungleichung)).  $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ ,  $f(0) = 0$ , strikt monoton,  $f(x) \rightarrow \infty, x \rightarrow \infty$ ,  $f$  stetig diffbar, also Umkehrfunktion  $f^{-1} = g$ . Mit den Stammfunktionen  $F(x) = \int_0^x f(t)dt$ ,  $G(y) = \int_0^y g(t)dt$  gilt  $xy \leq F(x) + G(y)$ .

Siehe Skizze. Spezialfall  $p = q = 2 : xy \leq \frac{x^2}{2} + \frac{y^2}{2} \iff (x - y)^2 \geq 0$

**Satz 31** (Hölder'sche Ungleichung). Für  $\frac{1}{p} + \frac{1}{q} = 1$  für  $p \in [1, \infty]$  (wenn  $p = 1$  ist  $q = \infty$ ) gilt für  $f \in \mathcal{L}^p, g \in \mathcal{L}^q$  (oder rechte Seite  $\infty$  möglich)

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

und daher  $f \cdot g \in \mathcal{L}^1$ .

*Proof.*  $p = 1$  bzw.  $p = \infty$  unmittelbar (UE)

$p \in (1, \infty)$  Definiere  $\tilde{f} := \frac{|f|}{\|f\|_p}, \tilde{g} := \frac{|g|}{\|g\|_q}$ . Nach Young  $\tilde{f}\tilde{g} \leq \frac{\tilde{f}^p}{p} + \frac{\tilde{g}^q}{q}$  eingesetzt ins Integral und Monotonie:

$$\begin{aligned} \int \tilde{f}\tilde{g} d\mu &\leq \frac{1}{p} \frac{\int |f|^p d\mu}{\|f\|_p^p} + \frac{1}{q} \frac{\int |g|^q d\mu}{\|g\|_q^q} = 1 \\ &\implies \int |fg| d\mu \leq \|f\|_p \|g\|_q \end{aligned}$$

□

Spezialfall:  $f, g \in \mathcal{L}^2 \implies \int |fg| d\mu \leq \|f\|_2 \|g\|_2$ . Für SG  $X, Y$  bezüglich  $(\Omega, \mathcal{A}, P)$  mit 2. Moment:  $\mathbb{E}XY \leq \mathbb{E}|X||Y| \leq \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}$ .

**Satz 32** (Minkowski-Ungleichung).  $f, g \in \mathcal{L}^p(\mu), 1 \leq p \leq \infty$  Dann gilt:

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

*Proof.*  $p = 1$  oder  $p = \infty$  Übung  
 $\infty > p > 1 :$

$$\begin{aligned} (f + g)^p &= (f + g)(f + g)^{p-1} = f(f + g)^{p-1} + g(f + g)^{p-1} \\ |f + g|^p &\leq |f| \cdot |f + g|^{p-1} + |g| \cdot |f + g|^{p-1} \text{ also durch integrieren} \\ \|f + g\|_p^p &\leq \int |f| \cdot |f + g|^{p-1} + \int |g| \cdot |f + g|^{p-1} \text{ mit Hölder} \\ &\leq \|f\|_p \cdot \|(f + g)^{p-1}\|_q + \|g\|_p \cdot \|(f + g)^{p-1}\|_q + \int |g| \cdot |f + g|^{p-1} \text{ mit } q = \frac{p}{p-1} \\ &= (\|f\|_p + \|g\|_p) \cdot \|(f + g)^{p-1}\|_q + \int |g| \cdot |f + g|^{p-1} = (\|f\|_p + \|g\|_p) \cdot \|f + g\|_p^{p-1} + \int |g| \cdot |f + g|^{p-1} \\ \|f + g\|_p^{p-1} &= \left( \int |f + g|^{(p-1)q} \right)^{\frac{p-1}{p}} = \left( \int |f + g|^p \right)^{\frac{p-1}{p}} = \|f + g\|_p^{p-1} \end{aligned}$$

Dividiert durch den letzten Wert ist

$$\|f + g\|_p = \frac{\|f + g\|_p^p}{\|f + g\|_p^{p-1}} \leq \|f\|_p + \|g\|_p$$

Die linke Seite ist definiert (also endlich) weil aus  $f, g \in \mathcal{L}^p$  folgt

$$|f + g|^p \leq (2 \cdot \max(|f|, |g|))^p \leq 2^p \cdot \underbrace{\max(|f|^p, |g|^p)}_{\leq |f|^p + |g|^p}.$$

□

Mit der Minkowski-Ungleichung ist jetzt  $L^p$  ein Vektorraum ( $f, g \in \mathcal{L}^p \implies f + g \in \mathcal{L}^p$ ) und  $\|f\|_p$  eine Norm mit der Dreiecksungleichung.

**Lemma 9** (Streuungsungleichung).  $SG X, Y \in \mathcal{L}^p$ .  $SD(X + Y) \leq SD(X) + SD(Y)$

Die Bezeichnung  $L_\infty$  kann man aus folgendem Zusammenhang rechtfertigen:

**Satz 33** (2.LP).  $\mu \dots$  endlich, dann gilt

$$\lim_{p \rightarrow \infty} \|f\|_p = \|f\|_\infty$$

(wenn  $f$  nicht  $p$ -fach integrierbar ist, gilt  $\infty$  auf beiden Seiten).

*Proof.*

$$\begin{aligned} \int |f|^p d\mu &\leq K^p \int d\mu = K^p \mu(\Omega) \quad K := \operatorname{ess\,sup}_c \{c | \mu(|f| > c) = 0\} (K = \infty \text{ möglich}) \quad |f| \leq K[\mu] \\ \implies \|f\|_p^p &\leq \|f\|_\infty^p \mu(\Omega) & \implies \|f\|_p &\leq \|f\|_\infty (\mu(\Omega)^{\frac{1}{p}}) \\ \limsup_{p \rightarrow \infty} \|f\|_p &\leq \|f\|_\infty & \text{da } \limsup_{p \rightarrow \infty} (\mu(\Omega)^{\frac{1}{p}}) &= 1 \end{aligned}$$

Die andere Ungleichung folgt aus der Markoff-Ungleichung  $\phi(x) = x^p$

$$\int |f|^p d\mu \geq C^p \mu(|f| > C) \forall C > 0 \quad \liminf_{p \rightarrow \infty} \|f\|_p \geq C \cdot \liminf_{p \rightarrow \infty} \underbrace{(\mu(|f| > C))^{\frac{1}{p}}}_{>0 \text{ wenn } C < \|f\|_\infty = K}$$

$$\begin{aligned} \implies \liminf_{p \rightarrow \infty} \|f\|_p &\geq C \forall 0 < C < \|f\|_\infty \implies \liminf_{p \rightarrow \infty} \|f\|_p \geq \|f\|_\infty \\ \text{Aus } \|f\|_\infty &\leq \liminf_p \|f\|_p \limsup_p \|f\|_p \leq \|f\|_\infty \implies \lim_p \|f\|_p = \|f\|_\infty \end{aligned}$$

□

Für endliche Maße gelten spezielle Aussagen, wie etwa:

**Satz 34** (3.LP, Schachtelung der  $L^p$ -Räume).  $\mu$  sei endlich.  $1 \leq p \leq \tilde{p} \leq \infty$ .

$$\mathcal{L}^{\tilde{p}}(\mu) \subset \mathcal{L}^p(\mu)$$

Also auf  $\|f\|_{\tilde{p}} < \infty$  folgt  $\|f\|_p < \infty$ .

*Proof.* Es sei  $\tilde{p} = \infty$ , da  $\|f\|_\infty = \int \{K > 0 | \mu(|f| > K) = 0\}$  also  $|f| \leq \|f\|_\infty [\mu]$ -f.s

$$\int |f|^p d\mu \leq \int \|f\|_\infty^p d\mu = \|f\|_\infty^p \mu(\mathbb{R}) < \infty$$

Sonst  $\tilde{p}$  bel.  $p < \tilde{p}$ : Da  $|f|^p = |f|^p \cdot 1_{(|f| \leq 1)} + |f|^p 1_{(|f| > 1)} \leq 1 + |f|^{\tilde{p}}$  und Minkowski-Ungl.

$$\int |f|^p d\mu \leq \int 1 d\mu + \int |f|^{\tilde{p}} d\mu = \mu(\Omega) + \|f\|_{\tilde{p}}^{\tilde{p}} < \infty$$

□

Für sigma-endliche (aber nicht endliche) Maße gilt diese Schachtelung i.a. nicht. Es gibt sogar sigma-endliche Maße  $\mu$ , wo die umgedrehte Schachtelung gilt  $\mathcal{L}^p(\mu) \subseteq \mathcal{L}^{\tilde{p}}(\mu)$ .

Weder die Schachtelung noch die Normkonvergenz gilt i.A. für nicht endliche Maße. Etwa  $f = 1$  ist für  $\lambda$ :  $\|f\|_p = 1 \forall p \geq 1, \|f\|_\infty = \infty$ .

Nach Satz 8.MS ist  $L_\infty$  ein Banachraum. Die Vollständigkeit wird für alle  $L_p$ -Räume nachgewiesen.

**Satz 35** (4.LP, Satz von Riesz-Fischer). *Eine  $L_p$  konvergente Folge  $f_n \in \mathcal{L}_p, 1 \leq p \leq \infty$  mit Grenzfunktion  $f$ ,  $\|f_n - f\|_p \rightarrow 0$ , führt auf  $f \in \mathcal{L}_p$ .  $L_p(\Omega, \mathcal{A}, \mu)$  sind für jedes  $p \geq 1$  Banachräume.*

*Proof.*  $p = \infty$  ist Satz 8.MS. Für endliches  $p \geq 1$  gehen wir den Umweg über die Konvergenz im Maß.  $f_n$  sei eine Cauchyfolge aus  $\mathcal{L}_p$ .  $\Rightarrow f_n$  ist Cauchyfolge im Maß.

Nach Satz 10.MS existiert eine  $\mu$ -fast glm. konvergente Teilfolge  $f_{n_k} \rightarrow f$   $\mu$ -fast glm.

Daher gilt auch  $f_{n_k} \rightarrow f$   $\mu$ -f.ü. Diese Teilfolge ist nach Voraussetzung auch  $L_p$ -konvergent.

Für  $\epsilon > 0 \exists n_\epsilon : \|f_n - f_m\|_p^p < \epsilon, n, m \geq n_\epsilon$  und für festes  $n \geq n_\epsilon \|f_n - f_{n_k}\|_p^p < \epsilon; n_k \geq n_\epsilon$ , dh.  $|f_n - f|^p = \liminf_{k \rightarrow \infty} |f_n - f_{n_k}|^p$

$$\int |f_n - f|^p d\mu = \int \liminf_k |f_n - f_{n_k}|^p d\mu \leq \liminf_k \int |f_n - f_{n_k}|^p d\mu < \epsilon \text{ mit Lemma von Fatou}$$

Somit ist  $f$  der Grenzwert der Folge  $f_n$ . Aus der Dreiecksungleichung folgt

$$\|f\|_p = \|f - f_n + f_n\|_p \leq \|f_n\|_p + \underbrace{\|f - f_n\|_p}_{< \epsilon^{\frac{1}{p}}} \leq \inf_n \|f_n\|_p + \epsilon^{\frac{1}{p}} < \infty$$

und  $f \in \mathcal{L}_p$ . □

$L_p$ -Räume sind Banachräume und  $L_2$  ist sogar ein Hilbertraum mit  $\langle f, g \rangle := \int f \cdot g d\mu$ .

Es folgt sofort, dass Grenzwerte in  $\mathcal{L}_p$   $\mu$ -f.ü. eindeutig sein müssen,

$$\begin{aligned} f_n \xrightarrow{L_p} f, f_n \xrightarrow{L_p} g : \|f - g\|_p &\leq \|f - f_n\|_p + \|f_n - g\|_p < \epsilon, n > n_\epsilon \\ \Rightarrow \|f - g\|_p &= 0, \int |f - g|^p d\mu = 0 \Rightarrow |f - g| = 0 \Rightarrow f = g \mu\text{-f.ü.} \end{aligned}$$

Die Existenz einer Teilfolge, die konvergiert, wird sich oft als sehr nützlich erweisen.

Das ist speziell für separable Räume (abzählbar dichte Teilmenge) der Fall.

Sei  $([0, 1], \mathcal{A}, \lambda)$ :  $L_p$  separabel  $p \geq 1$ ,  $L_\infty$  nicht!

#### 4.1.1 $L^p$ für $0 < p < 1$

TODO

## 4.2 Dualraum von $L_p$

Prinzipiell existiert zu einem normierten linearen Raum  $(V, \|\cdot\|)$  der Vektorraum der linearen Funktionale  $T : V \rightarrow \mathbb{R}, T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ , die auch stetig sind. Um die Stetigkeit definieren zu können, muss  $V$  ein Banachraum sein.

**Def** (Dualraum, beschränkt). *Der Dualraum  $V'$  zu  $V$  ist  $V' = \{F | F : V \rightarrow \mathbb{R}, \text{ linear und stetig}\}$ . Auch  $V'$  wird durch  $\|T\| := \sup\{|T(x)| : \|x\| \leq 1\}$  zu einem Banachraum.  $T$  heißt beschränkt, wenn  $\|T\| \leq C < \infty$ .*

$T$  ist genau dann linear und stetig, wenn  $T$  linear und beschränkt ist. (UE)

Wir haben den Satz von Riesz-Frechet für die Darstellung der Elemente von  $V'$  bereits verwendet. Wenn  $(V, \langle \cdot, \cdot \rangle)$  ein Hilbertraum ist, dann hat  $F \in V'$  eine eindeutige Darstellung durch ein  $f \in V$  mit  $F(v) = \langle v, f \rangle$ .

Für Hilberträume können  $V$  und  $V'$  in dem Sinn identifiziert werden, dass eine Isometrie zwischen  $V$  und  $V'$  existiert.

Für den Banachraum  $(L^p, \|\cdot\|_p)$  kann eine solche Isometrie gefunden werden.

**Lemma 10.**  $1 \leq p \leq \infty$  Zu  $f \in \mathcal{L}^p$  existiert eine Folge  $t_n \in \mathcal{L}^p$  von Treppenfunktionen mit  $\|t_n\|_p \leq \|f\|_p \forall n$  und  $\lim_{n \rightarrow \infty} \|f - t_n\|_p = 0$ .



*Proof.* Für die Konstruktion des Integrals wurde beispielsweise die Folge von Treppenfunktionen

$$t_n(w) := \begin{cases} n & f(w) \geq n \\ \frac{k-1}{2^n} & \frac{k-1}{2^n} \leq f(w) \leq \frac{k}{2^n} \end{cases}$$

verwendet.  $|t_n| \nearrow |f|$  mit  $|t_n| \leq |f|$ , daher  $\|t_n\|_p \leq \|f\|_p$  und  $|f - t_n|^p \leq 2^p |f|^p \in L^p$

Da  $\lim_{n \rightarrow \infty} |f - t_n|^p = 0$  gilt nach dem Satz der dominierten Konvergenz

$$\lim_{n \rightarrow \infty} \|f - t_n\|_p^p = \lim_{n \rightarrow \infty} \int |f - t_n|^p = \int \underbrace{\lim_{n \rightarrow \infty} |f - t_n|^p}_{=0} = 0$$

für  $p < \infty$ .

Bei  $p = \infty$  konvergieren obige  $t_n$  sogar gleichmäßig gegen  $f \in L_\infty$  und  $\|f - t_n\|_\infty \rightarrow 0$ . □

Zwischen  $L^q$  und  $(L^p)'$  sei die kanonische Abbildung

$$K : L^q \rightarrow (L^p)' \text{ mit } K(f)(g) := \int f \cdot g d\mu, f \in L^q, g \in L^p$$

$K$  ist eine lineare Abbildung.

**Satz 36** (8.LP). Für  $\frac{1}{p} + \frac{1}{q} = 1$  ist die kanonische Abbildung  $K$  eine Isometrie, also  $\|K(f)\|'_p = \|f\|_q$

*Proof.*  $K(f)$  ist aus dem Dualraum von  $L^p$  mit der Norm (des Dualraums)

$$\|K(f)\|'_p := \sup \left\{ \underbrace{|K(f)(g)|}_{| \int f g | \leq \int |f g| \leq \|f\|_q \|g\|_p = \|f\|_q} : \|g\|_p = 1, g \in L^p \right\}$$

mit  $f \in L^q$ .

Daher gilt  $\|K(f)\|'_p \leq \|f\|_q$

Für die umgekehrte Ungleichung betrachte

$$\begin{aligned} \|K(f)\|'_p &\geq \left| \int f \cdot \frac{g}{\|g\|_p} d\mu \right| \forall g \in L^p \\ &\implies \|K(f)\|'_p \|g\|_p \geq \left| \int f g d\mu \right| \end{aligned}$$

Ersetze  $g$  durch  $\tilde{g} := |g| \cdot \text{sign}(f)$ , dann ist  $|\tilde{g}| \leq |g|$  (wo  $f = 0$ , ist  $\tilde{g} = 0$ ) und auch  $\|\tilde{g}\|_p \leq \|g\|_p$

$$\implies \|K(f)\|'_p \|g\|_p \geq \|K(f)\|'_p \cdot \|\tilde{g}\|_p \geq \underbrace{\left| \int f \cdot \tilde{g} \right|}_{\geq 0} = \int |f g| d\mu = \|f g\|_1$$

Fallunterscheidung

1.  $q = 1$ , d.h.  $f \in L^1$ , wähle  $g = 1 \in \mathcal{L}_\infty$ .  $\implies \|K(f)\|'_p \geq \|f\|_1$

2.  $q \in (1, \infty)$  wähle  $g = |f|^{q-1} = |f|^{\frac{q}{p}}$  da  $\int g^p = \int |f|^q < \infty$ ,  $f \in L^q \implies g \in L^p$

$$\begin{aligned} \|K(f)\|'_p \|g\|_p &\geq \|f g\|_1 = \| |f|^q \|_1 = \|f\|_q^q = \|f\|_q \cdot \underbrace{\|f\|_q^{q-1}}_{(f|f|^q)^{\frac{q-1}{q}}} = \|f\|_q \cdot \left( \int g^p \right)^{\frac{1}{p}} = \|f\|_q \|g\|_p \\ &\implies \|K(f)\|'_p \geq \|f\|_q \end{aligned}$$

3.  $q = \infty$  also  $\|f\|_\infty < \infty$

Für jedes  $\epsilon > 0$  ist  $\{|f| > (1 - \epsilon)\|f\|_\infty\} =: A_\epsilon$  keine Nullmenge.  $0 < \mu(A_\epsilon) < \infty$ .

Wähle  $g = \frac{1}{\mu(A_\epsilon)} 1_{A_\epsilon}$  ist aus  $L_1$ ,  $\|g\|_1 = 1$

$$\implies \|K(f)\|'_p \geq \|f g\|_1 = \int |f g| d\mu = \int |f| |g| d\mu = \int \underbrace{|f| 1_{A_\epsilon}}_{> (1-\epsilon)\|f\|_\infty} \frac{1}{\mu(A_\epsilon)} d\mu$$

für alle  $\epsilon > 0$ , d.h.  $\|K(f)\|'_p \geq \|f\|_\infty$

In allen Fällen folgt mit oben  $\|K(f)\|'_p = \|f\|_q$ .

□

Die kanonische Abbildung liefert eine Isometrie und ist daher injektiv.  
Angenommen  $f, g \in L^q$  und  $f \neq g$  in  $L^q$ , d.h.  $\|f - g\|_q \neq 0$ .

$$\begin{aligned}\|K(f) - K(g)\|'_p &= \|K(f - g)\|'_p = \|f - g\|_q \neq 0 \\ \implies \|K(f) - K(g)\|'_p &\neq 0 \text{ und } K(f) \neq K(g)\end{aligned}$$

$K(\cdot)$  ist auch surjektiv, daher gilt:

**Satz 37** (9.LP).  $1 \leq p < \infty$  mit  $\frac{1}{p} + \frac{1}{q} = 1$ .

Dann ist  $L^q$  isomorph zu  $(L^p)'$ , dem Dualraum von  $L^p$ . Man kann dann  $(L^p)' \simeq L^q$  also  $(L^p)'$  mit  $L^q$  identifizieren.

ohne Beweis. Für den Nachweis der Surjektivität sei auf die Literatur (Buch der VO) verwiesen.

**Beispiel.** Wenn  $\Omega = \mathbb{N}$  und  $\mathcal{A} = 2^{\mathbb{N}}$ ,  $\mu = \zeta$  (Zählmaß), dann ist  $L^p$  ein Folgenraum mit der Bezeichnung  $l^p$ .  $(a_n) \in l^p$ , wenn  $\sum_{n=1}^{\infty} |a_n|^p < \infty$ .

Die kanonische Abbildung für  $a \in L^q, b \in L^p, a = (a_n)_{n \in \mathbb{N}}, b = (b_n)_{n \in \mathbb{N}}$  ist

$$K(a)(b) = \sum_{i=1}^{\infty} a_i \cdot b_i$$

und entsprechend dem letzten Satz gilt auch  $(l^p)' \simeq l^q$ .

Für endliches  $\Omega$ , etwa  $\Omega = \{1, \dots, m\}$  und  $a = (a_1, \dots, a_m)$  und  $b = (b_1, \dots, b_m)$  ist auch  $(L^\infty)' \simeq L^1$ .

Wenn  $a \in l^\infty$  (endliche Folge ist immer beschränkt) und  $b \in l^1$  (endliche Folge immer summierbar) existiert für jedes  $b \in l^1$  auch ein  $a \in l^\infty$  mit  $K(a) \simeq b$ .

Im Allgemeinen gilt Satz 9.LP für  $p = \infty$  nicht.

**Beispiel.**  $(\mathbb{N}, 2^{\mathbb{N}}, \zeta)$  und  $l^p$  sei der Folgenraum. Die Teilmenge  $(\hat{l}_\infty)$  der beschränkten Folgen  $l_\infty$  (Spur) sei die Menge der konvergenten Folgen. Definiere für  $(a_n) \in \hat{l}_\infty$

$$F((a_n)) = \lim_{n \rightarrow \infty} a_n$$

$F$  ist linear.  $F$  ist auch stetig:  $a_{n_k} \rightarrow b_n, k \rightarrow \infty$ , d.h.  $\limsup_{k \rightarrow \infty} |a_{n_k} - b_n| = 0$ .  $(b_n)$  konvergent gegen  $b_0$ .

$$\begin{aligned}|a_{n_k} - b_0| &\leq |a_{n_k} - b_n| + |b_n - b_0| \\ \limsup_k |a_{n_k} - b_0| &\leq \limsup_k \underbrace{|a_{n_k} - b_n|}_{\rightarrow 0} + \lim_{n \rightarrow \infty} \underbrace{|b_n - b_0|}_{\rightarrow 0} \\ &= \lim_k F((a_{n_k})) = F((b_n))\end{aligned}$$

$F$  ist nur auf der Teilmenge definiert, kann bei Beibehaltung der Eigenschaften (linear, stetig) auf  $l_\infty$  erweitert werden. Ein  $(b_n) \in l_1$  kann es nicht geben, sodass  $\sum a_n b_n = \lim a_n$  für  $a_n \in l_1 \cap l_\infty$ .

### 4.3 Gleichmäßige Integrierbarkeit

Für die Konvergenz im  $L_p$  werden zur punktweisen Konvergenz noch weitere strukturelle Eigenschaften der Familie von Funktionen aus  $L_p$  benötigt. Ein Fall, wo aus  $f_n \rightarrow f$   $\mu$ -f.ü. doch  $L_p$ -Konvergenz folgt, ist gegeben, falls die Integrale  $\|f_n\|_p$  konvergieren.

**Satz 38** (10.LP). Für  $1 \leq p < \infty$  und die Folge  $f_n \in L_p$  gelte  $f_n \rightarrow f$   $\mu$ -f.ü. und  $\|f_n\|_p \rightarrow \|f\|_p$  dann ist  $f_n$   $L_p$ -konvergent,  $\lim_{n \rightarrow \infty} \|f_n - f\|_p = 0$

*Proof.* Da  $|f + f_n|^p \leq (2 \cdot \max(|f|, |f_n|))^p \leq 2^p(|f|^p + |f_n|^p)$  gilt

$$2^p \left( \int |f|^p + \int |f_n|^p \right) = \int \liminf ((2^p(|f|^p + |f_n|^p)) - |f - f_n|^p) d\mu \text{ da } \lim |f - f_n|^p = 0$$

wegen Fatou ist weiter

$$\leq \liminf \left( 2^p \int |f|^p + 2^p \int |f_n|^p - \int |f - f_n|^p \right) = 2^p \int |f|^p + 2^p \lim \int |f_n|^p + \liminf (- \int |f - f_n|^p)$$

Insgesamt ist

$$2^p \left( \int |f|^p + \int |f_n|^p \right) \leq 2^p \left( \int |f|^p + \int |f_n|^p \right) - \limsup \int |f - f_n|^p$$

das ist nur möglich, wenn  $\limsup \int |f - f_n|^p = 0$  also  $\|f_n - f\|_p \rightarrow 0$ .  $\square$

Für  $p = \infty$  gilt obige Aussage auch für sigma-endliche Maße i.A. nicht.

**Beispiel.**  $(\mathbb{N}, 2^{\mathbb{N}}, \zeta)$  mit dem (auf  $\mathbb{N}$ ) sigma-endlichen Zählmaß  $\zeta$ . Sei  $f_n = 1_{\{1, \dots, n\}}$ , dann gilt  $\|f_n\|_{\infty} = 1$  und  $f_n \rightarrow f = 1$  auf  $\mathbb{N}$  mit  $\|f\|_{\infty} = 1$ , aber  $|f_n - f| = 1_{\{n+1, \dots\}}$  mit  $\|f_n - f\|_{\infty} = 1 \not\rightarrow 0$ .

Dass für  $p \in [1, \infty)$  die Umkehrung nicht ohne weiteres gilt, wurde im Beispiel A (LP 4) demonstriert. Dort war  $\|f_n - 0\|_p \rightarrow 0$  für jedes  $p$ , aber  $f_n$  konvergiert nicht punktweise.

Die Umkehrung gelingt, wenn punktweise Konvergenz durch die (schwächere) Konvergenz im Maß ersetzt wird. Im Beispiel A (LP 4) gilt

$$\lambda(|f_n - f| > \epsilon) = \lambda([a_n, b_n]) = \frac{1}{2[n] - 1} \rightarrow 0.$$

**Satz 39** (11.LP). Für  $1 \leq p < \infty$  konvergiert die Folge  $f_n$  im  $p$ -ten Mittel,  $\|f_n - f\|_p \rightarrow 0$  genau dann, wenn  $f_n \rightarrow^\mu f$  im Maß mit  $f \in L_p$  und  $\|f_n\|_p \rightarrow \|f\|_p$

*Proof.*  $\Rightarrow$  Aus  $\|f_n - f\|_p$  folgt Konvergenz im Maß (Satz 1.LP) und wegen der Vollständigkeit ist  $f \in L_p$ . Die Konvergenz der Integrale ergibt sich aus der (umgekehrten) Dreiecksungleichung.  $\|f\|_p \leq \|f - f_n\|_p + \|f_n\|_p$  und  $\|f_n\|_p \leq \|f - f_n\|_p + \|f\|_p$

$$|\|f\|_p - \|f_n\|_p| \leq \|f - f_n\|_p \rightarrow 0 \Rightarrow \|f_n\|_p \rightarrow \|f\|_p$$

Wenn  $f_n \rightarrow f$  im Maß, dann betrachtet man wieder eine Teilfolge  $f_{n_k}$  mit  $f_{n_k} \rightarrow f$   $\mu$ -f.ü. (sogar  $\mu$ -fast gleichmäßig). Nach dem vorigen Satz folgt daraus auch  $\|f_{n_k} - f\|_p \rightarrow 0$ .

Angenommen  $f_n$  konvergiert nicht in  $L_p$ , dann existiert eine Teilfolge  $f_{m_j}$ , die  $\|f - f_{m_j}\|_p \geq \epsilon \forall j$  aber  $f_{m_j} \rightarrow f$  im Maß, dann gibt es wieder eine Teilfolge  $f_{m_{j_k}}$  mit  $\|f - f_{m_{j_k}}\|_p \rightarrow 0, k \rightarrow \infty$ , wie oben. Das ist ein Widerspruch, und  $\|f_n - f\|_p \rightarrow 0$ .  $\square$

Aus der Konvergenz im Maß wird eine Konvergenzart von Maßen abgeleitet.

Wenn für alle  $A \in \mathcal{A}$   $\mu_n(A) \rightarrow \mu(A)$  für Maße  $\mu$  und  $\mu_n$  auf  $(\Omega, \mathcal{A})$  konvergiert  $\mu_n$  schwach gegen  $\mu$  (bzw. in Totalvariation), d.h.

$$\lim_{n \rightarrow \infty} \sup |\mu_n(A) - \mu(A)| = 0.$$

Bei LS-Maßen mit Verteilungsfunktionen  $F_n, F$  bedeutet das für  $A = (-\infty, x]$

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

wenn die Maße endlich sind. Es gibt gute Gründe hier die Konvergenz nur für die Stetigkeitspunkte  $C_F$  von  $F(\cdot)$  zu fordern,  $x \in C_F : \lim_{n \rightarrow \infty} F_n(x) = F(x)$ .

Allgemein kann die Konvergenz auch auf die Dichten (falls vorhanden) zurückgeführt werden.

**Satz 40** (12.LP, Satz von Scheffe).  $\nu_n, \nu$  sind endliche Maße mit  $\nu_n(\Omega) = \nu(\Omega)$ .  $\mu$  sei sigma-endlich und  $\nu_n \ll \mu, \nu \ll \mu$ . Dann gilt

$$\lim_{n \rightarrow \infty} \frac{d\nu_n}{d\mu} = \frac{d\nu}{d\mu} \Rightarrow \sup_{A \in \mathcal{A}} |\nu_n(A) - \nu(A)| \rightarrow 0$$

*Proof.* Die punktweise Konvergenz der RN-Dichten und Konvergenz der Integrale bringt nach Satz 10.LP

$$\left| \int_A \frac{d\nu_n}{d\mu} - \int_A \frac{d\nu}{d\mu} \right| \leq \int_A \left| \frac{d\nu_n}{d\mu} - \frac{d\nu}{d\mu} \right| d\mu \leq \int \left| \frac{d\nu_n}{d\mu} - \frac{d\nu}{d\mu} \right| d\mu = \left\| \frac{d\nu_n}{d\mu} - \frac{d\nu}{d\mu} \right\|$$

Da  $\left\| \frac{d\nu_n}{d\mu} \right\| = \nu_n(\Omega) \rightarrow \nu(\Omega) = \left\| \frac{d\nu}{d\mu} \right\|$  nach Voraussetzung liefert der vorige Satz die Aussage.  $\square$

Der Satz gilt auch (wie im Beweis), wenn  $\nu(\Omega) \rightarrow \nu(\Omega)$ . Meist wird aber die Verteilungskonvergenz von W-Maßen mit Dichten untersucht.

Diskrete Verteilung:  $(\mathbb{N}, 2^{\mathbb{N}})$  mit Zählmaß als dominierendes Maß. Es genügt die Konvergenz der Punktwahrscheinlichkeiten für Verteilungskonvergenz  $P_n \rightarrow P$ .

**Beispiel.**  $P_n \sim H_{N,M,k}$  Hypergeometrische Verteilung wobei für  $n, N_n, M_n, \in \mathbb{N}$  Folgen mit  $N_n \rightarrow \infty$  und  $\frac{M_n}{N_n} \rightarrow p, n \rightarrow \infty$  für  $0 < p < 1$ .

$$\text{Somit } P_n(\{i\}) = \frac{\binom{M_n}{i} \binom{N_n - M_n}{k-i}}{\binom{N_n}{k}}, i \in \mathbb{N}$$

Asymptotisch gilt für den Binomialkoeffizienten  $\binom{n}{x} \sim \frac{n^x}{x!}, n \rightarrow \infty$  da  $\binom{n}{x} = \frac{n^x}{x!} \frac{n(n-1)\cdots(n-x+1)}{n^x} \rightarrow \frac{n^x}{x!}, n \rightarrow \infty$ .

Daher

$$P_n(\{i\}) \sim \frac{\frac{M_n^i}{i!} \frac{(N_n - M_n)^{k-i}}{(k-i)!}}{\frac{N_n^k}{k!}} = \binom{k}{i} \left( \frac{M_n}{N_n} \right)^i \left( 1 - \frac{M_n}{N_n} \right)^{k-i}$$

$$\lim_{n \rightarrow \infty} P_n(\{i\}) = \binom{k}{i} p^i (1-p)^{k-i}$$

Binomial  $B_{k,p}$ . Die Hypergeometrische Verteilung wird durch die Binomialverteilung approximiert,  $N \rightarrow \infty, H_{N,M,k} \sim B_{k,p=M/N}$ . (Ziehung mit (Binomial) und ohne (Hypergeom.) Zurücklegen.) Erwartungswert  $k \frac{M}{N} \rightarrow kp$

**Beispiel.**  $P_n \sim B_{n,p_n}$  Binomialv.; es gelte  $n \rightarrow \infty$  und  $p_n \rightarrow 0$  mit  $np_n \rightarrow \theta$ .

$$P_n(\{k\}) = \binom{n}{k} p_n^k (1-p_n)^{n-k}$$

$$\Rightarrow P_n(\{k\}) \sim \frac{n^k}{k!} \underbrace{p_n^k}_{\rightarrow \theta^k} (1-p_n)^n \underbrace{(1-p_n)^{-k}}_{\rightarrow 1}$$

Für jede Nullfolge  $p_n \rightarrow 0$  gilt  $\lim_{p_n \searrow 0} (1-p_n)^{\frac{1}{p_n}} \rightarrow e^{-1}$  und wegen  $np_n \rightarrow \theta$  folgt

$$P_n(\{k\}) \sim \frac{n^k}{k!} e^{-\theta} \text{ Poisson mit Rate } \theta \quad B_{n,p_n} \sim Poi_\theta$$

Die Erwartungswerte für  $X_n \sim B_{n,p_n}, X \sim Poi_\theta$  konvergieren  $\mathbb{E}X_n = np_n \rightarrow \theta = \mathbb{E}X$ .

Auch bei stetigen Verteilungen werden Verteilungskonvergenzen über die Konvergenz der Dichten untersucht.

**Beispiel.**  $(\mathbb{R}, \mathcal{B}, \lambda), X_n \sim \gamma(n, 1)$  mit Dichte

$$\frac{dP^{X_n}}{d\lambda}(x) = \frac{x^{n-1}}{(n-1)!} e^{-x}, x \geq 0$$

Gamma-Verteilung mit  $\mathbb{E}X_n = n$

Sei  $Y_n := \frac{X_n - n}{\sqrt{n}} = \frac{X_n}{\sqrt{n}} - \sqrt{n}$  mit der Dichte nach dem Transformationssatz für Dichten

$$f_n := \frac{dP^{Y_n}}{d\lambda}(y) = \frac{(\sqrt{n}y + n)^{n-1} e^{-\sqrt{n}y} e^{-n}}{(n-1)!} \sqrt{n}$$

Die Sirling'sche Formel

$$n! \sim \sqrt{2\pi} \sqrt{n} n^n e^{-n}$$

ergibt

$$f_n(y) \sim \frac{1}{\sqrt{2\pi}} \frac{(\sqrt{n}y + n)^{n-1}}{n^{n-1}} e^{-\sqrt{n}y}$$

Wegen

$$n \log(1 + \frac{x}{\sqrt{n}}) - x\sqrt{n} \rightarrow -\frac{x^2}{2}$$

für  $n \rightarrow \infty$ , gilt

$$f_n(y) \sim \frac{1}{\sqrt{2\pi}} (1 + \frac{y}{\sqrt{n}})^n e^{-\sqrt{n}y} \text{ und} \\ \lim_{n \rightarrow \infty} f_n(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

Die Grenzverteilung ist die Standardnormalverteilung  $P^{Y_n} \rightarrow N(0, 1)$  und  $\mathbb{E}Y_n = \frac{n}{\sqrt{n}} - \sqrt{n} = 0$ .

Die Verteilung im letzten Beispiel entsteht als standardisierte Summe unabhängiger Exponentialverteilungen  $EX_1, Z_i \sim EX_1$  u.a.  $Y_n = \frac{\sum Z_i - \sum \mathbb{E}Z_i}{\sqrt{\sum \text{Var}(Z_i)}} \rightarrow N(0, 1)$  diese Summen konvergieren allgemein.

Ein großer Vorteil für eine Familie von Funktionen ist gegeben, wenn die Integrale der Funktionen "gleichgradig" sind, also sich einheitlich verhalten. Das wäre der Fall, wenn eine integrierbare Funktion  $g \in \mathcal{L}_1$  existiert, sodass  $|f_i| \leq g$ .

Eine Funktion  $f$  ist genau dann integrierbar, wenn ein  $g \in \mathcal{L}_1$  existiert, sodass

$$\int_{|f|>g} |f| d\mu = 0$$

Eine andere Konsequenz der Integrierbarkeit gibt das folgende Lemma.

**Lemma 11.**  $f \in \mathcal{L}_1$  auf  $(\Omega, \mathcal{A}, \mu)$ , dann existiert für jedes  $\epsilon > 0$  ein  $c_\epsilon > 0$  und ein  $A_\epsilon$  mit  $\mu(A_\epsilon) < \infty$  mit

$$\int_{|f|>c_\epsilon} |f| d\mu < \epsilon \text{ und } \int_{A_\epsilon^c} |f| d\mu < \epsilon$$

*Proof.* Da  $f \in L_1$  muss  $\mu([f = \infty]) = 0$  sein. Für die Folge  $f_n := |f|1_{[|f|>n]}$  gilt  $f_n$  fallend,  $f_n \rightarrow |f|1_{[|f|=\infty]}$  und  $f_n \rightarrow 0$   $\mu$ -f.ü.

$f_n \leq |f|$  und mit dominierter Konvergenz

$$\lim_n \int f_n d\mu = \lim \int_{[|f|>n]} |f| d\mu = 0$$

Genauso für  $A_n := [|f| \geq \frac{1}{n}]$  und  $A_n^c = [|f| < \frac{1}{n}]$ ,  $A_n^c \searrow [|f| = 0]$

$$\lim_{n \rightarrow \infty} \int_{A_n^c} |f| d\mu = \int \lim_{n \rightarrow \infty} |f| 1_{A_n^c} = 0$$

Da  $\int |f| \geq \int |f| 1_{[|f|>\frac{1}{n}]} \geq \frac{1}{n} \int 1_{[|f|>\frac{1}{n}]}$  ist  $\mu(A_n) < \infty$  und  $\mu(A_n^c) \rightarrow 0$ . □

Die Erweiterung obiger Eigenschaften auf eine Familie integrierbarer Funktionen ist folgende Definition.

**Definition 17** (gleichmäßig integrierbar). Die Familie  $f_i, i \in I$  auf  $(\Omega, \mathcal{A}, \mu)$  heißt gleichmäßig integrierbar oder gleichgradig integrierbar wenn für jedes  $\epsilon > 0$  ein  $c_\epsilon > 0$  und ein  $A_\epsilon \in \mathcal{A}$  mit  $\mu(A_\epsilon) < \infty$  existieren, sodass

$$\sup_{i \in I} \int_{|f_i|>c_\epsilon} |f_i| d\mu < \epsilon \text{ und } \sup_{i \in I} \int_{A_\epsilon^c} |f_i| d\mu < \epsilon$$

**Satz 41** (13.LP, Charakterisierungen der gmI). Für eine Familie  $\mathcal{F} = \{f_i, i \in I\}$  messbarer Funktionen auf  $(\Omega, \mathcal{A}, \mu)$  sind folgende Bedingungen äquivalent:

1.  $\mathcal{F}$  ist gleichmäßig integrierbar
2.  $\forall \epsilon > 0 \exists g_\epsilon \geq 0, g_\epsilon \in L_1$  mit  $\sup_{i \in I} \int_{|f_i| > g_\epsilon} |f_i| d\mu < \epsilon$
3.  $\forall \epsilon > 0 \exists g_\epsilon \geq 0, g_\epsilon \in L_1$  mit  $\sup_{i \in I} \int (|f_i| - g_\epsilon)^+ d\mu < \epsilon$
4.  $\sup \int |f_i| d\mu < \infty$  und  $\forall \epsilon > 0 \exists g_\epsilon \geq 0, g_\epsilon \in L_1$  und  $\delta > 0$  mit  $\int_A g_\epsilon d\mu < \delta \implies \sup_{i \in I} \int_A |f_i| d\mu < \epsilon$

*Proof.* i)  $\implies$  ii): Nach Voraussetzung existiert  $c_\epsilon$  und  $A_\epsilon^c$  wie in der Definition, so ist  $g_\epsilon := c_\epsilon 1_{A_\epsilon}$  ein Kandidat für ii)

$$\begin{aligned} \int g_\epsilon d\mu &= c_\epsilon \mu(A_\epsilon) < \infty \implies g_\epsilon \in \mathcal{L}_1 \\ \int_{[|f_i| > g_\epsilon]} |f_i| d\mu &= \int_{\emptyset \cap A_\epsilon} |f_i| + \int_{\emptyset \cap A_\epsilon^c} |f_i| \leq \underbrace{\int_{[|f_i| > c_\epsilon] \cap A_\epsilon} |f_i|}_{[|f_i| > c_\epsilon 1_{A_\epsilon}]} + \int_{A_\epsilon^c} |f_i| \leq \int_{|f_i| > c_\epsilon} |f_i| + \epsilon \leq 2\epsilon \end{aligned}$$

ii)  $\implies$  iii):  $|f_i| 1_{[|f_i| > g_\epsilon]} \geq (|f_i| - g_\epsilon)^+$  also gilt auch

$$\int_{|f_i| > g_\epsilon} |f_i| d\mu \geq \int (|f_i| - g_\epsilon)^+ d\mu$$

und iii) gilt, nach Anwendung von  $\sup_{i \in I}$ .

iii)  $\implies$  iv) Es gelte  $\sup_i \int (|f_i| - g_\epsilon)^+ < \epsilon$ .

$$\begin{aligned} \int |f_i| &= \int_{|f_i| > g_\epsilon} |f_i| - g_\epsilon + \int_{|f_i| > g_\epsilon} g_\epsilon + \int_{|f_i| \leq g_\epsilon} |f_i| \leq \int (|f_i| - g_\epsilon)^+ + \int_{|f_i| > g_\epsilon} g_\epsilon + \int_{|f_i| \leq g_\epsilon} g_\epsilon \leq \\ &\epsilon + \int g_\epsilon < \infty \text{ da } g_\epsilon \in \mathcal{L}_1 \end{aligned}$$

$\implies c := \sup_i \int |f_i| < \infty$ , somit gilt a).

Für b) sei  $A \in \mathcal{A}$  mit  $\int_A g_\epsilon < \delta = \frac{\epsilon}{2}$ .

$$\begin{aligned} \int_A |f_i| &= \int_{A \cap [|f_i| > g_\epsilon]} (|f_i| - g_\epsilon) + \int_{A \cap [|f_i| > g_\epsilon]} g_\epsilon + \int_{A \cap [|f_i| \leq g_\epsilon]} |f_i| \leq \\ &\int (|f_i| - g_\epsilon)^+ + \int_{A \cap [|f_i| > g_\epsilon]} g_\epsilon + \int_{A \cap [|f_i| \leq g_\epsilon]} g_\epsilon \leq \delta + \int_A g_\epsilon \leq \epsilon + \delta \end{aligned}$$

daher ist auch  $\sup_i \int_A |f_i| < 2\epsilon$

iv)  $\implies$  i) Sei  $C = \sup \int |f_i| < \infty$  und  $a > \frac{C}{\delta}$  dann gilt

$$\begin{aligned} \int_{|f_i| > ag_\epsilon} g_\epsilon &\leq \int_{|f_i| > ag_\epsilon} |f_i| \leq \int |f_i| \leq C \\ \implies \int_{|f_i| > ag_\epsilon} g_\epsilon &\leq \frac{C}{a} \leq \delta \end{aligned}$$

und nach Voraussetzung für iv) folgt daraus  $\sup_i \int_{|f_i| > ag_\epsilon} |f_i| \leq \epsilon$ .

Mit  $g_\epsilon \in \mathcal{L}_1$  ist auch  $h := ag_\epsilon \in \mathcal{L}_1$ . Damit wurde ein  $A_\epsilon := [|f_i| \leq ag_\epsilon]$  mit endlichem Maß  $\mu(A_\epsilon) < \infty$  gefunden, sodass  $\sup_i \int_{A_\epsilon^c} |f_i| < \epsilon$ .

Für den Nachweis einer Konstanten  $c_\epsilon$  dient das vorige Lemma:

Wenn  $h = a \cdot g_\epsilon \in \mathcal{L}_1$ , dann existiert ein  $c_\epsilon$  und ein  $A_\epsilon$  sodass  $\int_{|h| > c_\epsilon} |h| < \epsilon$  und  $\int_{A_\epsilon^c} |h| < \epsilon$ .

Dann ist

$$\int_{|f_i| > c_\epsilon} |f_i| = \int_{[h \geq |f_i| > c_\epsilon]} |f_i| + \underbrace{\int_{[h < |f_i|] \cap [|f_i| > c_\epsilon]} |f_i|}_{\max(h, c_\epsilon) < |f_i| \implies |f_i| > h} \leq \int_{[h \geq |f_i| > c_\epsilon]} h + \int_{|f_i| > h} |f_i| \leq \int_{[h > c_\epsilon]} h + \epsilon \leq 2\epsilon.$$

damit auch  $\sup_i \int_{|f_i| > c_\epsilon} |f_i| < \epsilon$ , die erste Bedingung für gmI.

Auch die zweite Bedingung für gmI folgt aus dem Lemma für Integrierbarkeit.

$$\int_{A_\epsilon^c} |f_i| = \int_{[h \geq |f_i|] \cap A_\epsilon^c} |f_i| + \int_{[|f_i| > h] \cap A_\epsilon^c} |f_i| \leq \int_{[h \geq |f_i|] \cap A_\epsilon^c} |f_i| + \int_{|f_i| > h} |f_i| \leq \int_{A_\epsilon^c} h + \epsilon < 2\epsilon$$

nach dem Lemma bzw wegen der Voraussetzung für iv).

Daher gilt auch  $\sup_i \int_{A_\epsilon^c} |f_i| < \epsilon \forall \epsilon > 0$  und auch die zweite Bedingung für gmI ist erfüllt.  $\square$

Es ergeben sich unmittelbar viele Situationen (Familien), wo gmI gilt.

- Ist die Familie  $\mathcal{F}$  durch  $g \in \mathcal{L}_1$  dominiert, d.h.  $|f_i| \leq g \forall i$ , dann ist  $\mathcal{F}$  gmI, wegen 3.

$$0 = \sup \int (|f_i| - g)^+ < \epsilon$$

- Eine endliche Familie integrierbarer Funktionen  $\mathcal{F}$  ist immer gmI.
- Wird die Familie  $\mathcal{F}$  durch die Familie  $\mathcal{G}$  dominiert, also zu jedem  $f \in \mathcal{F}$  existiert ein  $g \in \mathcal{G}$  mit  $|f| \leq |g|$ , dann ist auch  $\mathcal{F}$  gmI, vorausgesetzt  $\mathcal{G}$  ist gmI.
- Die Summe  $\mathcal{F} + \mathcal{G} = \{f + g | f \in \mathcal{F}, g \in \mathcal{G}\}$  ist auch gmI, wenn  $\mathcal{F}$  und  $\mathcal{G}$  gmI sind.
- Die Integrale sind gleichmäßig über  $\mathcal{F}$  (gmI) beschränkt, da  $C = \sup_{f \in \mathcal{F}} \int |f| d\mu < \infty$  eine Folgerung aus gmI ist.

Damit ist etwa  $\mathcal{F} := \{f(x) = \frac{1}{x^{1+\epsilon}} 1_{(1,\infty)(x)} | \epsilon > 0\}$  auf  $(\mathbb{R}, \mathcal{B}, \lambda)$  keine gmI Familie, obwohl jedes  $f \in \mathcal{F}$  integrierbar ist und  $\sup_{f \in \mathcal{F}} \|f\|_\infty = 1 < \infty$ .

Das wäre anders, wenn statt  $\lambda$  ein endliches Maß  $\mu$  vorliegen würde.

Ist  $\mu$  endlich und  $\sup_{f \in \mathcal{F}} \|f\|_\infty < K$  ist wegen 3. sofort auch  $\mathcal{F}$  gmI mit  $g_\epsilon = K$ .

- Jede Familie  $\mathcal{F}$ , die gmI ist wegen  $\sup_{f \in \mathcal{F}} \int |f| < \infty$  Teilmenge von  $\mathcal{L}_1$ ,  $\mathcal{F} \subseteq \mathcal{L}_1$  für  $(\Omega, \mathcal{A}, \mu)$ .

Für endliche Maße (also alle W-Maße) wird der Nachweis der gmI wesentlich einfacher.

**Satz 42** (14.LP). *Die Familie  $\mathcal{F}$  messbarer Funktionen auf dem endlichen Maßraum  $(\Omega, \mathcal{A}, \mu)$  ist genau dann gmI, wenn*

$$C = \sup_{i \in I} \int |f_i| < \infty \text{ oder } \lim_{c \rightarrow \infty} \sup_i \mu(|f_i| \geq c) = 0$$

$$\text{und } \forall \epsilon > 0 \exists \delta > 0 \text{ mit } \mu(A) < \delta \implies \sup_i \int_A |f_i| < \epsilon$$

*Proof.*  $\implies$  :  $C = \sup \int |f_i| < \infty$  folgt aus gmI, dann gilt auch  $\mu(|f_i| > c) \leq \frac{1}{c} \int |f_i| \leq \frac{C}{c} \rightarrow 0$  für  $c \rightarrow \infty$  somit  $\lim_{c \rightarrow \infty} \sup_i \mu(|f_i| \geq c) = 0$ .

Das  $\epsilon$ - $\delta$ -Kriterium b) gilt ebenfalls, zu  $\epsilon > 0$  existiert wegen gmI ein  $c_\epsilon > 0$  mit  $\sup_i \int_{|f_i| > c_\epsilon} |f_i| < \frac{\epsilon}{2}$  und daher gilt wenn  $\mu(A) \leq \delta := \frac{\epsilon}{2c_\epsilon}$ :

$$\int_A |f_i| = \int_{A \cap [|f_i| \leq c_\epsilon]} |f_i| + \int_{A \cap [|f_i| > c_\epsilon]} |f_i| \leq c_\epsilon \mu(A) + \frac{\epsilon}{2} < \epsilon$$

$\Leftarrow$  : Es gelte  $\sup_i \mu(|f_i| \geq c) \rightarrow 0$ , es existiert ein  $c_\delta$  mit  $\mu(|f_i| \geq c_\delta) < \delta$ , da auch das  $\epsilon$ - $\delta$ -Kriterium gilt, folgt für  $A_{\epsilon,i} := [|f_i| > c_\delta]$

$$\sup_j \int_{A_{\epsilon,i}} |f_j| d\mu < \epsilon$$

daher auch  $\sup_i \sup_j \int_{A_{\epsilon,i}} |f_j| d\mu < \epsilon$  und  $\int_{|f_i| > c_\delta} |f_i| d\mu < \epsilon$ , die zweite Bedingung für gmI.  $\square$

Wenn die Funktionen  $f_i$  als Dichten verstanden werden von  $\nu_i(A) := \int_A |f_i| d\mu$ .

Diese (endliche) Maße  $\nu_i$  sind daher gleichmäßig absolut stetig bezüglich dem endlichen Maß  $\mu$ .

Für den Nachweis der gml einer Familie  $\mathcal{F}$  messbarer Funktionen auf dem endlichen Maßraum  $(\Omega, \mathcal{A}, \mu)$  genügt es, eine (stärker als  $x \rightarrow x$  wachsende) Funktion zu finden, die gleichmäßig über  $\mathcal{F}$  integrierbar ist. Dieses Kriterium ist oft einfacher zu prüfen.

**Lemma 12.**  $\mathcal{F} \subseteq \mathcal{L}_1$  auf dem endlichen Maßraum  $(\Omega, \mathcal{A}, \mu)$  ist genau dann gml, wenn eine Funktion  $H : [0, \infty) \rightarrow [0, \infty)$  mit

$$\lim_{x \rightarrow \infty} \frac{H(x)}{x} = \infty \text{ und } \sup_{f \in \mathcal{F}} \int H(|f|) d\mu < \infty$$

existiert.

Die Funktion  $H$  ist als monoton wachsend und konvex wählbar.

Für SG  $X$  bzw. eine Familie SGn  $\mathcal{F}$  reduziert sich die Bedingung für gml auf

$$\sup_{X \in \mathcal{F}} \mathbb{E}[|X| 1_{|X| > K}] < \epsilon$$

wobei zu  $\epsilon > 0$  ein  $K \geq 0$  mit obiger Bedingung existiert.

**Lemma 13.** Folge  $X_n$  SGn mit  $\mathbb{E}|X_n| < \infty$  ist gml, genau wenn

$$\lim_{x \rightarrow \infty} (\sup_{n \geq 1} \mathbb{E}[|X_n| 1_{|X_n| > x}]) = 0$$

*Proof.*  $\Rightarrow$  :  $X_n$  sei gml, da  $\sup_n \mathbb{E}|X_n| < \infty$  gilt mit der Markoff-Ungleichung für jedes  $x > 0$

$$xP(|X_n| > x) \leq \mathbb{E}|X_n| \forall n$$

Für  $\delta > 0$  und ausreichend großes  $x$  gilt

$$\sup_n P(\underbrace{|X_n| > x}_{A_{\epsilon, n}}) \leq \frac{1}{x} \sup_{n \geq 1} \mathbb{E}|X_n| < \delta$$

gilt  $\mathbb{E}|X_n| 1_{A_{\epsilon, n}} \leq \epsilon$  und daher auch

$$\limsup_{x \rightarrow \infty} \sup_n \mathbb{E}[|X_n| 1_{|X_n| > x}] < \epsilon.$$

$\Leftarrow$  : Es gelte  $\lim_{x \rightarrow \infty} (\sup_n \mathbb{E}[|X_n| 1_{|X_n| > x}]) = 0$ .

Für  $x$  groß genug gilt

$$\sup_n \mathbb{E}|X_n| \leq \sup_n \{\mathbb{E}|X_n| 1_{|X_n| > x} + x\} < \infty$$

Für die zweite Bedingung der gml wird für  $\epsilon > 0$  ein  $x$  gewählt, sodass  $\sup_n \mathbb{E}[|X_n| 1_{|X_n| > x}] < \frac{\epsilon}{2}$ .

Für  $\delta < \frac{\epsilon}{2x}$  und  $A \in \mathcal{A}$  mit  $P(A) < \delta$  folgt

$$\mathbb{E}|X_n| 1_A = \mathbb{E}|X_n| 1_{A \cap \{|X_n| > x\}} + \mathbb{E}|X_n| 1_{A \cap \{|X_n| \leq x\}} \leq \underbrace{\mathbb{E}|X_n| 1_{|X_n| > x}}_{< \frac{\epsilon}{2}} + \underbrace{\mathbb{E}x 1_{A \cap \{|X_n| \leq x\}}}_{\leq 1_A} \leq \epsilon.$$

( $\delta x < \frac{\epsilon}{2}$ )

□

**Beispiel.** Die Familie SGn  $\mathcal{F} = \{X_i | i \in I\}$  enthalte quadratisch integrierbare Funktionen,  $\mathcal{F} \subseteq \mathcal{L}_2$  auf  $(\Omega, \mathcal{A}, P)$ . Wenn die 2. Momente beschränkt sind, etwa  $\sup_i \text{Var}(X_i) < \infty$  dann ist  $\mathcal{F}$  gml. Aus  $(\mathbb{E}|X_i|)^2 \leq \mathbb{E}X_i^2$  (wegen der Jensen Ungl.) ist auch  $\sup \mathbb{E}|X_i| < \infty$ .

Weiters gilt

$$\sup_i \int |X_i|^2 d\mu = \sup_i ((\mathbb{E}|X_i|)^2 + \text{Var}(X_i)) < \infty$$

Mit  $H(x) = x^2$  sind die Bedingungen der letzten Proposition erfüllt. Eigentlich würde die Aussage auch für  $H(x) = x^{1+a} \forall a > 0$  gelten.



Wie es aus den vorigen Aussagen vermuten lässt, ist die Äquivalenz von  $L_p$ -Konvergenz und Konvergenz im Maß, wenn die Familie (Folge) auch gml ist.

Als Vorbereitung dienen die folgenden Hilfssätze.

**Lemma 14.**  $f_n$  ist eine gml-Folge auf  $(\Omega, \mathcal{A}, \mu)$  dann gilt das Lemma von Fatou in der Form

$$\int \liminf_n f_n \leq \liminf_n \int f_n \leq \limsup_n \int f_n \leq \int \limsup_n f_n$$

*Proof.* Für die Anwendung des Lemmas von Fatou müssen integrierbare Funktionen mit  $f_n \geq g_1$  bzw.  $f_n \leq g_2$  existieren.

Da die  $f_n$  gml sind, existiert zu  $\epsilon > 0$  ein  $g \geq 0$ ,  $g \in \mathcal{L}_1$  und  $\sup_n \int_{|f_n| > g} |f_n| d\mu < \epsilon$  und

$$\begin{aligned} \int f_n &= \int_{f_n > -g} f_n + \int_{f_n \leq -g} f_n \geq \int_{f_n > -g} f_n - \epsilon \\ \text{Da } \int_{f_n \leq -g} f_n &= - \int_{f_n \leq -g} |f_n| \geq - \int_{|f_n| > g} |f_n| \geq -\epsilon \end{aligned}$$

Betrachte  $\tilde{f}_n := f_n 1_{[f_n > -g]} \geq -g$ .  $\tilde{f}_n$  erfüllt die Voraussetzungen des Lemmas von Fatou.

$$\int \liminf_n \tilde{f}_n \leq \liminf_n \int \tilde{f}_n = \liminf_n \underbrace{\int f_n 1_{f_n > -g}}_{\leq \sup \int |f_n| < \infty}$$

Da  $f_n \leq \tilde{f}_n \implies \liminf f_n \leq \liminf \tilde{f}_n$  somit  $\int \liminf f_n \leq \int \liminf \tilde{f}_n \leq \liminf \int \tilde{f}_n$ .

Nach oben gilt  $\int \tilde{f}_n \leq \int f_n + \epsilon$  also ist der letzte Term  $\leq \liminf \int f_n + \epsilon$ .

Die Ungleichung für  $\limsup$  wird analog wegen  $\limsup f_n = -\liminf(-f_n)$  geführt.  $\square$

**Lemma 15.**  $f_n$  sei gml Folge. Aus der Konvergenz im Maß  $f_n \xrightarrow{\mu} f$  folgt  $f \in \mathcal{L}_1$ ,  $\lim \int f_n = \int f$  und  $\lim_n \|f_n - f\|_1 = 0$ .

Die Folgerungen treffen auch für  $f_n \rightarrow f \mu$ -f.ü. zu.

*Proof.* Es gelte  $f_n \rightarrow f \mu$ -f.ü. und daher auch  $|f_n| \rightarrow |f|$ . Nach dem vorigen Lemma (von Fatou) ist

$$\begin{aligned} \int \lim |f_n| &\leq \underbrace{\liminf \int |f_n|}_{=\lim \int |f_n| \leq C = \sup \int |f_n|} \leq \limsup \int |f_n| \leq \int \lim |f_n| \end{aligned}$$

Also ist  $f \in \mathcal{L}_1$ . Da  $\|f_n\|_1 \rightarrow \|f\|_1$  und  $f_n \rightarrow f \mu$ -f.ü. gilt nach Satz 10.LP auch  $\|f_n - f\|_1 \rightarrow 0$ .

Gilt  $f_n \rightarrow f$  im Maß, dann existiert wieder eine Teilfolge  $f_{n,k}, f_{n,k} \rightarrow f \mu$ -f.ü. für diese Teilfolge gilt  $\|f_{n,k} - f\|_1 \rightarrow 0$  nach vorigem Argument.

Angenommen  $\|f_n - f\|_1 \not\rightarrow 0$  müsste es eine andere Teilfolge geben, die  $\|f_{m,j} - f\|_1 > \epsilon$  erfüllt für alle  $j \geq 1$ .  $f_{m,j} \rightarrow f$  im Maß gilt trotzdem. So müsste  $f_{m,j}$  wieder eine Teilfolge  $f_{m,j,l}$  besitzen, die wiederum  $\|f_{m,j,l} - f\|_1 \rightarrow 0$  erfüllt. Das ist ein Widerspruch.  $\square$

Die letzten Aussagen über gml Folgen werden zum wesentlichen Kriterium für  $L_p$ -Konvergenz zusammengefügt.

**Satz 43** (15.LP, Vitali-Kriterium). Die Folge  $f_n \in \mathcal{L}_p, 1 \leq p < \infty$  konvergiert in  $L_p$   $\|f_n - f\|_p \rightarrow 0 \iff f_n \xrightarrow{\mu} f$  und  $|f_n|^p$  ist gml.

*Proof.*  $\implies$  : Aus  $\|f_n - f\|_p \rightarrow 0$  folgt nach Satz 11.LP  $f_n \xrightarrow{\mu} f$  und  $\|f_n\|_p \rightarrow \|f\|_p$  mit  $f \in \mathcal{L}_p$ .

Daher ist  $C = \sup \int |f_n|^p < \infty$ .

Für die zweite Bedingung der gml sei  $\epsilon > 0$  und  $\delta := \frac{\epsilon^{1/p}}{2}$  sowie  $g := |f|^p + \sup_{i=1}^{n_\epsilon} |f_i|^p$  wobei  $n_\epsilon$  so bestimmt wird, dass  $\underbrace{\|f_n - f\|_p}_{\rightarrow \int |f_n - f|^p < \epsilon/2} < \delta \forall n \geq n_\epsilon$ .

Damit gilt für ein  $A \in \mathcal{A}$

$$\int_A g d\mu < \delta \implies \int_A |f_n|^p d\mu < \epsilon$$

für  $n \leq n_\epsilon$  nach Definition von  $g$ .

Wenn  $n > n_\epsilon$  betrachte

$$\|f_n 1_A\|_p \leq \underbrace{\|(f_n - f) 1_A\|_p}_{< \frac{\epsilon}{2}} + \underbrace{\|f 1_A\|_p}_{< \frac{\epsilon}{2} \text{ wenn } \int_A g < \delta}$$

Auch in diesem Fall folgt aus  $\int_A g d\mu < \delta \implies \int_A |f_n|^p d\mu < \epsilon$

$\Leftarrow$  :  $|f_n|^p$  sei gmI und Maß-konvergent. Es gibt wieder eine Teilfolge  $f_{n,k}$  mit  $f_{n,k} \rightarrow f$   $\mu$ -f.ü.  
 $\implies |f_{n,k}|^p \rightarrow |f|^p$   $\mu$ -f.ü.  $|f_{n,k}|^p$  ist gmI, nach vorigem Lemma muss  $\|f_{n,k}\|_p \rightarrow \|f\|_p$  und  $\|f\|_p < \infty$ .

Nach Satz 10.LP ist dann auch  $\|f_{n,k} - f\|_p \rightarrow 0$ . Würde  $\|f_n - f\|_p \not\rightarrow 0$  müsste eine Teilfolge  $f_{m,j}$  mit  $\|f_{m,j} - f\|_p > \epsilon$  die aber wieder  $f_{m,j} \rightarrow f$  im Maß erfüllt. Davon müsste eine nächste Teilfolge  $f_{m,j,k}$  mit  $\|f_{m,j,k} - f\|_p \rightarrow 0$  geben, was ein Widerspruch ist.  $\square$

Die Sätze 11.LP und 15.LP zusammengefasst

$$\|f_n - f\|_p \rightarrow 0 \iff f_n \xrightarrow{\mu} f \wedge f \in L_p \wedge \|f_n\|_p \rightarrow \|f\|_p \iff f_n \xrightarrow{\mu} f \wedge |f_n|^p \text{ ist gmI}$$

kann als Kriterium für gmI verwendet werden:

Für eine im Maß konvergente Folge  $f_n \xrightarrow{\mu} f$  sind  $|f_n|^p$  gmI  $\iff f \in L_p \wedge \|f_n\|_p \rightarrow \|f\|_p, \|f\|_p < \infty$ .

Wie die  $L_p$ -Konvergenz von einer Metrik (sogar einer Norm), lässt sich auch die Konvergenz im Maß (für sigma-endliches Maß) aus einer Metrik erzeugen.

**Lemma 16.**  $\mu$  sei ein sigma-endliches Maß mit  $A_n \nearrow \Omega, \mu(A_n) < \infty$ . Zu messbaren Funktionen  $f, g : \Omega \rightarrow \mathbb{R}$  sei

$$d(f, g) = \sum_{n=1}^{\infty} \frac{2^{-n}}{1 + \mu(A_n)} \int_{A_n} \min(1, |f - g|) d\mu$$

Dann ist  $d$  eine Metrik, sodass für die Folge  $f_n, f$  gilt  $f_n \xrightarrow{\mu} f \iff d(f_n, f) \rightarrow 0$ .

Eine weitere Bedingung für f.ü. Konvergenz aus  $L_p$ -Konvergenz abzuleiten ist die schnelle Konvergenz.

**Satz 44** (15.LP, Schnelle Konvergenz).  $f_n$  und  $f$  aus  $L_p$  für ein  $1 \leq p < \infty$ . Dann folgt aus

1.  $\sum_{n \geq 1} \|f_n - f\|_p < \infty$  oder

2. Für  $A \in \mathcal{A}$  mit  $\mu(A) < \infty$  und jedes  $\epsilon > 0$  gilt  $\sum_{n \geq 1} \mu(A \cap [|f_n - f| > \epsilon]) < \infty$

dass  $f_n \rightarrow f$   $\mu$ -f.ü.

*Proof.* Aus 1. folgt wegen

$$\mu(A \cap [|f_n - f| > \epsilon]) \leq \mu(|f_n - f| > \epsilon) \leq \frac{1}{\epsilon^p} \|f_n - f\|_p^p$$

sofort 2. Für 2. wird ein messbares  $A \in \mathcal{A}$  mit  $\mu(A) < \infty$  angenommen, o.B.d.A sei  $A = \Omega$ . Mit  $D_n := [|f_n - f| > \epsilon]$  und  $D_\epsilon := \limsup_n D_n \rightarrow \infty$  folgt nach dem Lemma von Borel-Cantelli

$$\mu(D_\epsilon) = \mu\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} [|f_k - f| > \epsilon]\right) = 0$$

Sei  $N = \bigcup_{m=1}^{\infty} D_{1/m}$ , daher  $\mu(N) = 0$  und  $f_n \rightarrow f$  auf  $N^c$ .  $\square$

## 4.4 Konvergenzarten schematisch

TODO LP Seite 54

## 5 Bedingte Erwartung und Verteilung

Bedingte Wahrscheinlichkeit soll von einem Ereignis  $A$  mit  $P(A) > 0$  verallgemeinert werden. Für festes  $A \in \mathcal{A}$  gilt  $P(M|A) := \frac{P(M \cap A)}{P(A)}$ .

**Def** (elementarer bedingter Erwartungswert). Diese Verteilung führt für eine messbare Funktion  $X \in \mathcal{L}^1(P)$  auf einen elementaren bedingten Erwartungswert  $\mathbb{E}[X|A] = \int X(\omega) dP(\omega|A) = \frac{1}{P(A)} \mathbb{E}(1_A X)$

Wohldefiniertheit elementarer bedingter Erwartungswert. TODO □

Zusammensetzung eines Erwartungswerts auf disjunkte Teilmengen  $B_i$  mit  $\bigcup_i B_i = \Omega$ :  $\mathbb{E}X = \mathbb{E}(X|B_1)P(B_1) + \mathbb{E}(X|B_2)P(B_2) + \dots$ .

Bedingte Wahrscheinlichkeit  $P(A|B) = \mathbb{E}(1_A|B)$ .

**Def** (bedingte Erwartung unter  $\mathcal{F}$ ).  $(B_i)$  sei eine Zerlegung (höchstens abzählbar). Es sei  $\mathcal{F} := \sigma((B_i)_i)$  von der Zerlegung erzeugt (Wenn es eine endliche Zerlegung ist, dann besteht  $\mathcal{F}$  nur aus allen Vereinigungen von  $B_i$ ).

Definiere als bedingte Erwartung unter  $\mathcal{F}$  eine messbare Funktion  $Y = \mathbb{E}[X|\mathcal{F}]$  mit  $Y : \Omega \rightarrow \mathbb{R}$ ,  $Y(\omega) = \mathbb{E}[X|B_i](\omega) \iff \omega \in B_i$  also  $Y = \sum_{i=1}^{\infty} \mathbb{E}[X|B_i]1_{B_i}(\omega)$ .

**Lemma 17.** •  $Y$  ist  $\mathcal{F}$  messbar

•  $Y$  ist  $\mathcal{L}^1(P)$ , hat endliche Erwartung und für jedes  $A \in \mathcal{F}$  gilt  $\int_A Y dP = \int_A X dP$ .

*Proof.* TODO □

**Lemma 18** (vollständige Wahrscheinlichkeit). Für bedingte Wahrscheinlichkeit, also  $X = 1_B$  mit  $B \in \mathcal{A}$  ergibt sich der Satz der vollständigen Wahrscheinlichkeit

$$P(B) = \sum_{j \in J'} \underbrace{P(B|B_j)P(B_j)}_{P(B \cap B_j)}$$

Diesen bedingten Erwartungswert kann man als Erwartung einer bedingten Verteilung auffassen.

**Beispiel.** TODO

Eine Verallgemeinerung (motiviert durch den obigen diskreten Zugang) sollte auch Nullmengen für  $B_i$  vorsehen, etwa von stetigen SG erzeugt.

**Definition 18** (bedingte Erwartung).  $(\Omega, \mathcal{A}, P) \dots$  Wahrscheinlichkeitsraum,  $\mathcal{F} \subseteq \mathcal{A}$ ,  $\mathcal{F}$  beliebige Teil Sigmaalgebra. Die SG heißt bedingte Erwartung  $X_0 = \mathbb{E}[X|\mathcal{F}]$  von  $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$  wenn

1.  $X_0$   $\mathcal{F}$  messbar ist

2.  $\forall A \in \mathcal{F}$  gilt  $\mathbb{E}[X1_A] = \int_A X dP = \int_A X_0 dP = \mathbb{E}[X_01_A]$

Die bedingte Wahrscheinlichkeit ist  $P(A|\mathcal{F}) = \mathbb{E}[1_A|\mathcal{F}]$ ,  $A \in \mathcal{A}$ .

Wenn  $\mathcal{F} = \sigma(Y)$  von einer SG erzeugt wird, bezeichnet  $\mathbb{E}[X|Y] = \mathbb{E}[X|\mathcal{F}]$  bedingte Erwartung von  $X$  unter  $Y$ .

**Lemma 19** (Existenz und Eindeutigkeit).  $X_0 = \mathbb{E}[X|\mathcal{F}]$  existiert und ist eindeutig,  $[P|_{\mathcal{F}}]$  f.s.

*Proof.* TODO □

Die Aussage der Eindeutigkeit (wie alle Aussagen über  $\mathbb{E}(X|\mathcal{F})$ ) ist auf  $[P]$ -f.s. bzw.  $[P|_{\mathcal{F}}]$ -f.s zu betrachten, man bezeichnet daher  $\mathbb{E}[X|\mathcal{F}]$  als eine Version der bedingten Erwartung.

$\mathbb{E}[X|\mathcal{F}]$  besitzt die wichtigen Eigenschaften des Integrals, also des Erwartungswerts.

**Satz 45** (Eigenschaften bedingter Erwartungswert). Es sei  $\mathcal{G} \subset \mathcal{F} \subset \mathcal{A}$  Teil Sigma-Algebren,  $X, Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$  dann gilt:

1. Monotonie:  $X \geq Y \implies X_0 = \mathbb{E}[X|\mathcal{F}] \geq \mathbb{E}[Y|\mathcal{F}] = Y_0$
2. Linearität:  $\mathbb{E}[aX + Y|\mathcal{F}] = a\mathbb{E}[X|\mathcal{F}] + \mathbb{E}[Y|\mathcal{F}]$
3. Dreiecksungleichung:  $\mathbb{E}[|X||\mathcal{F}] \geq |\mathbb{E}[X|\mathcal{F}]|$
4. Totale Erwartung:  $\mathbb{E}[X_0] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}X$
5. Turmeigenschaft:  $\mathcal{G} \subset \mathcal{F} \subset \mathcal{A}$ ,  $\mathcal{G}$  ist gröber als  $\mathcal{F}$ ,  $\implies \mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{G}]$
6. Faktorisierung  $X \cdot Y \in \mathcal{L}^1(P)$ ,  $Y$  sei messbar bez.  $\mathcal{F}$ ,  $X$  messbar bez.  $\mathcal{F} \implies \mathbb{E}[X|\mathcal{F}] = X$ .  
 $\mathbb{E}[XY|\mathcal{F}] = Y \cdot \mathbb{E}[X|\mathcal{F}]$ .
7. Monotone Konvergenz:  $X_n \nearrow X$ ,  $X_n \in \mathcal{L}^1 \implies \mathbb{E}[X_n|\mathcal{F}] \nearrow \mathbb{E}[X|\mathcal{F}]$
8. Majorisierte Konvergenz:  $X_n$  Folge mit  $|X_n| \leq Y$  und  $X_n \rightarrow X$  f.s. und  $Y \in \mathcal{L}^1$  integrierbar, dann gilt  $\lim_{n \rightarrow \infty} \mathbb{E}[X_n|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}]$
9. Unabhängigkeit:  $X$  sei unabhängig von  $\mathcal{F}$ ,  $\sigma(X)$  u.a.  $\mathcal{F} \implies \mathbb{E}[X|\mathcal{F}] = \text{const.} = \mathbb{E}[X]$
10. Trivialitätseigenschaft: Wenn  $\mathcal{F}$  aus  $P$ -trivialen Ereignissen besteht, also für  $F \in \mathcal{F}$ , ist  $P(F) = 0$  oder 1, dann gilt  $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}X = \text{const.}$

Proof. TODO □

**Lemma 20** (Lemma von Fatou).  $X_n$  Folge SGN auf  $(\Omega, \mathcal{A}, P)$ . Es gibt ein messbares  $U \leq X_n$  und  $\mathbb{E}U^- < \infty$ . Für  $\mathcal{A}' \subseteq \mathcal{A}$ , Teil Sigma-Algebra gilt

$$\mathbb{E}[\liminf_n X_n|\mathcal{A}'] \leq \liminf_n \mathbb{E}[X_n|\mathcal{A}'] P\text{-f.s.}$$

Entsprechend gilt für  $X_n \leq U'$ ,  $\mathbb{E}U'^+ < \infty$

$$\limsup_n \mathbb{E}[X_n|\mathcal{A}'] \leq \mathbb{E}[\limsup_n X_n|\mathcal{A}'] P\text{-f.s.}$$

Proof. TODO □

**Satz 46** (Dominierte Konvergenz für bedingte Erwartung).  $X_n$  ist  $P$ -f.s. konvergente Folge auf  $(\Omega, \mathcal{A}, P)$  mit  $|X_n| \leq Y \forall n$  und  $Y$  ist integrierbar. Für die Teilsigma-Algebra  $\mathcal{A}' \subseteq \mathcal{A}$  gilt  $P$ -f.s.:  $\mathbb{E}(\lim X_n|\mathcal{A}') = \lim_n \mathbb{E}(X_n|\mathcal{A}')$

Proof. TODO □

Wie der Grenzwert von  $X_n$  im Satz der dominierten Konvergenz (ohne bedingte Erwartung) ist auch hier  $\mathbb{E}[X|\mathcal{A}'] < \infty P\text{-f.s.}$

**Satz 47** (1.BE). Es sei  $X \cdot Y \in \mathcal{L}^1$ ,  $Y$  messbar bezgl.  $\mathcal{F}$  dann gilt  $\mathbb{E}[X \cdot Y|\mathcal{F}] = Y\mathbb{E}[X|\mathcal{F}]$ .

**Beberkung.** Diese Aussage gilt auch unter geringeren Voraussetzungen als  $\mathbb{E}|XY| < \infty$ , etwa  $X, Y \geq 0$ , mit allerdings spezielleren Bedingungen.

Proof. TODO □

BE Seite 12

**Beberkung.** Die Unabhängigkeitseigenschaft zweier SGN  $X, Y$  überträgt sich nicht (unkorreliert auch nicht):

$X$  u.a.  $Y$  bez.  $P$ , dann gilt  $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$  aber nicht unbedingt  $\mathbb{E}[XY|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}]\mathbb{E}[Y|\mathcal{F}]$ . (Nur wenn  $\sigma(X), \sigma(Y)$  und  $\mathcal{F}$  unabhängig sind).

Unabhängige SGN von  $Y$  tragen keine Information i.d.S.  $\mathbb{E}[X|Y] = \mathbb{E}X$ ,  $X$  u.a.  $Y$ .

Wenn mehrere u.a. SGN von  $X$  existieren, müssen die aber auch gemeinsam unabhängig sein. Wenn  $Y$  oder  $Z$  u.a. von  $X$  sind, dann muss  $(Y, Z)$  nicht informationslos sein, also i.a.  $\mathbb{E}[X|Y, Z] \neq \mathbb{E}[X]$ .

**Satz 48** (2.BE).  $X$  SG ( $\mathbb{E}X$  existiert) auf  $(\Omega, \mathcal{A}, P)$ ,  $\mathcal{A}'$  ist Sub-Sigmaalgebra wie  $\mathcal{C}$  auch,  $\mathcal{A}', \mathcal{C} \subseteq \mathcal{A}$ . Wenn  $\sigma(\sigma(X) \cup \mathcal{A}')$  unabhängig von  $\mathcal{C}$  ist, gilt

$$\mathbb{E}[X|\sigma(\mathcal{C} \cup \mathcal{A}')] = \mathbb{E}[X|\mathcal{A}']$$

*Proof.* TODO □

Die Forderung für den bedingten Erwartungswert, dass für alle  $A \in \mathcal{A}', \mathcal{A}' \subseteq \mathcal{A}$

$$\int_A X dP = \int_A \mathbb{E}[X|\mathcal{A}'] dP$$

kann auf die Erzeugermengen  $A \in \mathcal{E}$  mit  $\mathcal{A}' = \sigma(\mathcal{E})$  reduziert werden, wenn  $\mathcal{E}$  ein Halbring mit  $\Omega \in \mathcal{E}$  ist. (Semi-Algebra)

**Satz 49** (3.BE).  $X$  sei integrierbar auf  $(\Omega, \mathcal{A}, P)$  und Teil-Sigmaalgebra  $\mathcal{A}' = \sigma(\mathcal{E})$ , wobei  $\mathcal{E}$  ein Halbring ist und  $\Omega \in \mathcal{E}$ . Wenn  $Y$   $\mathcal{A}'$ -messbar ist und  $\forall M \in \mathcal{E} : \int_M X dP = \int_M Y dP$ , dann gilt  $Y = \mathbb{E}[X|\mathcal{A}']$   $P$ -f.s.

*Proof.* TODO □

**Satz 50** (4.BE).  $X$  sei SG auf  $(\Omega, \mathcal{A}, P)$  und  $\mathcal{A}' \subseteq \mathcal{A}$  ist Teil-Sigmaalgebra.  $X_0$  sei  $\mathcal{A}'$  messbar.

1. Sei  $X \geq 0$ , dann ist  $X_0$  genau dann eine Version von  $\mathbb{E}[X|\mathcal{A}'] = X_0$   $P$ -f.s., wenn für alle  $\mathcal{A}'$ -messbaren  $Z \geq 0$ :  $\int Z X_0 dP = \int Z X dP$ .
2. Sei  $X$  integrierbar, dann ist  $X_0$  genau dann eine Version von  $\mathbb{E}[X|\mathcal{A}'] = X_0$  wenn für alle beschränkten und  $\mathcal{A}'$  messbaren  $Z$ :  $\int Z X_0 dP = \int Z X dP$ .

*Proof.* TODO □

Wenn im vorigen Satz  $X \in \mathcal{L}^p$  (statt  $\mathcal{L}^1$ ), dann gilt  $\mathbb{E}XZ = \mathbb{E}X \cdot Z$  für alle  $Z \in \mathcal{L}^q$  mit  $\frac{1}{p} + \frac{1}{q} = 1$ , (und  $\mathcal{A}'$ -messbar) also  $X_0 \in \mathcal{L}^p(\Omega, \mathcal{A}', P)$  erfüllt

$$X_0 = \mathbb{E}(X|\mathcal{A}') \iff \int X_0 Z dP = \int X Z dP$$

für alle  $Z \in \mathcal{L}^q(\Omega, \mathcal{A}', P)$ .

Beweis in Übung.

**Satz 51** (Jensen-Ungleichung für bedingte Erwartung).  $X$  ist integrierbare SG auf  $(\Omega, \mathcal{A}, P)$  mit  $X : \Omega \rightarrow (a, b)$ . Sei  $\phi : (a, b) \rightarrow \mathbb{R}$  konvex. Dann gilt für  $\mathcal{A}' \subseteq \mathcal{A}$ , Sub-Sigmaalgebra

$$\phi(\mathbb{E}(X|\mathcal{A}')) \leq \mathbb{E}(\phi(X)|\mathcal{A}') P\text{-f.s.}$$

*Proof.* TODO □

Hier wurde das Beweisprinzip der unbedingten Jensen-Ungleichung übernommen und für den bedingten Erwartungswert adaptiert. Genauso lassen sich weitere Ungleichungen auf bedingte Erwartungswerte anwenden.

**Satz 52** (Markoff-Ungleichung für bedingte Erwartung).  $\eta : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  und monoton wachsend. Für  $c > 0$  und  $\mathcal{A}' \subseteq \mathcal{A}$  (Sub-Sigmaalgebra) gilt  $P$ -f.s.

$$P(|X| > c|\mathcal{A}') \leq \frac{\mathbb{E}(\eta(|X|)|\mathcal{A}')}{\eta(c)}$$

*Proof.* TODO □

Eine sofortige Folgerung aus der bedingten Markoff-Ungleichung ist wieder die bedingte Tschebyscheff-Ungleichung.

**Lemma 21** (Tschebyscheff-Ungleichung).  $\phi : [0, \infty) \rightarrow [0, \infty)$  und monoton steigend,  $\phi(x) > 0$  für  $x > 0$ . Bei  $c > 0$  gilt

$$P[|X| > c|\mathcal{A}'] \leq \frac{\mathbb{E}[\phi(|X|)|\mathcal{A}']}{\phi(c)}$$

Da  $\phi$  monoton wächst, gilt  $[|X| > c] \geq [\phi(|X|) \geq \phi(c)]$  und die vorige Markoff-Ungleichung für  $\phi_0|X|$  das Resultat.

Die "klassische Tschebyscheff-Ungleichung" wird für  $\phi(x) = x^2$  auf  $[0, \infty)$  formuliert,

$$P[|X - \mathbb{E}X| > c|\mathcal{A}'] \leq \frac{\mathbb{V}(X|\mathcal{A}')}{c^2}.$$

Das lässt sich für andere Momente auch aufschreiben

$$P[|X - \mathbb{E}X| > c|\mathcal{A}'] \leq \frac{\mathbb{E}(|X - \mathbb{E}X|^k|\mathcal{A}')}{c^k}.$$

Aus der Jensen-Ungleichung für bedingte Erwartung erhält man auch für (bedingte)  $p$ -fache Integrale ( $p \geq 1$ ) und  $X \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$

$$|\mathbb{E}(X|\mathcal{A}')|^p \leq \mathbb{E}(|X|^p|\mathcal{A}') P\text{-f.s. und } \mathbb{E}(|\mathbb{E}(X|\mathcal{A}')|^p) \leq \mathbb{E}|X|^p$$

damit ist  $\mathbb{E}(X|\mathcal{A}')$  auch in  $\mathcal{L}^p(\Omega, \mathcal{A}', P)$ .

Auch für weitere für  $\mathcal{L}^p$ -Räume essentielle Ungleichungen werden die Beweisprinzipien auf die bedingte Erwartung übertragen.

**Lemma 22** (Hölder-Ungleichung für bedingte Erwartung). *Es seien  $X$  und  $Y$  SGn auf  $(\Omega, \mathcal{A}, P)$  mit der Sub-Sigmaalgebra  $\mathcal{A}' \subseteq \mathcal{A}$ . Dann gilt die Hölder-Ungleichung für bedingte Erwartung:  $\mathbb{E}|X|^p < \infty$ ,  $\mathbb{E}|Y|^q < \infty$ ,  $1 \leq p, q \leq \infty$  und  $\frac{1}{p} + \frac{1}{q} = 1$ , dann gilt  $P$ -f.s.*

$$|\mathbb{E}(XY|\mathcal{A}')| \leq \mathbb{E}(|XY||\mathcal{A}') \leq (\mathbb{E}(|X|^p|\mathcal{A}'))^{\frac{1}{p}} (\mathbb{E}(|Y|^q|\mathcal{A}'))^{\frac{1}{q}} < \infty$$

**Lemma 23** (Cauchy-Schwarz Ungleichung (Hölder,  $p = q = 2$ )).

$$|\mathbb{E}(XY|\mathcal{A}')| \leq \sqrt{\mathbb{E}(|X|^2|\mathcal{A}')} \sqrt{\mathbb{E}(|Y|^2|\mathcal{A}')} \text{ oder} \\ |\text{Cov}(X, Y|\mathcal{A}')| \leq \sqrt{\mathbb{V}(X|\mathcal{A}')} \sqrt{\mathbb{V}(Y|\mathcal{A}')}$$

**Lemma 24** (Minkowski-Ungleichung).  $\mathbb{E}|X|^p < \infty$ ,  $\mathbb{E}|Y|^p < \infty$  dann gilt die Dreiecksungleichung  $P$ -f.s.

$$(\mathbb{E}(|X + Y|^p|\mathcal{A}'))^{\frac{1}{p}} \leq (\mathbb{E}(|X|^p|\mathcal{A}'))^{\frac{1}{p}} + (\mathbb{E}(|Y|^p|\mathcal{A}'))^{\frac{1}{p}} < \infty$$

auch für bedingte Erwartung.

Beweise als Übung.

**Lemma 25** (Austauschbarkeitseigenschaften). *Wenn  $X, Y \in \mathbb{R}^k$  SGn mit identischen Verteilungen sind, i.Z.  $X \sim Y$ , dann gilt auch  $g(X) \sim g(Y)$  für messbare  $g(\cdot)$ .*

$$P[g(X) \in A] = P^X[g^{-1}(A)] = P^Y[g^{-1}(A)] = P[g(Y) \in A]$$

**Lemma 26.**  $X, Y$  SGn mit identischer Verteilung. Wenn  $(X, Z) \sim (Y, Z)$  für SG  $Z$ , dann gilt  $\mathbb{E}[g(X)|Z] = \mathbb{E}[g(Y)|Z] P$ -f.s.

*Proof.* TODO □

**Beispiel.** TODO

Für den bedingten Erwartungswert bestimmt nur die Sigmaalgebra  $\mathcal{A} \mathbb{E}(X|\mathcal{A})$ , nicht die konkrete SG. Wenn  $\sigma(Y_1) = \sigma(Y_2)$  (etwa bei  $Y_1, Y_2$  iid) ist  $P$ -f.s.  $\mathbb{E}(X|Y_1) = \mathbb{E}(X|Y_2)$ .

Die Information in der Bedingung ist in der Sigma-Algebra enthalten, die konkrete SG (bei  $Y_1 \sim Y_2$ ) ist nicht wesentlich.

Im Fall einer diskret verteilten SG  $Y$  für  $\mathbb{E}(X|Y)$  (der für die Motivation des bedingten Erwartungswerts erklärt wurde) ist  $\mathbb{E}(X|Y = y) = h(y)$  eine Funktion von  $Y$ .

Das bleibt im allgemeinen Fall erhalten. Sei  $\mathcal{A}' = \sigma(T)$  und TODO Grafik

Nach dem Faktorisierungslemma ist eine  $\mathcal{A}'$ -messbare Funktion eine Funktion von  $T$ .  $\exists h$ , sodass  $\mathbb{E}[X|T] = h \circ T$  mit  $\mathbb{E}[X|T = t] = h(t) P$ -f.s. Die bedingte Wahrscheinlichkeit führt für jedes  $A \in \mathcal{A}$  auf eine messbare Funktion  $h_A(\cdot)$  mit  $P(A|T = t) = \mathbb{E}[1_A|T = t] = h_A(t)$ .

Wir werden (erfolgreich) versuchen  $P(.|T = .)$  als Wahrscheinlichkeitsmaß für die Bestimmung von  $\mathbb{E}(X|T)$  als Erwartungswert bezüglich  $P(.|T)$  zu verwenden.

Das ursprüngliche Maß  $P(.)$  wird durch das "bedingte Wahrscheinlichkeitsmaß"  $P(.|T)$  ersetzt und damit das von  $X$  induzierte W-Maß  $P(X \in .|T)$  und Erwartungswerte  $\mathbb{E}(g(X)|T)$  etc. ermittelt. Damit ist eine konstruktive Bestimmung von  $\mathbb{E}(X|T)$  möglich.

Das folgende einfache Beispiel gibt ein Prinzip für die Bestimmung von  $\mathbb{E}(.|\mathcal{A}')$  bei endlichem  $\mathcal{A}'$ .

**Beispiel. TODO**

### 5.0.1 Geometrische Interpretation

Für einen Hilbertraum  $(H, \langle ., . \rangle)$  mit Skalarprodukt  $\langle x, y \rangle \in \mathbb{R}$  ist  $x_0$  die Projektion von  $X$  auf einen abgeschlossenen Unterraum  $(H_0, \langle ., . \rangle)$  mit  $H_0 \subseteq H$  definiert.

$x_0$  ist die einzige Lösung von  $\|x - x_0\| = \inf_{z \in H_0} \|x - z\|$  und  $x_0 \in H_0$ . Die Projektion erfüllt  $x - x_0 \perp x_0$ , d.h.  $\langle x - x_0, x_0 \rangle = 0$  für alle  $x \in H$ .

Zu  $H = L_2(\Omega, \mathcal{A}, P)$  mit  $\langle X, Y \rangle = \mathbb{E}XY$  ist  $H_0 := L_2(\Omega, \mathcal{A}_0, P|_{\mathcal{A}_0})$  bei  $\mathcal{A}_0 \subseteq \mathcal{A}$  ein abgeschlossener Unterraum. Die Projektion von  $X \in L_2(\Omega, \mathcal{A}, P)$  auf  $L_2(\Omega, \mathcal{A}_0, P)$  ist  $X_0$  eine  $\mathcal{A}_0$ -messbare Funktion, die auch  $\mathbb{E}(X_0(X - X_0)) = \mathbb{E}(X_0X - X_0^2) = 0$  erfüllt.  $X_0 = \mathbb{E}(X|\mathcal{A}_0)$  ist  $\mathcal{A}_0$ -messbar und

$$\mathbb{E}(X_0(X - X_0)) = \mathbb{E}(X_0X) - \underbrace{\mathbb{E}(X_0\mathbb{E}(X|\mathcal{A}_0))}_{\mathbb{E}(X_0X|\mathcal{A}_0)} = \mathbb{E}(X_0X) - \underbrace{\mathbb{E}[\mathbb{E}(X_0X|\mathcal{A}_0)]}_{\mathbb{E}X_0X \text{ (totale Erwartung)}} = 0$$

Das tatsächlich  $X_0 = \mathbb{E}(X|\mathcal{A}_0)$  die Projektion von  $X \in L_2$  auf  $L_2(\Omega, \mathcal{A}_0, P)$  ist, zeigt der nächste Satz.

### 5.0.2 Bedingte Erwartung als Projektion

**Satz 53** (5.BE).  $X \in \mathcal{L}^2(P)$ . Für jedes  $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$  gilt:

$$\|X - Y\|_2^2 \geq \|X - \mathbb{E}(X|\mathcal{F})\|_2^2$$

Gleichheit für  $Y = \mathbb{E}(X|\mathcal{F})$  P-f.s.

*Proof.* TODO □

In einem Hilbertraum ist die Projektion auf einen Unterraum, das Element mit kleinstem Abstand (Norm) aus dem Unterraum.

**Folgerung** (Prognose, Residuum).  $X, Y$  SG aus  $\mathcal{L}^2$ , dann ist die SG  $\mathbb{E}(X|Y) = H(Y)$  die Prognose von  $X$  mit kleinster Varianz, wenn  $Y$  beobachtet wurde  $\hat{X} = \mathbb{E}[X|Y]$

Residuum ist die SG des Restes oder Fehler  $X - \hat{X} = X - \mathbb{E}[X|Y] = R$ .

$R$  und  $\hat{X}$  sind unkorreliert

$$\mathbb{E}R = \mathbb{E}[X - \mathbb{E}[X|Y]] = \mathbb{E}X - \underbrace{\mathbb{E}[\mathbb{E}(X|Y)]}_{\mathbb{E}X} = 0$$

$$\mathbb{E}\hat{X} = \mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}X = \mu$$

$$\text{Cov}(\hat{X}, R) = \mathbb{E}[\underbrace{(X - \mathbb{E}(X|Y))}_R \underbrace{(\mathbb{E}(X|Y) - \mu)}_{\hat{X}}] = \underbrace{\mathbb{E}[X\mathbb{E}(X|Y)]}_{\mathbb{E}[\mathbb{E}(X|Y)]} - \mathbb{E}(\mathbb{E}(X|Y)^2) - \underbrace{\mu \mathbb{E}X}_{\mu} + \underbrace{\mu \mathbb{E}\mathbb{E}[X|Y]}_{\mu} = 0$$

**Beispiel. TODO**

### 5.0.3 Reguläre bedingte Wahrscheinlichkeit

Bei diskreten SGn  $X$  bzw.  $Y$  ist  $P(X = i|Y = k)$  tatsächlich eine Wahrscheinlichkeitsverteilung für jedes  $k$ . Mit dieser bedingten Verteilung ist dann

$$\mathbb{E}[X|Y = k] = \sum_m mP[X = m|Y = k]$$

der Erwartungswert dieser bedingten Verteilung.

Bei stetigen Verteilungen von  $X, Y$  mit Dichten  $\frac{dP^X}{d\lambda} = f_X$ ,  $\frac{dP^Y}{d\lambda} = f_Y$ ,  $\frac{dP^{(X,Y)}}{d\lambda} = f_{X,Y}$  wurde die bedingte Dichte bereits durch

$$f(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

für  $f_Y(y) > 0$   $P^Y$ -f.s. Für die Funktion  $H : Y \rightarrow \mathbb{E}[X|Y = g] := \int x f(x|y) d\lambda(x)$  gilt, dass  $H$  messbar bezüglich  $\sigma(Y)$  ist.

Der Schnitt  $\left(\frac{f_{X,Y}(x,y)}{f_Y(y)}\right)_y$  ist messbar und integrierbar, daher ist  $f(x|y)$  messbar und wegen  $\mathbb{E}X < \infty$  auch  $H$  messbar bezüglich  $\sigma(Y)$ .

Für  $B$  mit  $Y^{-1}(B) \in \mathcal{B}$ :

$$\begin{aligned} \int_B H(y) dP^Y &= \int_B \int_{\mathbb{R}} x f(x|y) d\lambda(x) dP^Y = \int_B \int_{\mathbb{R}} x f(x|y) d\lambda(x) f_Y(y) d\lambda(y) = \\ &= \int_{B \times \mathbb{R}} \underbrace{x f(x|y) f_Y(y)}_{f(x,y)} d\lambda \times \lambda = \int_B x \underbrace{\int_{\mathbb{R}} f(x,y) d\lambda(y)}_{f_X(x)} d\lambda(x) = \int_B x dP^X \end{aligned}$$

Transformiert gilt

$$\begin{aligned} \int_{Y^{-1}(B)} H dP &= \int_{Y^{-1}(B)} X dP \text{ also ist} \\ \mathbb{E}(X|Y) &= \mathbb{E}(X|\sigma(Y)) = \int x f(x|Y) d\lambda(x) \end{aligned}$$

Bei Verteilungen mit Dichten bezüglich  $\lambda$  auf  $(\mathbb{R}, \mathcal{B})$  existiert eine bedingte Wahrscheinlichkeit, deren Erwartung tatsächlich dem bedingten Erwartungswert entspricht.

In wie weit kann  $P(A|\mathcal{A}_0) := \mathbb{E}[1_A|\mathcal{A}_0]$  immer als bedingte Wahrscheinlichkeitsverteilung verstanden werden?

Für die Subalgebra  $\mathcal{A}_0 \subseteq \mathcal{A}$  sei  $P(A|\mathcal{A}_0)$  eine Version von  $\mathbb{E}(1_A|\mathcal{A}_0)$  für  $A \in \mathcal{A}$ .

Die Bedingungen für eine Wahrscheinlichkeitsverteilung sind:

1.  $P(\Omega|\mathcal{A}_0) = 1$   $P$ -f.s.

Das ist wegen  $1_\Omega = 1$  und der Definition des bedingten Erwartungswerts. Genauso

2.  $P(\emptyset|\mathcal{A}_0) = 0$   $P$ -f.s.

Da  $0 \leq 1_A \leq 1$  gilt auch

3.  $0 \leq P(A|\mathcal{A}_0) \leq 1$   $P$ -f.s

Bleibt die sigma-Additivität,  $A_i$  disjunkte Folge aus  $\mathcal{A}$

4.  $P(\bigcup_{i=1}^{\infty} A_i|\mathcal{A}_0) = \sum_{i=1}^{\infty} P(A_i|\mathcal{A}_0)$

Mit  $B_n := \bigcup_{i=1}^n A_i$  gilt  $\mathbb{E}(1_{B_n}|\mathcal{A}_0)$  und der monotonen Konvergenz  $1_{B_n} \nearrow 1_{\bigcup A_i}$  für bedingte Erwartung folgt

$$P\left(\bigcup_{i=1}^{\infty} A_i|\mathcal{A}_0\right) = \mathbb{E}[\lim 1_{B_n}|\mathcal{A}_0] = \sum_{i=1}^{\infty} \mathbb{E}(1_{A_i}|\mathcal{A}_0) = \sum P(A_i|\mathcal{A}_0) \text{ } P\text{-f.s.}$$

Die Suche nach einer (einheitlichen) bedingten Wahrscheinlichkeit, die hinter dem bedingten Erwartungswert steht, scheint erfolgreich. Das Problem liegt aber bei 4. daran, dass die Nullmenge, für die 4) nicht gilt, also nicht Sigma-Additivität gilt, von der Folge  $A_i$  abhängt. Es gibt insgesamt überabzählbar viele Nullmengen, für die 4) nicht gilt bzw. eventuell nicht gilt.

Zur Klärung dieses Sachverhalts betrachten wir Übergangswahrscheinlichkeiten bzw. Markov-Kerne.

**Definition 19** (Übergangskern / Markov-Kern).  $(\Omega_1, \mathcal{A}_1), (\Omega_2, \mathcal{A}_2) \dots$  Messräume  $K : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, \infty]$  heißt (sigma-endlicher) Übergangskern / Markov-Kern von  $\Omega_1$  auf  $\Omega_2$ , wenn

1.  $\omega_1 \mapsto K(\omega_1, A_2)$  ist eine messbare Abbildung ( $\mathcal{A}_1$  messbar) für jedes  $A_2 \in \mathcal{A}_2$
2.  $A_2 \mapsto K(\omega_1, A_2)$  ist für jedes  $\omega_1 \in \Omega_1$ , ein sigma-endliches Maß.

Ist das Maß  $K(\omega_1, \cdot)$  für jedes  $\omega_1 \in \Omega_1$  ein  $W$ -Maß, dann heißt  $K(\cdot, \cdot)$  auch Übergangswahrscheinlichkeit.



**Beispiel. TODO**

**Lemma 27** (Zusammensetzung von Kernen).  $K_i \in K(\Omega_i, \Omega_{i+1}), i = 1, 2$

$K(x_1, A_3) = K_1 \circ K_2(x_1, A_3) = \int_{\Omega_2} K_2(x_2, A_3) K_1(x_1, dx_2)$  ist Übergangskern von  $\Omega_1$  in  $\Omega_3$ .

Bedeutung: TODO Grafik

Aussagen:

- $K$  ist Markov-Kern (stochastisch)
- Wenn  $h \geq 0$  messbar bezüglich  $(\Omega_2, \mathcal{A}_2)$ , dann ist  $x_1 \rightarrow \int_{\Omega_2} h(x_2) K(x_1, dx_2)$  messbar bez.  $\mathcal{A}_1$ .

**Beispiel. TODO**

Wie im Beispiel mit Markov-Ketten, lassen sich Kerne zu gemeinsamen (Produkt-)Maßen zusammensetzen:

Es sei  $(\Omega_1, \mathcal{A}_1, \mu)$  ein endlicher Maßraum und  $K$  ein endlicher Übergangskern  $(\Omega_1 \mapsto \Omega_2, \mathcal{A}_2)$ . Es existiert ein eindeutiges, sigma-endliches Maß auf  $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2)$  mit der Bezeichnung  $\mu \times K := \mu \times K(A_1 \times A_2) = \int_{A_1} K(\omega_1, A_2) \mu(d\omega_1)$ .

Produkt von  $\mu \times K$  TODO BE Seite 36

**Definition 20** (Reguläre bedingte Wahrscheinlichkeit).  $(\Omega, \mathcal{A}, P), \mathcal{F} \subset \mathcal{A}$  Sigma-Unteralgebra.

- Eine Version von  $\mathbb{E}(1_A | \mathcal{F}) = P(A | \mathcal{F})(x)$  für alle  $A \in \mathcal{A}$  die durch einen stochastischen Kern  $(x, A) \mapsto K(x, A) = P(A | \mathcal{F})(x)$  definiert werden kann, heißt reguläre bedingte Wahrscheinlichkeit.
- $Y : \Omega \rightarrow (E, \mathcal{E})$  messbare SG. Ein stochastischer Kern  $K : (\Omega, \mathcal{F}) \rightarrow E$  mit  $K(\omega, B) = P[Y \in B | \mathcal{F}](\omega)[P]$  heißt reguläre bedingte Verteilung von  $Y | \mathcal{F}$ .
- Wird  $\mathcal{F} = \sigma(X)$  von  $X$  erzeugt mit  $X : \Omega \mapsto (E', \mathcal{E}')$ . Ein stochastischer Kern  $K(x, A) = P[Y \in A | X = x], x \in E', A \in \mathcal{E}$  von  $(E' \times \mathcal{E})$  heißt reguläre bedingte Verteilung von  $Y$  gegeben  $X$ .

Im allgemeinen muss eine solche Version nicht existieren.

Wenn der Bildraum der messbaren Funktion  $X$  ein polnischer Raum ist (vollständig, separabler metrischer topologischer Raum), dann existiert eine reguläre bedingte Wahrscheinlichkeit.  $\text{SGn } X : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^k, \mathcal{B}_k)$  erfüllen die Bedingungen.

**Satz 54** (6.BE).  $X : (\Omega, \mathcal{A}) \rightarrow (\Gamma, \mathcal{B})$  ( $\Gamma$  polnischer Raum mit Borel-Algebra). Dann existiert ein Markovkern  $K$  von  $(\Omega, \mathcal{A}_0) \rightarrow (\Gamma, \mathcal{B})$  für die Teil-Sigmaalgebra  $\mathcal{A}_0 \subseteq \mathcal{A}$  der Gestalt  $K(\omega, A) = P[X^{-1}(A) | \mathcal{A}_0](\omega)P$ -f.s. für jedes  $A \in \mathcal{B}$ .

Proof. TODO

□

Eine Borel-Äquivalenz wie oben existiert etwa für  $(\Gamma = \mathbb{R}^\infty, \mathcal{B} = \mathcal{B}_\infty)$ , reguläre bedingte Wahrscheinlichkeiten für Folgen  $\text{SGn}$ .

Damit ist  $\mathbb{E}[X | \mathcal{A}_0](\omega)$  eine Version der Erwartung der regulären bedingten Wahrscheinlichkeit (wie bei Dichten bereits bestätigt).

**Satz 55** (7.BE). Unter den Voraussetzungen des vorigen Satzes ist  $P$ -f.s.

$$\mathbb{E}[X | \mathcal{A}_0](\omega) = \int X(\omega') K(\omega, d\omega') = \int X(\omega) P[d\omega' | \mathcal{A}_0]$$

Proof. TODO

□

Im regulären Fall gilt das (bequeme) Einsetzungsprinzip.

**Satz 56** (8.BE). Bei Gültigkeit der Bedingungen wie im letzten Satz, d.h. für SG  $X_1, X_2$  existiert eine reguläre bedingte Verteilung  $P^{X_2 | X_1}$ . Dann gilt für  $g \in L_1, g : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$

$$\begin{aligned} \mathbb{E}[g(X_1, X_2) | X_1 = x_1] &= \int g(x_1, x_2) P(dx_2 | X = x_1) \forall x \in \Omega_1 \text{ und} \\ &\mathbb{E}[g(X_1, X_2) | X_1 = x_1] = \mathbb{E}[g(x_1, X_2)] \end{aligned}$$

Der Wert von  $X_1 = x_1$  wird in den bedingten Erwartungswert eingesetzt.

**Beispiel. TODO**

**Beispiel. TODO**

**Beispiel. TODO**

## 6 Gesetz der großen Zahlen

Ausgehend von einem wiederholbaren Experiment soll eine Methode zur Bestimmung der Wahrscheinlichkeitsverteilung, die dem Prozess zugrunde liegt, entwickelt werden. Eigenschaften und Charakteristika von  $P$  sollen auch aus der Beobachtung abgeleitet werden.

Oft genügt es, Momente zu bestimmen, um Aussagen über Wahrscheinlichkeiten zu erhalten.

Obwohl alle Begriffe längst definiert wurden (besonders in den Kapiteln EI, WR, LP, UN) und auch mannigfaltig verwendet wurden, sei zunächst eine Wiederholung zu Momenten vorangestellt.

- zentrales Moment:  $\mathbb{E}(X - \mathbb{E}X)^n$
- $n = 2$  Varianz:  $\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}X)^2$   
Eigenschaften:  $\mathbb{V}(aX + b) = a^2\mathbb{V}(X)$
- Kovarianz  $Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y$  oder  $\langle X, Y \rangle = Cov(X, Y) = \mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)$
- Standardisierte SG:  $X_0 = \frac{(X - \mathbb{E}X)}{\sqrt{\mathbb{V}(X)}}$
- Streuung oder Standardabweichung  $\sqrt{\mathbb{V}(X)}$
- Korrelation ist die Kovarianz der standardisierten SGn  $Cor(X, Y) = \frac{\mathbb{E}(X - \mathbb{E}X)(Y - \mathbb{E}Y)}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}} = \frac{\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y}{\sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}}$   
 $\mathbb{V}(X + Y) = \mathbb{V}(X) + \mathbb{V}(Y) + 2Cov(X, Y)$   
 $X, Y$  u.a.  $\implies Cov(X, Y) = 0$  unkorreliert  
 $\mathbb{V}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbb{V}(X_i) + \sum_{i \neq j} Cov(X_i, X_j) = \sum \mathbb{V}(X_i) + 2 \sum_{i < j} Cov(X_i, X_j)$   
 $Cov(aX + b, cY + d) = acCov(X, Y)$ ,  $|Cor(aX + b, cY + d)| = |Cor(X, Y)|$   
Wegen der Cauchy-Schwarz-Ungleichung gilt  $-1 \leq Cor(X, Y) \leq 1$  und  $|Cor(X, Y)| = 1 \implies X = aY + b$  für ein  $a, b \in \mathbb{R}$   
 $|Cov(X, Y)| = \sqrt{\mathbb{V}(X)}\sqrt{\mathbb{V}(Y)}Cor(X, Y)$
- Spezielle höhere Momente:  $\sigma_X = \sqrt{\mathbb{V}(X)}$
- Schiefe  $\mathbb{E}\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)^3$  (0 bei  $N(\mu, \sigma^2)$ )
- Kurtosis  $\mathbb{E}\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)^4$  (3 bei  $N(\mu, \sigma^2)$ )  
Daher wird meistens  $\mathbb{E}\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)^4 - 3$  als Kurtosis (Vergleich zu N) definiert.

**Beispiel.** TODO

**Beispiel.** TODO

Als Gesetz der großen Zahlen ist eigentlich eine Konvergenz von Mittelwerten  $\bar{X}_n$  gegen den Erwartungswert gemeint.

**Satz 57** (1.GZ Schwaches Gesetz der großen Zahlen).  $(X_n)_{n \in \mathbb{N}}$  ist eine unkorrelierte Folge SGn mit  $\sup_n \mathbb{V}(X_n) < \infty$ . Dann gilt für  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}X_i)\right| > \epsilon\right) = 0$$

Ist  $X_n$  schwach stationär, also  $\mathbb{E}X_i = \mathbb{E}X$ , dann gilt

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mathbb{E}X| > \epsilon) = 0$$

*Proof.* TODO

□

Die Existenz der Varianz ist (wie wir später sehen werden) für die Aussage nicht erforderlich, macht aber die Anwendungen sehr praktisch.

Wenn beispielsweise  $X_n$  eine iid. Folge ist, und die Erwartung bis auf  $\pm\epsilon$  angenähert werden soll, dann kann die Anzahl  $n$  an Beobachtungen abgeschätzt werden.

$$P(|\bar{X}_n - \mathbb{E}X| > \epsilon) \leq \frac{\mathbb{V}X}{n\epsilon^2} = \alpha$$

mit "Unsicherheit"  $\alpha$ , dann liefert  $n \geq \frac{\mathbb{V}X}{\alpha\epsilon^2}$  eine Mindestprobenanzahl.  $\alpha = 0.05$  ist üblich.

Auch allgemeine Momente der iid. Folge konvergieren in Wahrscheinlichkeit,

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i^k - \mathbb{E}X_i^k)\right| > \epsilon\right) \rightarrow 0, n \rightarrow \infty$$

wofür  $\mathbb{E}X^{2k} < \infty$  gefordert wird.

Eine wichtige Auswirkung des GGZ ist die Schätzbarkeit der Verteilungsfunktion.

$X_i$  ist iid Folge,  $Y_i := 1_{(X_i \leq t)}$  für festes  $t$ ,  $Y_i$  ist alternativverteilt mit  $p = P(X_i \leq t) = F(t)$ ,  $\mathbb{V}(Y_i) = p(1-p)$ . Da  $\sum_{i=1}^n Y_i \sim B_{n,p}$  ist  $\mathbb{V}(\sum Y_i) = np(1-p)$  und  $P(|\bar{Y}_n - p| > \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \rightarrow 0$ .

$\bar{Y}_n = \frac{\#\{X_i | X_i \leq t\}}{n}$  ist die relative Häufigkeit der Werte unter  $t$ . Die Genauigkeit hängt von  $\mathbb{V}(\sum Y_i)$  ab, d.h. wenn  $t = \text{med}(x)$ , der Median der Verteilung, dann ist die Genauigkeit minimal, da für  $p = \frac{1}{2}$   $p(1-p)$  maximal wird.

Allgemein formuliert ist das eine Variante des

**Satz 58** (Empirischen Gesetz der großen Zahlen). *Relative Häufigkeiten konvergieren gegen die entsprechenden Wahrscheinlichkeiten.*

$F_n^*(t) = \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(X_i)$  heißt empirische Verteilungsfunktion und für eine iid. Folge  $X_n$  gilt demnach  $\lim_{n \rightarrow \infty} F_n^*(t) = F(t) \forall t$ .

Konvergenz gilt nicht nur in der Wahrscheinlichkeit auch P-f.s. und sogar gleichmäßig:

$$P\left(\lim_{n \rightarrow \infty} \sup_t |F_n^*(t) - F(t)| = 0\right) = 1$$

(Satz von Glivenko-Cantelli - Fundamentalsatz der Statistik)

**Beispiel. TODO**

Aus der Theorie der Martingale haben wir bereits die Kolmogoroff-Ungleichung (MA 18) hergeleitet.  $S_n := \sum_{i=1}^n (Z_i - \mathbb{E}Z_i)$  für unabhängige  $Z_i$  ist ein Martingal. Dabei hätte auch genügt, dass die Summen unkorrelierte Zuwächse (statt unabhängiger) haben.

**Satz 59** (2.GZ, Kolmogoroff-Ungleichung).  $X_i \in \mathcal{L}_2$  und für  $S_n := \sum_{i=1}^n (X_i - \mathbb{E}X_i)$  gelte  $\int_C S_i(S_n - S_i) dP = 0$  für alle  $C \in \sigma(X_1, \dots, X_n)$ . Dann gilt für alle  $\epsilon > 0$

$$P\left[\max_{1 \leq k \leq n} |S_k| \geq \epsilon\right] \leq \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{V}(X_i)$$

Für unabhängige  $X_i$  sind die Summen  $\sum_{i=1}^k (X_i - \mathbb{E}X_i)$  und  $\sum_{i=k+1}^n (X_i - \mathbb{E}X_i)$  auch unabhängig und die Bedingung oben ist natürlich erfüllt.

Mit der Kolmogoroff-Ungleichung lässt sich folgende allgemein nützliche Konvergenzaussage zeigen.

**Satz 60** (3.GZ). Die stochastische Folge  $X_i$  sei unabhängig und aus  $L_2$ . Wenn  $\sum_{i=1}^{\infty} \mathbb{V}(X_i) < \infty$ , dann ist  $\sum_{i=1}^n (X_i - \mathbb{E}X_i) = S_n$  P-f.s. konvergent gegen  $S_{\infty} = \sum_{i=1}^{\infty} (X_i - \mathbb{E}X_i) \in L_2$ .

*Proof. TODO* □

Die folgenden Lemmata werden für das starke Gesetz der großen Zahlen benötigt.

**Lemma 28.** 1.  $\sum_{i=k}^{\infty} \frac{1}{n^2} \leq \frac{2}{k}$

2.  $a_n \in \mathbb{R}, a_n \rightarrow a$  dann gilt  $\lim_n \frac{1}{n} \sum_{i=1}^n a_i = a$

*Proof. TODO* □

Dieses Lemma ist ein Sonderfall des

**Lemma 29** (Kronecker).  $b_n > 0$  und  $b_n \nearrow \infty$ , dann gilt für eine summierbare Folge  $a_i \in \mathbb{R}$

$$\sum_{i=1}^{\infty} a_i \in \mathbb{R} \implies \lim_n \frac{1}{b_n} \sum_{i=1}^n a_i b_i = 0$$

Mit dem Kronecker-Lemma und Satz 3.GZ erhält man:

**Satz 61** (4.GZ, 1. Gesetz der großen Zahlen).  $X_i$  sei eine Folge unabhängiger SGn,  $X_i \in L_2$  und  $\sum_{i=1}^{\infty} \frac{V(X_i)}{n^2} < \infty$ , so gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (X_i - \mathbb{E}X_i) = 0 \text{ P-f.s.}$$

*Proof.* TODO □

Bilden die  $X_i$  eine u.a. Folge mit identischem Erwartungswert  $\mathbb{E}X_i = \mathbb{E}X$  gilt demnach  $\bar{X}_n \rightarrow \mathbb{E}X$  P-f.s.

Für die Gültigkeit des letzten GGZ ist die Existenz der Varianz nicht erforderlich:

**Satz 62** (5.GZ, 2. Gesetz der großen Zahlen). Die Folge  $X_i, i \geq 1$  SGn sei eine iid. Folge mit Erwartungswert  $\mathbb{E}X_i = \mathbb{E}X$ ,  $\mathbb{E}X$  existierte. Dann gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}X \text{ P-f.s.}$$

*Proof.* TODO □

**Beispiel.** TODO

Die für das schwache Gesetz gebrachten Beispiele gilt auch das starke Gesetz.

Wenn etwa  $\mathbb{E}X^k < \infty$  für eine iid. Folge  $X_n \sim X$ , dann gilt  $\frac{1}{n} \sum X_i^k \rightarrow \mathbb{E}X^k$  P-f.s. ( $\mathbb{E}X^{2k} < \infty$  ist nach dem 2.GGZ nicht erforderlich).

Auch das empirische Gesetz der großen Zahlen gilt in der Version mit f.s.- Konvergenz.

Wenn  $h_n(A)$  die relative Häufigkeit des Ereignisses  $A$  von  $n$  unabhängigen Versuchen bezeichnet, dann gilt P-f.s.  $h_n(A) = \frac{1}{n} \sum_{i=1}^n 1_A \rightarrow \mathbb{E}(1_A) = P(A)$  die Konvergenz gegen die Wahrscheinlichkeit.

Auch die empirische Verteilungsfunktion  $F_n^*$  für eine iid. Folge  $X_n$  konvergiert P-f.s. gegen die Verteilungsfunktion  $F(\cdot)$  von  $X$ .

$$\text{P-f.s. } \lim_n F_n^*(t) = \lim_n \frac{1}{n} \sum_{i=1}^n 1_{(-\infty, t]}(X_i) = F(t)$$

Auch bei Anwendungen, wie der Monte-Carlo-Integration, gilt das starke GGZ, da  $\int_a^b f(t)dt = I < \infty$  und  $Y_i = f(U_i), U_i \sim U_{a,b}$

Auf ähnliche Art lassen sich verschiedene Sätze, beispielsweise Approximationssätze zeigen.

**Satz 63** (Approximationssatz von Weierstraß).  $f : [0, 1] \rightarrow \mathbb{R}$  ist eine stetige Funktion. Dann existiert zu jedem  $\epsilon > 0$  ein Polynom  $p$  mit  $\|f - p\|_{\infty} = \max_{x \in [0, 1]} |f(x) - p(x)| \leq \epsilon$

*Proof.* TODO □

Für das starke Gesetz der großen Zahlen gibt es eine Umkehrung.

**Satz 64** (6.GZ).  $X_n$  sei eine iid. Folge. Wenn das Stichprobenmittel gegen einen Grenzwert  $m$  konvergiert  $\bar{X}_n \rightarrow m$  P-f.s., dann sind die  $X_n$  integrierbar und  $\mathbb{E}X_n = m$ .

*Proof.* TODO □

Diese Umkehrung liefert ein zusätzliches Argument für die Bedeutung des empirischen GGZ, die asymptotisch festgelegte Wahrscheinlichkeit ist eindeutig.

Für abhängige Folgen gibt es Ergodensätze für die Konvergenz des Stichprobenmittels.

## 7 Schwache Konvergenz und zentraler Grenzwertssatz

Hier wird die Konvergenz von Wahrscheinlichkeitsmaßen und Verteilungen behandelt. Es scheint naheliegend, einfach für eine Folge von Wahrscheinlichkeitsmaßen  $P_n$  und einen Grenzwert (Grenzmaß)  $P$   $P_n(A) \rightarrow P(A)$  für alle  $A \in \mathcal{A}$  zu fordern. Tatsächlich ist diese Forderung zu stark, die Eindeutigkeit des Grenzwerts und die Übertragbarkeit auf von stochastischen Folgen  $X_n$  induzierten Wahrscheinlichkeitsmaßen  $P^{X_n}$ , bereiten so Schwierigkeiten. Es wird sich herausstellen, dass  $P_n(A) \rightarrow P(A)$  nur für  $A$  mit  $P(\partial A) = 0$  ( $\partial A = \bar{A} \setminus \dot{A}$ , der Rand von  $A$ ) gefordert wird.

Dazu eine klärende Darstellung der Äquivalenz von Wahrscheinlichkeitsmaßen auf Borelräumen. Es sei jetzt immer  $(\Omega, \mathcal{B}(\Omega))$  ein Messraum mit einer Metrik  $d$  auf  $\Omega$  und der Borel-Sigmaalgebra, also etwa  $(\mathbb{R}, \mathcal{B})$  oder  $(\mathbb{R}^k, \mathcal{B}_k)$ .

**Lemma 30.** Die W-Maße  $P_1, P_2$  auf  $(\Omega, \mathcal{A} = \mathcal{B}(\Omega))$  sind genau dann äquivalent  $P_1 = P_2$ , wenn für alle  $f \in C_b$  (beschränkt und stetig)  $\int f dP_1 = \int f dP_2$ .

*Proof.* TODO □

**Definition 21** (schwach konvergent).  $P_n$  sei eine Folge von W-Maßen auf  $(\Omega, \mathcal{B}(\Omega))$  wie auch  $P$ , dann heißt  $P_n$  schwach konvergent gegen  $P$ , wenn für alle  $f \in C_b(\Omega)$  (stetig und beschränkt) gilt

$$\lim_{n \rightarrow \infty} \int f dP_n = \int f dP$$

$P_n \xrightarrow{w} P$  weak convergence.

Mit dieser Definition ist die Konvergenz der W-Maße auf die Stetigkeit eines Funktional auf dem Raum  $C_b$  zurückgeführt.

**Beispiel.** TODO

**Definition 22** (Stetigkeitsmenge). Für das W-Maß  $P$  auf  $(\Omega, \mathcal{A})$  heißt  $s \in \mathcal{A} = \mathcal{B}(\Omega)$  eine Stetigkeitsmenge wenn  $P(\partial S) = P(\bar{S} \setminus \dot{S}) = 0$ , also  $P(\bar{S}) = P(\dot{S})$ . ( $\dot{S}$  bezeichnet das Innere von  $S$ ).

**Satz 65** (1.KG, Portemanteau-Theorem).  $P, P_n$  eine Folge von W-Maßen auf  $(\Omega, \mathcal{A} = \mathcal{B}(\Omega))$ . Dann sind äquivalent:

1.  $P_n \xrightarrow{w} P$
2. Für jedes offene  $U$  gilt  $\liminf_n P_n(U) \geq P(U)$
3. Für jedes abgeschlossene  $G$  gilt  $\limsup_n P_n(G) \leq P(G)$
4. Für jede  $P$ -Stetigkeitsmenge  $S$  gilt  $\lim_{n \rightarrow \infty} P_n(S) = P(S)$

*Proof.* TODO □

Mit der Bezeichnung "Portemanteau" oder auch "Portmanteau" wird hier nicht Bezug auf eine Person genommen, der der Satz zuzuordnen ist. Es bedeutet "Handkoffer" und gemeint ist, dass es die wesentlichen Werkzeuge für "Konstruktionen" von Aussagen zur schwachen Konvergenz enthält.

Dafür ist folgende äquivalente Definition für schwache Konvergenz ein gutes Beispiel

**Definition 23** (schwach konvergent).  $F_n$  sei eine Folge von Verteilungsfunktionen.  $F_n$  heißt schwach konvergent gegen die VF  $F$ , wenn für jede Stetigkeitsstelle  $x \in \mathbb{R}$  von  $F$   $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ . Notation:  $F_n \Rightarrow F$ .

Dass diese Definition von schwacher Konvergenz äquivalent zur ersten Definition ist, zeigt der folgende Satz.

**Satz 66** (2.KG).  $F_n$  ist eine Folge von VF, die zu W-Maßen  $P_n$  gehören, wie auch  $F$  zu  $P$  gehört. Dann ist äquivalent:

1.  $P_n \xrightarrow{w} P$
2.  $F_n \Rightarrow F$

*Proof.* TODO □

**Definition 24** (Verteilungskonvergenz). Werden die Verteilungsfunktionen  $F_n$  SGN  $X_n$  zugeordnet  $F_n = F_{X_n}$  und  $F = F_X$  dann heißt  $X_n$  verteilungskonvergent oder konvergent in Verteilung. In Zeichen  $X_n \xrightarrow{D} X$ , wenn  $F_{X_n} \implies F_X$ , schwache Konvergenz der VFn bzw.  $P^{X_n} \xrightarrow{w} P^X$  schwache Konvergenz der induzierten W-Maße.

Alle drei Konvergenzen entsprechen der selben Art einer stochastischen Konvergenz. Die Konvergenz ist wohldefiniert, aus der Darstellung (Lemma, 1KG) folgt die Eindeutigkeit von  $P$ , wenn  $P_n \xrightarrow{w} P$ . Auch in der Version über die Verteilungsfunktionen ist die Grenzverteilung eindeutig:  $F_n \implies F$  und  $F_n \implies G$ , dann gilt  $F = G$ .

$F$  und  $G$  haben insgesamt nur höchstens abzählbar viele Unstetigkeitsstellen. Für jedes  $x \in \mathbb{R}$  existiert eine Folge  $x_n$  von gemeinsamen Stetigkeitspunkten mit  $x_n \searrow x$ . Wegen der Rechtsstetigkeit von  $F$  und  $G$  gilt

$$F(x) = \lim_k F(x_k) = \lim_k \lim_{n \rightarrow \infty} F_n(x_k) = \lim_k G(x_k) = G(x)$$

Die Verteilungskonvergenz auch für Unstetigkeitsstellen zu verlangen, führt aus dem Raum der Verteilungen eventuell heraus.

**Beispiel.** TODO

Für diskrete Verteilungen gilt die schwache Konvergenz, wenn die Punktwahrscheinlichkeiten konvergieren  $P(X_n = k) \rightarrow P(X = k)$ . (Übung)

Wenn  $P_n, P$  durch ein sigma-endliches Maß  $\mu$  dominiert wird,  $P_n \ll \mu, P \ll \mu$ , dann folgt aus der Konvergenz der Dichten  $\frac{dP_n}{d\mu} := f_n \rightarrow f = \frac{dP}{d\mu}$   $\mu$ -f.ü. die Konvergenz  $P_n \xrightarrow{w} P$ . Das folgt sofort aus dem Satz von Scheffe (Satz 12.LP):

$$\lim_{n \rightarrow \infty} f_n = f \mu\text{-f.ü.} \implies \lim_n \sup_{A \in \mathcal{A}} |P_n(A) - P(A)| = 0$$

das heißt die Totalvariation konvergiert gegen 0.

Totalvariation von W-Maßen  $P, Q$

$$d_{TV}(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$$

Da  $d_{TV}(P_n, P) \rightarrow 0 \implies P_n \xrightarrow{w} P$  (Konvergenz nur für Stetigkeitsmengen) ist die Konvergenz der Dichten stärker als die schwache Konvergenz.

Bisher war unter den Konvergenzarten die Konvergenz in der Wahrscheinlichkeit die "schwächste", folgte also aus den anderen Konvergenzen. Die Bezeichnung "schwache" Konvergenz wird dadurch gerechtfertigt, dass Konvergenz in W. stärker als die schwache Konvergenz ist. Dazu ein vorbereitendes Lemma:

**Lemma 31.**  $X, Y$  SGN. Dann gilt für alle  $x, y \in \mathbb{R}$

$$P(Y \leq y) \leq P(X \leq x) + P(|X - Y| > x - y)$$

*Proof.* TODO □

**Satz 67** (3.KG). Die Folge  $X_n$  konvergiere in Wahrscheinlichkeit gegen  $X$ ,  $X_n \xrightarrow{W} X$ . Dann gilt auch  $F_{X_n} \implies F_X$  die schwache Konvergenz.

*Proof.* TODO □

Die Umkehrung kann i.a. nicht gelten. Etwa wenn  $X_0, X_1, X_2, \dots$  eine iid Folge mit  $V(X_0) \neq 0, F_n = F_0$  für die VFn  $F_n = F_{X_n}$  aber  $X_n \not\xrightarrow{W} X_0$  in irgendeiner anderen Konvergenzart.

Auch bei einer abhängigen Folge, etwa  $P(X_0 = k) = \frac{1}{2}$  für  $k = -1 \vee 1$  und  $X_n = (-1)^n X_0$ .  $F_{X_n} = F_{X_0}$ , also  $F_{X_n} \implies F_{X_0}$  aber  $P(|X_{2n+1} - X_0| > 1) = 1 \not\xrightarrow{W} 0$ .

Eine spezielle Ausnahme ist die Konvergenz gegen ein Dirac-Maß, also gegen eine Konstante.

**Satz 68** (4.KG). Stochastische Folge  $X_n$  und  $c \in \mathbb{R}$ . Dann sind äquivalent

1.  $X_n$  konvergiert in Wahrscheinlichkeit gegen  $c$
2.  $X_n$  ist Verteilungs-konvergent gegen  $c$ .

*Proof.* TODO □

Aus den vorigen Beispielen ist zu erkennen, dass die Konvergenz in Verteilung keine Konvergenz zugehöriger SGn, wie etwa  $X_n \rightarrow X_0$  P-f.s., ergeben kann. Es lassen sich aber (in einem anderen W-Raum) SGn  $X_n$  und  $X$  mit identischer Verteilung zu  $F_n, F$  mit  $X_n \rightarrow XP$ -f.s. konstruieren.

Die Grundlage dafür ist die verallgemeinerte Inverse eine VF  $F^{-1}(y) := \inf\{x | F(x) \geq y\}$ , die wir schon zur Konstruktion von Zufallszahlen einer Verteilung mit VF  $F$  betrachtet haben. Wenn  $F$  eine VF ist, dann ist  $F^{-1}$  linksstetig und es gilt  $0 < y < 1 \implies y \leq F(x) \iff F^{-1}(y) \leq x$ .

Aus diesen Eigenschaften wird folgendes Lemma abgeleitet.

**Lemma 32.**  $F_n \implies F$ , wobei  $F_n, F$  VFn sind. Dann gilt für die verallgemeinerte Inversen  $F_n^{-1}$  und  $F^{-1}$ , dass  $F_n^{-1}(x) \rightarrow F^{-1}(x)$  für alle Stetigkeitspunkte  $x$  von  $F^{-1}$ .

ohne Beweis.

**Satz 69** (5.KG, Darstellungssatz von Skorochod).  $F, F_n$  seien VFn und  $F_n \implies F$ . Dann existiert ein W-Raum  $(\Omega, \mathcal{A}, P)$  und SGn  $X_n$  und  $X$  auf  $(\Omega, \mathcal{A}, P)$  mit  $X_n \sim F_n$  und  $X \sim F$  mit  $\lim_{n \rightarrow \infty} X_n = XP$ -f.s.

*Proof.* TODO □

Für die oben konstruierte Folge  $X_n$  aus Satz 5.KG gilt natürlich auch  $X_n \xrightarrow{P} X$  die Konvergenz in Wahrscheinlichkeit. Trotzdem können wir uns damit nicht Satz 3.KG ersparen, da dort  $X_n$  die VF  $F_n$  besitzt, hier (bei Satz 5.KG) nur eine Folge von SGn mit identischer Verteilung aber auf einem andren Raum sind.

Für die weitere Diskussion der schwachen Konvergenz ist die Fourier-Transformation wesentlich.

## 7.1 Charakteristische Funktion

Zu einer SG  $X$  mit W-Maß  $P^X$  wurde bereits die Momenterzeugende Funktion  $M(t) = \mathbb{E}e^{tX}, t \in \mathbb{R}$  und für  $X \geq 0$  auch die Laplace-Transformierte  $\mathcal{L}(t) = \mathbb{E}e^{-tX}, t \geq 0$  eingeführt. Diese Maß-Transformationen erleichtern die Analyse des Verhaltens der W-Maße und charakterisieren diese auch. Beide Transformationen sind nicht für alle SGn brauchbar,  $M(\cdot)$  muss nicht existieren (außer  $t = 0$ ) und  $\mathcal{L}(\cdot)$  verlangt  $X \geq 0$ . Die komplexwertige Version von  $M(\cdot)$  hat diese Nachteile nicht.

**Definition 25** (Fouriertransformierte, charakteristische Funktion).  $\mu$  sei ein endliches Maß auf  $(\mathbb{R}, \mathcal{B})$ , dann heißt  $\phi(t) := \int e^{itx} d\mu(x)$  die Fouriertransformierte von  $\mu$ . Ist  $\mu = P^X$  von einer SG  $X$  so ist  $\phi(t) = \mathbb{E}e^{itX}$  die charakteristische Funktion von  $X$ .

Die Integration komplexwertiger Funktionen unterliegt folgenden Eigenschaften  $\int f d\mu = \int R(f) d\mu + i \int I(f) d\mu$  (Realteil  $R(\cdot)$ , Imaginärteil  $I(\cdot)$ ) und  $f$  ist integrierbar, genau dann, wenn  $R(f)$  und  $I(f)$  integrierbar ist, damit auch  $|f| = \sqrt{f \cdot \bar{f}}$ , ( $\bar{f}$  konjugierte von  $f$ ) weil  $\max(R(f), I(f)) \leq |f| \leq R(f) + I(f)$ .

Es gilt auch die Dreiecks-Ungleichung  $|\int f d\mu| \leq \int |f| d\mu$  und  $\int \bar{f} d\mu = \overline{\int f d\mu}$ .

Elementare Eigenschaften der CF (charakteristischen Funktion) von SGn sin:

1.  $\phi_X$  sei CF von  $X$ ,  $Y := aX + b, a, b \in \mathbb{R}$  dann ist  $\phi_Y(t) = e^{itb} \phi_X(at)$

TODO

2. Ist  $X$  symmetrisch um 0 verteilt, also  $X \sim -X \sim P^X$ , dann ist  $\phi_X$  reellwertig, da TODO

3.  $X$  u.a.  $Y$ ,  $S = X + Y$  dann ist die CF von  $S$   $\phi_S(t) = \phi_X(t) \phi_Y(t)$

TODO

4. Die Fouriertransformierte eines W-Maßes existiert für alle  $t \in \mathbb{R}$ , ist beschränkt  $|\phi(t)| \leq 1$  und ist gleichmäßig stetig.

TODO

5. Die Momente von  $X$  werden mit der Ableitung von  $\phi_X$  berechnet.

TODO



Der Name "charakteristische Funktion" rechtfertigt, dass zu jedem W-Maß  $P$  genau eine CF  $\phi$  existiert. Dafür zunächst:

**Satz 70** (6.KG, Umkehrsatz). *Zum W-Maß  $P$  auf  $(\mathbb{R}, \mathcal{B})$  sei  $F$  die VF und  $\phi$  die charakteristische Funktion.*

*Für Stetigkeitspunkte  $a < b$  von  $F$  gilt*

$$F(b) - F(a) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt$$

ohne Beweis.

**Satz 71** (7.KG, Eindeutigkeitssatz). *Die Verteilungsfunktion  $F$  auf  $\mathbb{R}$  eines W-Maßes wird durch die CF  $\phi$  eindeutig festgelegt.*

*Proof.* TODO □

Wenn Dichten für die W-Maße existieren, kann aus der CF auch die Dichte bestimmt werden.

**Satz 72** (8.KG). *Die CF  $\phi$  des W-Maßes  $P$  auf  $(\mathbb{R}, \mathcal{B})$  sei integrierbar. Dann ist  $P \ll \lambda$  und die Dichte ist*

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

*Die Dichte  $\frac{dP}{d\lambda}$  ist gleichmäßig stetig und beschränkt.*

ohne Beweis.

**Beispiel.** TODO

Für den Zusammenhang von charakteristischen Funktionen und schwacher Konvergenz ist eine Form der "Kompaktheit" von W-Maßen verantwortlich.

**Definition 26** (straff). *Auf  $(\Omega, \mathcal{B}(\Omega))$  heißt eine Familie von W-Maßen  $\mathcal{P}$  straff, wenn es eine kompakte Menge  $K \in 2^\Omega$  zu jedem  $\epsilon > 0$  gibt, sodass für alle  $P \in \mathcal{P}$ :  $P[K] \geq 1 - \epsilon$*

Die W-Maße in  $\mathcal{P}$  haben gleichmäßig einen gemeinsamen kompakten Träger bis auf  $\epsilon$ . Bei Zahlen gibt es in einer kompakten Menge eine konvergente Teilfolge jeder Folge.

**Beispiel.** TODO

Entsprechend wird auch die Straffheit einer Familie von VFen  $\{F_i | F_i \text{ ist VF, } i \in I\}$  durch  $|F_i(M) - F_i(-M)| > 1 - \epsilon$  für alle  $i \in I$  für  $K = [-M, M], M < \infty$  definiert und die Familie  $X_t$  von SGn heißt straff, wenn  $\{P^{X_t}\}$  bzw.  $\{F_{X_t}\}$  straff sind.

Im allgemeinen sind Familien von Verteilungen nicht straff.

**Beispiel.** TODO

Die Erwartungswerte der Verteilungen von  $|X|$  im letzten Beispiel waren unbeschränkt.

**Satz 73** (9.KG). *Die Familie  $X_t, t \in I$  von SGn ist straff, wenn  $C := \sup\{\mathbb{E}|X_t| | t \in I\} < \infty$ .*

*Proof.* TODO □

**Satz 74** (10.KG). *Jede schwach konvergente Folge von Verteilungen,  $P_n \xrightarrow{w} P$ , ist eine straffe Folge.*

*Proof.* TODO □

Das Teilfolgeprinzip für schwach konvergente Folgen charakterisiert die Straffheit.

**Satz 75** (11.KG, Satz von Helly). *Für die Folge  $P_n$  von W-Verteilungen sind äquivalent:*

1.  $P_n$  ist eine straffe Folge von Verteilungen
2. Jede Teilfolge von  $P_n$  besitzt eine schwach konvergente Teilfolge.

*Proof.* TODO □



**Satz 76** (12.KG, Teilfolgenprinzip).  $P_n, P$  seien  $W$ -Verteilungen. Dann sind äquivalent:

1.  $P_n \xrightarrow{w} P$
2. Jede Teilfolge von  $P_n$  besitzt eine Teilfolge, die schwach gegen  $P$  konvergiert.

*Proof.* TODO □

Nun kann man den Zusammenhang von schwacher Konvergenz und der punktweisen Konvergenz der charakteristischen Funktion belegen:

**Satz 77** (13.KG). Sei  $P_n$  eine Folge von  $W$ -Verteilungen und  $\phi_n$  bezeichne die zugehörigen charakteristischen Funktionen. Wenn  $P_n$  straff ist und  $\phi_n$  punktweise konvergiert  $\phi_n(t) \rightarrow \phi(t)$ , mit  $\phi : \mathbb{R} \rightarrow \mathbb{C}$ , dann gilt:

- $\phi$  ist die charakteristische Funktion einer Verteilung  $P$ .
- Die Folge  $P_n \xrightarrow{w} P$  konvergiert schwach gegen  $P$ .

*Proof.* TODO □

**Satz 78** (14.KG, Stetigkeitssatz von Levy). Die Folge  $P_n$  von Verteilungen besitze die CFn  $\phi_n$  und die CF  $\phi$  gehöre zu  $P$ . Dann sind äquivalent:

1.  $P_n \xrightarrow{w} P$
2.  $\phi_n \rightarrow \phi$  konvergiert punktweise.

*Proof.* TODO □

**Beispiel.** TODO

Die schwache Konvergenz bleibt unter gewissen Transformationen erhalten. Es ist zu erwarten, dass stetige Transformationen stochastischer Größen die Verteilungskonvergenz erhält, da stetige Funktionen Grenzwerte erhalten.

**Satz 79** (15.KG, Continuous Mapping Theorem).  $X, X_i, i \geq 1$  sind SGn auf  $(\Omega, \mathcal{A}, P)$  und die Funktion  $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$  ist  $P$ -f.s. stetig. Dann gilt:

1.  $X_n \xrightarrow{\mathcal{D}} X$  (Verteilungskonvergenz)  $\implies f(X_n) \xrightarrow{\mathcal{D}} f(X)$
2.  $X_n \xrightarrow{P} X$  (Konvergenz in  $W$ .)  $\implies f(X_n) \xrightarrow{P} f(X)$
3.  $X_n \rightarrow XP$ -f.s.  $\implies f(X_n) \rightarrow f(X)P$ -f.s.

*Proof.* TODO □

**Folgerung.** Aus  $(X_n, Y_n) \xrightarrow{\mathcal{D}} (X, Y)$  folgt  $X_n + Y_n \xrightarrow{\mathcal{D}} X + Y$ ,  $X_n Y_n \rightarrow XY$  etc.

**Beberkung.** Aus  $X_n \xrightarrow{\mathcal{D}} X$  und  $Y_n \xrightarrow{\mathcal{D}} Y$  folgt nicht automatisch  $X_n + Y_n \rightarrow X + Y$ .

**Beispiel.** TODO

**Satz 80** (16.KG, Slutsky-Theorem).  $X_n, Y_n, X, n \geq 1$  sind SGn auf  $(\Omega, \mathcal{A}, P)$ . Es gilt

1.  $X_n \xrightarrow{\mathcal{D}} X$  und  $(X_n - Y_n) \xrightarrow{P} 0 \implies Y_n \xrightarrow{\mathcal{D}} X$
2.  $X_n \xrightarrow{P} X$  und  $(X_n - Y_n) \xrightarrow{P} 0 \implies Y_n \xrightarrow{P} X$
3.  $X_n \rightarrow XP$ -f.s. und  $(X_n - Y_n) \rightarrow 0P$ -f.s.  $\implies Y_n \rightarrow XP$ -f.s.

*Proof.* TODO □

Oft werden die Folgerungen aus dem letzten Satz als Slutsky-Theorem bezeichnet. Die folgenden Aussagen folgen aus den beiden letzten Sätzen (15.KG und 16.KG)

- $X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{\mathcal{D}} c$  (oder auch  $Y_n \xrightarrow{P} c$ )  $\implies (X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$

- $X_n \xrightarrow{P} X, Y_n \xrightarrow{P} c \implies (X_n, Y_n) \xrightarrow{P} (X, c)$
- $X_n \rightarrow XP\text{-f.s.}, Y_n \rightarrow cP\text{-f.s.} \implies (X_n, Y_n) \rightarrow (X, c)P\text{-f.s.}$
- $X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{\mathcal{D}} c \implies (X_n, Y_n) \xrightarrow{\mathcal{D}} (X, c)$  und für stetiges  $f$ :  $f(X_n, Y_n) \xrightarrow{\mathcal{D}} (f(X, c))$
- $X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{P} c$ , dann gilt  $X_n Y_n \xrightarrow{\mathcal{D}} Xc, X_n + Y_n \xrightarrow{\mathcal{D}} X + c, c \neq 0: \frac{X_n}{Y_n} \rightarrow \frac{X}{c}$

Das Verteilungsverhalten von standardisierten Summen führt auf ein allgemeines empirisches Gesetz.

### Beispiel. TODO

Das war die erste Version des zentralen Grenzwertsatzes von Abraham de Moivre (1730). Pierre Simon-Laplace verallgemeinerte die Konvergenzaussage für beliebiges  $p \in (0, 1)$ , formuliert als Satz von Moivre-Laplace im Jahr 1812.

Die Verallgemeinerung ist der Gauß zugeschriebene aber eigentlich erst von Lyapunov so formulierte

**Satz 81** (17.KG, Zentraler Grenzwertsatz).  $X_n \in \mathcal{L}_2$  ist eine iid. Folge auf  $(\Omega, \mathcal{A}, P)$ ,  $\mathbb{E}X_n = \mu$ ,  $\mathbb{V}X = \sigma^2 < \infty$ . Dann gilt

$$\sum \frac{X_i - \mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} Z \sim N(0, 1).$$

Proof. TODO □

**Beberkung.** Für eine iid Folge  $X_n$  mit  $\mathbb{E}X = \mu$  und  $\mathbb{V}X = \sigma^2 < \infty$  folgt mit obigem Konzept auch sofort das (schwache) Gesetz der großen Zahlen. Sei jetzt  $S_n = \frac{1}{n} \sum_{i=1}^n X_i$  und

$$\phi_{S_n}(t) = (\phi_X(\frac{t}{n}))^n = \underbrace{(1 - \frac{t^2}{2n^2} + o(\frac{1}{n}))^n}_{\rightarrow 1}$$

d.h.  $\lim \phi_{S_n}(t) = 1 \forall t \in \mathbb{R}$ .

Die Dirac-Verteilung in 0 hat CF  $e^{it0} = 1$  und für  $\mu \neq 0$  ist  $\phi_{S_n}(t) = e^{it\mu}$  entspricht  $\delta_\mu$ .

Aus Verteilungskonvergenz folgt hier auch Konvergenz in Wahrscheinlichkeit.

Eine allgemeine Version des ZGVS unter schwächeren Bedingungen wurde von J. Lindeberg (und Levy) formuliert.

**Definition 27** (Lindeberg-Bedingung). Die Folge  $X_n$  unabhängiger SGN auf  $(\Omega, \mathcal{A}, P)$  mit endlichen Varianzen  $\mathbb{V}X_n = \sigma_n^2 \neq 0$  und  $S_n^2 = \sum_{i=1}^n \sigma_i^2$  erfüllt die Lindeberg-Bedingung, wenn  $\forall \epsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{k=1}^n \int_{|X_k - \mathbb{E}X_k| > \epsilon S_n} (X_k - \mathbb{E}X_k)^2 dP = 0$$

Eine iid. Folge erfüllt diese Bedingung.

**Satz 82** (18.KG, ZGVS von Lindeberg). Die unabhängige Folge  $X_k$  auf  $(\Omega, \mathcal{A}, P)$  erfüllt die Lindeberg-Bedingung mit  $S_n^2 = \sum_{k=1}^n \mathbb{E}(X_k - \mathbb{E}X_k)^2$ , dann gilt

$$\sum_{k=1}^n \frac{X_k - \mathbb{E}X_k}{S_n} \xrightarrow{\mathcal{D}} Z \sim N(0, 1)$$

ohne Beweis.

Aus der Lindeberg-Bedingung folgt

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} P\left(\left|\frac{X_k - \mathbb{E}X_k}{S_n}\right| \geq \epsilon\right) = 0$$

(Die Folge  $X_n$  ist dann "asymptotisch vernachlässigbar").

Unter dieser Bedingung ist die Lindeberg-Bedingung sogar notwendig:

Ist die Folge  $X_n$  asymptotisch vernachlässigbar und gilt der ZGVS

$$\sum_{k=1}^n \frac{X_k - \mathbb{E}X_k}{S_n} \xrightarrow{\mathcal{D}} Z \sim N(0, 1)$$

dann erfüllt die Folge  $X_n$  die Lindeberg-Bedingung.