

Maß- und Wahrscheinlichkeitstheorie Skript Felsenstein W2022

Ida Hönigmann

November 15, 2022

1 Produkträume und Maße

Auf dem kartesischen Produkt von Grundmengen

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

wird eine Produkt-(Sigma)Algebra konstruiert, wobei $(\Omega_i, \mathcal{A}_i)$ Messräume sind. Es soll $(\Omega_i, \mathcal{A}_i)$ in (Ω, \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 $\mathcal{E}_i \subset 2^{\Omega_i}$. Ω_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

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

^aMan 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 $\Omega = \prod_i \Omega_i$.

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

\implies : Alle π_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 Ω_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 π_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 μ_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. μ_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}_∞ .

2. Ist μ_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 Ω_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 μ_1, μ_2 sigma-endlich sind, dann sind die Funktionen $f_B(x_1)$ bzw. $f_B(x_2)$ die "Dichten" bezüglich den Randmaßen μ_1 bzw. μ_2 . μ_1, μ_2 werden auch als marginale Maße bezeichnet.

Satz 4 (4.PR). 1. μ_2 sei sigma-endlich. Dann definiert

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

das Produktmaß auf (Ω, \mathcal{A}) .

2. Sind beide Maße μ_1, μ_2 sigma-endlich, dann ist μ 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 μ ein Maß auf (Ω, \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.

μ ist auf dem (durchschnitts-stabilen) Erzeuger definiert, da die Maße sigma-endlich sind, ist μ eindeutig (Eindeutigkeitssatz). Vice versa gelten alle Gleichungen analog bez. μ_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 $\Omega \mapsto \mathbb{R}$, μ_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} μ_2 -integrierbar $[\mu_1]$ f.ü. und f_{x_2} μ_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

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

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

ϕ_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 ϕ_1 integrierbar sein (wie auch ϕ_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). $\Omega = \Omega_1 \times \Omega_2, \mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2, \mu = \mu_1 \otimes \mu_2$
 ν sei absolut stetig bzgl. μ ($\nu \ll \mu$), es existiert ein $f = \frac{d\nu}{d\mu} \geq 0$ mit $\nu(A) = \int_A f d\mu$ (f integrierbar)
 μ sei sigma-endlich, ν 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 ϕ_1, ϕ_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 ν_1 $P[X \in A] = \int_A \frac{2}{\pi} \sqrt{1-y^2} d\lambda(y)$

ν 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 μ 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 ∞ -dim. Produkträume

Die Konstruktion der ∞ -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 $f(x, y) = f_1(x) \cdot \Phi\left(\frac{y}{\sigma_y}\right)$ (Φ 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). μ_F und μ_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). μ_1, μ_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 μ_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). μ_1, μ_2 seien Maße mit Dichten bezüglich λ : $\mu_1 = \int f_1 d\lambda$ 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 λ 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 (≥ 0) bis zum k -ten Erfolg (≤ 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). μ sei ein Maß mit Dichte f bez. λ . μ daher $\mu \ll \lambda$ und ν 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. μ sei eine Exponentialverteilung E_{x_1} mit Dichte $f(s) = e^{-s} 1_{(0, \infty)}(s)$. ν 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 λ_1, λ_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 ν ist ein signiertes Maß, wenn auf (Ω, \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 ν eine Dichte $f : \Omega \rightarrow \mathbb{R}$ bezüglich eines Maßes μ 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. ν , wenn $\forall B \subseteq A^+ : \nu(B) \geq 0$ erfüllt, A^- ist ν -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 Δ -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 Ω 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 Ω .

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 Ω . 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 ν bilden Maße μ_1, μ_2 mit

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

wobei μ_1, μ_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 ν^+, ν^- eine Jordanzerlegung von $\nu = \nu^+ - \nu^-$ und $\nu = \mu_1 - \mu_2$ für beliebige Maße μ_1, μ_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 ν 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 ν^+, ν^- , 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 ν 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). *ν und μ seien sigma-endliche Maße auf (Ω, \mathcal{A}) . Eine Lebesgue-Zerlegung von ν besteht aus Maßen ν_c, ν_s , sodass*

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

also ν_c absolut stetig bezüglich μ und ν_s singulär zu μ ist.

ν ist in Bezug auf μ 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 ν 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 μ, ν 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 $g \leq 1$ γ -f.ü.

Es gilt γ -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 μ -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 μ_1, μ_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 μ_1 und μ_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 ζ bestimmt.

3 Dichten und absolute Stetigkeit

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

$$\frac{dF}{dx} = \frac{d}{dx} \mu((-\infty, x]) = f \text{ und}$$

$$\mu((-\infty, x]) = \int_{-\infty}^x f(t) dt.$$

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 λ) 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 λ -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 μ 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 δ , 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

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

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

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

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). *content...*