Maß- und Wahrscheinlichkeitstheorie Skript Felsenstein W2022

Ida Hönigmann

November 10, 2022

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)**.**

$$\pi_i : \Omega \mapsto \Omega_i \ mit \ \pi_i(\omega_1, ..., \omega_n) = \omega_i$$

Die Produktalgebra wird von den Projektionen erzeugt.

Definition 1 (Produktalgebra)**.**

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

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

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

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].*

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

Dann gilt

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

[a] 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 $\Omega = \bigtimes_i \Omega_i$.

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 $E_i \in \mathcal{E}_i$

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

Da

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

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

$$F_k = E_{1,k} \times E_{2,k} \times ... \times E_i \times ... \times E_{n,k} \nearrow \pi_i^{-1}(E_i)$$
$$\lim F_k = \bigcup_{k=1}^{\infty} F_k = \pi_i^{-1}(E_i) \in \sigma(\mathcal{E}) \subset \tilde{\mathcal{A}}$$

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 $\sigma(\pi_1, ..., \pi_n) \subset \sigma(\mathcal{E})$. Nach obigem setze $\tilde{\mathcal{A}} = \sigma(\pi_1, ..., \pi_n) \implies \sigma(\mathcal{E}) \subseteq \sigma(\pi_1, ..., \pi_n)$ also $\sigma(\mathcal{E}) = \sigma(\pi_1, ..., \pi_n)$. ☐

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

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

Folgerung.

$$f_i : \Omega_0 \mapsto \Omega_i, f = (f_1, ..., f_n) =: \otimes_i f_i$$
$$f : \Omega_0 \mapsto \bigtimes_i \Omega_i = \Omega$$

Dann gilt:

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

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.

$\Longleftarrow$: $A \in \otimes \mathcal{A}_i$ 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 $\sigma(f) = \sigma(f_1, ..., f_n) \subseteq \mathcal{A}_0$ also ist $f \mathcal{A}_0 - \otimes \mathcal{A}_i$ messbar. ☐

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. ☐

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. TODO □

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 $(\Omega, \mathcal{A})$.

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. TODO □

Beispiel. *Insbesonders 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).

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

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

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. TODO □

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:

- $f \geq 0$*: 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* $\mathcal{A}_1$
- 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.*
- $\int_{\Omega_1} \int_{\Omega_2} |f_{x_1}| d\mu_2 d\mu_1 < \infty$ *oder* $\int_{\Omega_2} \int_{\Omega_1} |f_{x_2}| d\mu_1 d\mu_2 < \infty$*: Daraus folgt* f *integrierbar.*

Proof. TODO □

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

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

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 (Ordinatenmenge, Graph)**.** *Die Punkte unter einer positiven Funktion $f \geq 0$ heißt Ordinatenmenge $O_f = \{(x, y) | 0 \leq < \leq f(x)\}$ 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,

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

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 $(X_1, ..., X_n)$ sind auf $(\times_i \Omega_i, \otimes_i \mathcal{A}_i)$ mit

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

also für Zylinder $Z = \pi^{-1}_{1,...,n}(C_n)$, $C_n \in \otimes_i \mathcal{A}_i$ gilt $P[Z] = P_n(Cn)$.
Wenn diese Wahrscheinlichkeitsverteilung verträglich sind, d.h. die Randverteilungen (bzw. alle Teilmengen 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 ($P_j = P_k \pi^{-1}_{k,j}, J \subset K, K$ 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)**.** *$(\Omega_i, \mathcal{A}_i, \mu_i), i = 1, 2$ seinen sigma-endliche Maßräume. Wenn die messbaren Funktionen $f_i : (\Omega_i, \mathcal{A} : i) \to (\mathbb{R}, \mathcal{B})$ 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 $(f_1 \cdot f_2)_{x_1} = f_1(x_1) \cdot f_2(.)\ \mathcal{A}_2$ messbar ist und nach Fubini

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

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. □

Die Betrachtung unabhängiger SGn $X_i, i = 1, ..., k$ 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 $X = (X_1, ..., X_k)$ ein Vektor unabhängiger SGn X_i ist, gilt

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

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

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

und die Randverteilung von X_i ist mit Dichte

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

und wieder erhält man die Randdichte

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

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

Wenn die X_i unabhängig sind ist die gemeinsame Dichte

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

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 gemainsame Dichte zerfällt (siehe EI24)

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

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

$$\mathbb{E}XY = \int\int xyf(x, y)d\lambda_2 = \int \underbrace{\int yf_2(x, y)dy}_{m} \, xf_1(x)dx = \int mxf_1(x)dx = \rho \int x^2 f_1(x)dx = \rho\frac{\sigma_y}{\sigma_x}\sigma_x^2 = \rho\sigma_x\sigma_y$$

X,Y sind unkorreliert $\iff$ $\rho = 0$, dann ist $f(x, y) = f_1(x) \cdot \Phi(\frac{y}{\sigma_y})$ (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 seinen Lebesgue-Stieltes 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

$$\lim_{\triangle\to 0}\frac{F(x+\triangle)-F(x)}{\triangle}=f_X(x) \text{ oder}$$
$$\frac{P[X\in[x-\triangle,x]]}{\triangle}\to f_X(x) \text{ für } \triangle\to 0 \text{ oder}$$
$$P[X\in[x-\triangle,x]]\sim f_X(x)\triangle$$

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

$$P[X\in[x-\triangle,x]|Y\in[y-\triangle,y]]=\frac{P[X\in[x-\triangle,x],Y\in[y-\triangle,y]]}{P[Y\in[y-\triangle,y]]}=$$
$$\frac{\int_{x-\triangle}^{x}\int_{y-\triangle}^{y}f_{X,Y}(s,t)dsdt}{\int_{[y-\triangle,y]}f_Y(t)d\lambda(t)}\sim\frac{f_{X,Y}(x,y)\triangle^2}{f_Y(y)\triangle}=\frac{f_{X,Y}(x,y)\triangle}{f_Y(y)}$$

Definition 3 (bedingte Dichte)**.** *$X:\Omega\to\mathbb{R}$ und $Y:\Omega\to\mathbb{R}$ 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 $f(.|y)$ eine Wahrscheinlichkeitsdichte und tatsächlich eine Verteilung festgelegt.

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

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

bedingter Erwartungswert. Dann ist $h(Y):=\mathbb{E}[X|Y]$ eine PY^{-1}-f.s. messbare Funktion. $h(.)$ 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*

$$f(x|y)=\frac{1}{\sqrt{2\pi v}}exp\left(-\frac{1}{2}\left(\frac{x-m}{v}\right)^2\right)$$
$$\textit{mit } m=\mu_x+\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y) \textit{ und } v^2=\sigma_x^2(1-\rho^2)$$

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:

$$\mathbb{E}H(Y) = \int \mathbb{E}[X|Y=y]dPY^{-1} = \int\int xf(x|y)dxf_Y(y)dy$$
$$= \int\int x \underbrace{f(x|y)f_Y(y)}_{f(x,y)} dydx = \int\int xf(x,y)dydx = \int xf_X(x)dx = \mathbb{E}X.$$

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

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

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

$$(\mu_1 * \mu_2) * \mu_3(A) = \int \mu_1 * \mu_2(A-z)d\mu_3(z) = \int \left(\int \mu_2(A-z-x)d\mu_1(x)\right) d\mu_3 =$$
$$\int\int \mu_2(A-z-x)d\mu_3(z)d\mu_1 = \int \mu_2 * \mu_3(A-x)d\mu_1(x) = (\mu_2 * \mu_3) * \mu_1(A) = \mu_1 * (\mu_2 * \mu_3)(A)$$

  □

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

  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 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. TODO □

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. TODO □

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)**.** *TODO*