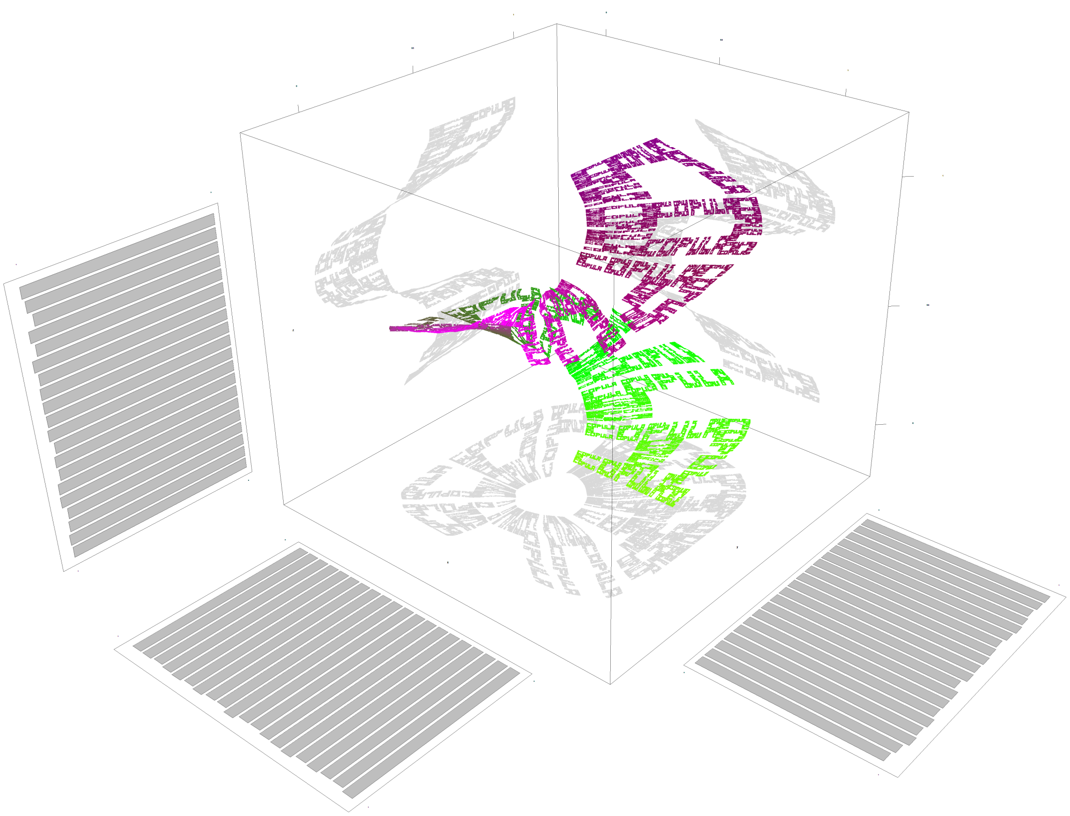
Rohversion Skriptum zur Vorlesung

Abhängigkeitsmodellierung

LV-Nr. 405.552

Ass.-Prof. Dr. Wolfgang Trutschnig
www.trutschnig.net
Fachbereich Mathematik
Paris Lodron Universität Salzburg
Hellbrunner Strasse 34
A-5020 Salzburg



Inhaltsverzeichnis

Kapitel 1

Ergänzungen zur Maß/Wahrscheinlichkeitstheorie

Um die größtenteils maßtheoretisch formulierten Resultate in [2] sauber verstehen und beweisen zu können, bedarf es insbesondere zweier Hauptheoreme der Maßtheorie/Wahrscheinlichkeitstheorie, die in den Grundvorlesungen leider oft zu kurz kommen.

1.1 Lebesgue-Zerlegung und der Satz von Radon-Nikodym

Eine sehr spezielle Version der Lebesgue-Zerlegung und des Satzes von Radon-Nikodym kennen Sie aus der Statistik Bachelor Vorlesung: Jede eindimensionale Verteilungsfunktion lässt sich als Konvexkombination einer absolut stetigen, einer diskreten und einer singulären Verteilungsfunktion darstellen. Im Folgenden betrachten wir diese Zerlegungseigeschaft auf einem allgemeinen Messraum $(\Omega, \mathcal{A})$ ohne jegliche topologische Struktur. Mit $\mathcal{M}(\Omega, \mathcal{A})$ bezeichnen wir (im gesamten Skriptum) die Menge aller endlichen Maße auf $\mathcal{A}$. Die Resultate in diesem Kapitel gelten großteils auch für σ-endliche positive und für komplexe Maße - nachdem der Fokus der LV auf Copulas, und damit auf speziellen Wahrscheinlichkeitsmaßen liegt, reicht es, mit $\mathcal{M}(\Omega, \mathcal{A})$ zu arbeiten.

Definition 1.1 *Gegeben seien zwei endliche Maße $\nu, \mu \in \mathcal{M}(\Omega, \mathcal{A})$. Dann heiß ν absolut stetig bezüglich μ (wir schreiben $\nu \ll \mu$) genau dann, wenn für jedes $E \in \mathcal{A}$ aus $\mu(E) = 0$ auch $\nu(E) = 0$ folgt.*

Definition 1.2 *Sei $\mu \in \mathcal{M}(\Omega, \mathcal{A})$. Falls eine Menge $E \in \mathcal{A}$ existiert, sodass für jedes $F \in \mathcal{A}$ die Gleichheit*

$$\mu(F) = \mu(E \cap F)$$

gilt, dann sagen wir, dass μ auf E konzentriert ist (oft auch: μ 'lebt' auf E).

Bemerkung 1.3 Gelte $\mu(E) = 0$, dann ist μ offensichtlich auf E^c konzentriert. Zusätzlich gilt: μ ist auf E konzentriert genau dann wenn $\mu(E) = \mu(\Omega) =: \|\mu\|$.

Definition 1.4 *Zwei Maße $\nu, \mu \in \mathcal{M}(\Omega, \mathcal{A})$ heißen singulär zueinander (wir schreiben $\nu \perp \mu$), falls zwei disjunkte Mengen $E_1, E_2 \in \mathcal{A}$ existieren, sodass ν auf E_1 und μ auf E_2 konzentriert ist.*

Bemerkung 1.5 Offensichtlich gilt $\nu \perp \mu$ genau dann wenn eine Menge $E \in \mathcal{A}$ existiert, für die $\mu(E) = \|\mu\|$ (also $\mu(E^c) = 0$) und $\nu(E) = 0$ gilt.

Das folgende Lemma fasst einige grundlegende Eigenschaften der gerade eingeführten Begriffe zusammen:

Lemma 1.6 *Für $\nu, \nu_1, \nu_2, \mu \in \mathcal{M}(\Omega, \mathcal{A})$ gelten folgende Aussagen:*

1. *Aus $\nu_1 \ll \mu$ und $\nu_2 \ll \mu$ folgt $\nu_1 + \nu_2 \ll \mu$.*
2. *Aus $\nu_1 \perp \mu$ und $\nu_2 \perp \mu$ folgt $\nu_1 + \nu_2 \perp \mu$.*
3. *Aus $\nu_1 \ll \mu$ und $\nu_2 \perp \mu$ folgt $\nu_1 \perp \nu_2$.*
4. *Aus $\nu \ll \mu$ und $\nu \perp \mu$ folgt $\nu = 0$ (also $\nu(E) = 0$ für jedes $E \in \mathcal{A}$).*

Beweis: Einfache Übungsaufgabe

Satz 1.7 (Lebesgue Zerlegung und Satz von Radon-Nikodym für endliche Maße) *Seien $\nu, \mu \in \mathcal{M}(\Omega, \mathcal{A}) =: \mathcal{M}$, dann gelten folgende Aussagen:*

1. *Es existiert ein eindeutiges Paar $(\nu_a, \nu_s) \in \mathcal{M}^2$ mit*
$$\nu = \nu_a + \nu_s, \quad \nu_a \ll \mu, \quad \nu_s \perp \mu \tag{1.1}$$
2. *Es existiert ein eindeutiges $f \in L^1(\Omega, \mathcal{A}, \mu)$ mit*
$$\nu_a(E) = \int_E f d\mu \tag{1.2}$$
für jedes $E \in \mathcal{A}$. Dieses f heißt Radon-Nikodym Ableitung/Dichte von ν_a bezüglich μ.

Beweis: Wir definieren ein neues Maß $\varphi \in \mathcal{M}$ durch
$$\varphi(E) := \nu(E) + \mu(E)$$
für jedes $E \in \mathcal{A}$. Für jedes $h \in L^2(\Omega, \mathcal{A}, \varphi)$ gilt offensichtlich
$$\Big| \int_\Omega h d\nu \Big| \le \int_\Omega |h| d\nu \le \int_\Omega |h| d\varphi \le \underbrace{\left(\int_\Omega h^2 d\varphi \right)^{1/2}}_{=:\|h\|_{L^2(\Omega,\mathcal{A},\varphi)}} \big(\varphi(\Omega)\big)^{1/2} < \infty. \tag{1.3}$$

Die Abbildung $T : L^2(\Omega, \mathcal{A}, \varphi) \to \mathbb{R}$, definiert durch $T(h) = \int_\Omega h d\nu$, ist wohldefiniert und linear. Für $h_1, h_2 \in L^2(\Omega, \mathcal{A}, \varphi)$ gilt wegen (1.3)
$$|T(h_1) - T(h_2)| \le \|h_1 - h_2\|_{L^2(\Omega,\mathcal{A},\varphi)} \big(\varphi(\Omega)\big)^{1/2},$$
T ist also (Lipschitz-)stetig, und damit insgesamt ein stetiges lineares Funktional auf dem Hilbertraum $L^2(\Omega, \mathcal{A}, \varphi)$. Anwendung des Darstellungssatzes von Riesz liefert die Existenz eines eindeutigen $g \in L^2(\Omega, \mathcal{A}, \varphi)$ mit
$$\int_\Omega h \, d\nu = T(h) = \langle h, g \rangle_{L^2(\Omega,\mathcal{A},\varphi)} = \int_\Omega hg \, d\varphi \tag{1.4}$$

für jedes $h \in L^2(\Omega, \mathcal{A}, \varphi)$. Setzen wir $h = \mathbf{1}_E$ für ein $E \in \mathcal{A}$ mit $\varphi(E) > 0$, dann folgt via (1.4) sofort

$$\nu(E) = \int_\Omega \mathbf{1}_E \, d\nu = \int_\Omega \mathbf{1}_E \, g \, d\varphi = \int_E g \, d\varphi,$$

und damit

$$0 \leq \frac{1}{\varphi(E)} \int_E g \, d\varphi = \frac{\nu(E)}{\varphi(E)} \leq 1.$$

Nachdem $E \in \mathcal{A}$ mit $\varphi(E) > 0$ beliebig war, erhalten wir $g(x) \in [0,1]$ für φ-fast jedes $x \in \Omega$. O.B.d.A. können wir annehmen (falls notwendig, ändern wir g auf einer φ-Nullmenge), dass $g(x) \in [0,1]$ sogar für jedes $x \in \Omega$ gilt.
Unter Verwendung der Definition von φ kann Gleichung (1.4) umgeformt werden zu

$$\int_\Omega h(1-g) \, d\nu = \int_\Omega hg \, d\mu. \tag{1.5}$$

Wir setzen wir $A = g^{-1}([0,1)) \in \mathcal{A}$ (also $A^c = g^{-1}(\{1\})$) und definieren $\nu_a, \nu_s \in \mathcal{M}$ durch

$$\nu_a(E) = \nu(A \cap E), \quad \nu_s(E) = \nu(A^c \cap E). \tag{1.6}$$

Das Maß ν_a ist konzentriert auf A, ν_s auf A^c, und offensichtlich gilt $\nu = \nu_a + \nu_s$. Für $h = \mathbf{1}_{A^c}$ liefert (1.5)

$$0 = \int_{A^c} 1 - g \, d\nu = \int_{A^c} g \, d\mu = \mu(A^c),$$

es gilt also $\nu_s \perp \mu$.
Um $\nu_a \ll \mu$ und die Darstellung (1.2) zu beweisen, gehen wir wiefolgt vor: Für beliebiges $E \in \mathcal{A}, n \in \mathbb{N}$ und $h = (1 + g + g^2 + \ldots + g^n) \, \mathbf{1}_E$ liefert (1.5)

$$L_n := \int_E (1 - g^{n+1}) d\nu = \int_E \underbrace{g(1 + g + g^2 + \ldots + g^n)}_{=:I_i} \, d\mu =: R_n. \tag{1.7}$$

Für $x \in E \cap A$ gilt $1 - g^{n+1}(x) \to 1$ für $n \to \infty$, für $x \in E \cap A^c$ gilt offensichtlich $1 - g^{n+1}(x) = 0$, insgesamt erhalten wir also (Satz von der majorisierten Konvergenz oder Satz von der monotonen Konvergenz) $\lim_{n\to\infty} L_n = \nu(E \cap A) = \nu_a(E)$. Der Integrand I_n der rechten Seite ist monoton nicht fallend in n und konvergiert (in $[0,\infty]$) punktweise gegen eine nicht-negative messbare Funktion $f : \Omega \to [0,\infty]$ (beachten Sie, dass wir f sogar explizit berechnen können - wie?). Anwendung des Satzes von der monotonen Konvergenz liefert $\lim_{n\to\infty} R_n = \int_E f \, d\mu$, und wir erhalten insgesamt

$$\infty > \nu_a(E) = \int_E f \, d\mu,$$

also genau die geünschte Darstellung für ν_a ($E \in \mathcal{A}$ war beliebig). Für den Fall $E = \Omega$ folgt wegen $\int_\Omega |f| \, d\mu = \int_\Omega f \, d\mu = \nu_a(\Omega) < \infty$ die Integrierbarkeit von f. Damit ist der Beweis bis auf die in der ersten Aussage behauptete Eindeutigkeit des Paares $(\nu_a, \nu_s) \in \mathcal{M}^2$ (Übungsaufgabe) komplett. ■

Bemerkung 1.8 Für $\nu \ll \mu$ schreiben wir in der Folge auch oft $f = \frac{d\nu}{d\mu}$ für die Radon-Nikodym Dichte/Ableitung von ν bezüglich μ, und verwenden zusätzlich die in der Literatur übliche Schreibweise $d\nu = f d\mu$.

Satz 1.9 (Rechenregeln für Radon-Nikodym Ableitungen/Dichten)
Seien $\mu_1, \mu_2, \mu_3 \in \mathcal{M}(\Omega, \mathcal{A})$ und es gelte $\mu_1 \ll \mu_2 \ll \mu_3$. Dann gilt die sogenannte Kettenregel für Radon-Nikodym Ableitungen, i.e.

$$\frac{d\mu_1}{d\mu_3} = \frac{d\mu_1}{d\mu_2} \cdot \frac{d\mu_2}{d\mu_3} \quad [\mu_3] \tag{1.8}$$

Falls $\mu_1 \ll \mu_2$ und $\mu_2 \ll \mu_1$ gilt, dann folgt

$$\frac{d\mu_2}{d\mu_1} = \left(\frac{d\mu_1}{d\mu_2}\right)^{-1} = \frac{1}{\frac{d\mu_1}{d\mu_2}} \; [\mu_1]. \tag{1.9}$$

Beweis: Übungsaufgabe

Der folgende abschießende Satz begründet die Verwendung des Wortes 'Stetig' im Zusammenhang mit $\nu \ll \mu$.

Satz 1.10 *Für $\nu, \mu \in \mathcal{M}(\Omega, \mathcal{A})$ sind die folgenden beiden Aussagen äquivalent:*

1. $\nu \ll \mu$
2. *Für jedes $\varepsilon > 0$ existiert ein $\delta > 0$ sodass $\nu(E) < \varepsilon$ für jedes $E \in \mathcal{A}$ mit $\mu(E) < \delta$.*

Beweis: Übungsaufgabe

1.2 Bedingte Erwartungen

Wir betrachten ein motivierendes (nicht-diskretes) Beispiel, in der die Berechnung des bedingten Erwartungswerts zumindest intuitiv klar ist.

Beispiel 1.11 Der Zufallsvektor (X, Y) sei absolut stetig (bezüglich des Lebesgue-Maßes λ_2 auf $\mathcal{B}(\mathbb{R}^2)$) mit Dichte

$$f(x, y) = \tfrac{1}{\pi}\mathbf{1}_{B(0,1)}(x, y),$$

wobei $B(0, 1) = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}$ die offene Einheitskugel bezeichnet. Wie würden Sie $\mathbb{P}(Y \in B|Y = x)$ sowie $\mathbb{E}(Y|X = x)$ für $B \in \mathcal{B}(\mathbb{R})$ und $x \in \mathbb{R}$ definieren? Intuitiv würde man zweifelsohne

$$Y|(X = x) \sim \mathcal{U}(-\sqrt{1 - x^2}, \sqrt{1 - x^2})$$

für jedes $x \in [-1, 1]$ setzen und damit $\mathbb{E}(Y|X = x) = 0$ erhalten. Intuitiv ebenfalls naheliegend wäre, dass

$$\mathbb{P}((X, Y) \in E \times F) = \mathbb{P}^{(X,Y)}(E \times F) = \int_E \mathbb{P}(Y \in F|X = x)\, d\mathbb{P}^X(x)$$

für jedes $G \in \mathcal{B}(\mathbb{R}^2)$ sowie

$$\mathbb{E}(Y) = \int_{[-1,1]} \mathbb{E}(Y|X = x)\, d\mathbb{P}^X(x)$$

gilt. Tatsächlich lassen sich die letzten beiden Identitäten auch formal beweisen - beides sind Spezialfälle der sogenannten Disintegration, die im nächsten Kapitel behandelt wird.

Das nachfolgende Beispiel deutet an, in welche Richtung es gehen wird.

Beispiel 1.12 Wir betrachten den Wurf mit einem fairen Würfel, modelliert durch $(\Omega, \mathcal{A}, \mathbb{P}) = (\{1, 2, \dots, 6\}, p(\{1, 2, \dots, 6\}), \mathbb{P})$ mit $\mathbb{P}(\{i\}) = \frac{1}{6}$ für jedes $i \in \Omega$ sowie die Ereignisse $A = \{1, 2, 3\}, B = \{1, 3, 5\}$. Die bedingte Wahrscheinlichkeit von A gegeben B berechnet sich bekanntlich zu

$$\mathbb{P}_B(A) := \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}.$$

Für jede andere Teilmenge $E \in \mathcal{A}$ kann $\mathbb{P}_B(E)$ analog berechnet werden. Sei nun X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Wie kann der Erwartungswert $\mathbb{E}(X|B)$ von X unter B berechnet werden? Naheliegenderweise ist $\mathbb{E}(X|B)$ definiert als der Erwartungswert von X unter dem (neuen) Wahrscheinlichkeitsmaß $\mathbb{P}_B$. Wir erhalten also

$$\begin{aligned}
\mathbb{E}(X|B) &= \int_\Omega X d\mathbb{P}_B = \sum_{\omega=1}^{6} X(\omega)\mathbb{P}_B(\{\omega\}) = \sum_{\omega \in B} X(\omega) \frac{\mathbb{P}(\{\omega\})}{\mathbb{P}(B)} \\
&= \frac{1}{\mathbb{P}(B)} \sum_{\omega \in \Omega} X(\omega) \mathbf{1}_B(\omega) \mathbb{P}(\{\omega\}) \\
&= \frac{1}{\mathbb{P}(B)} \mathbb{E}(X \mathbf{1}_B) = \frac{(X(1) + X(3) + X(5)) \frac{1}{6}}{\frac{1}{2}} = \frac{X(1) + X(3) + X(5)}{3}
\end{aligned}$$

Für B^c erhalten wir analog

$$\mathbb{E}(X|B^c) = \frac{1}{\mathbb{P}(B^c)} \mathbb{E}(X \mathbf{1}_{B^c}).$$

Setzen wir $\mathcal{B} := \mathcal{A}_\sigma(\{B, B^c\}) = \{\emptyset, \Omega, B, B^c\} \subseteq \mathcal{A}$ und definieren wir $\mathbb{E}(X|\mathcal{B}) : \Omega \to \mathbb{R}$ durch

$$\mathbf{E}(X|\mathcal{B})(\omega) = \begin{cases} \mathbb{E}(X|B) & \text{für } \omega \in B, \\ \mathbb{E}(X|B^c) & \text{für } \omega \in B^c, \end{cases} \tag{1.10}$$

dann hat $\mathbf{E}(X|\mathcal{B})$ die folgenden Eigenschaften:

1. $\mathbb{E}(X|\mathcal{B})$ ist $\mathcal{B}$-messbar (und ein Element von $L^1(\Omega, \mathcal{B}, \mathbb{P})$).

2. Für jedes $C \in \mathcal{B}$ gilt
$$\int_C \mathbb{E}(X|\mathcal{B}) d\,\mathbb{P} = \int_C X d\,\mathbb{P}. \tag{1.11}$$
Tatsächlich liefert die Wahl von $C = B$ sofort
$$\begin{aligned}
\int_B \mathbb{E}(X|\mathcal{B}) d\,\mathbb{P} &= \int_B \mathbb{E}(X|B) d\,\mathbb{P} = \int_B \frac{\mathbb{E}(X\,\mathbf{1}_B)}{\mathbb{P}(B)} d\,\mathbb{P} = \frac{\mathbb{E}(X\,\mathbf{1}_B)}{\mathbb{P}(B)} \int_B \mathbf{1} d\mathbb{P} \\
&= \int_B X d\mathbb{P},
\end{aligned}$$
die Gleichheit für $C = B^c$ und $C = \Omega$ folgt analog, und für $C = \emptyset$ ist die Gleichheit trivial. $\mathbb{E}(X|\mathcal{B})$ kann also interpretiert werden als geglättete (aggregierte) Version von X.

Im obigen Beispiel hätten wir jeden beliebigen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und eine Menge $B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ nehmen können - die Konstruktion der *bedingten Erwartung* $\mathbb{E}(X|\mathcal{B})$ wäre gleich verlaufen.

Ist nun I eine abzählbare Menge, $(B_i)_{i\in\mathbb{N}}$ eine messbare Partition von Ω bestehend aus Mengen mit echt positiver Wahrscheinlichkeit und X eine integierbare Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, dann kann $\mathbb{E}(X|\mathcal{B})$ mit $\mathcal{B} = \mathcal{A}_\sigma(\{B_1, B_2, \ldots\})$ naheliegenderweise definiert werden durch

$$\mathbb{E}(X|\mathcal{B})(\omega) = \sum_{i\in I} \mathbb{E}(X|B_i)\mathbf{1}_{B_i}(\omega) = \sum_{i\in I} \frac{\mathbb{E}(X\mathbf{1}_{B_i})}{\mathbb{P}(B_i)}\mathbf{1}_{B_i}(\omega). \tag{1.12}$$

Das nachfolgende Lemma bestätigt, dass die so erhaltene Zufallsvariable $\mathbb{E}(X|\mathcal{B}) : \Omega \to \mathbb{R}$ dieselben Eigeschaften hat wie $\mathbb{E}(X|\mathcal{B})$ im vorigen Beispiel:

Lemma 1.13 *Die Zufallsvariable $\mathbb{E}(X|\mathcal{B})$ gemäß (1.12) ist diskret und hat folgende Eigenschaften:*

1. *$\mathbb{E}(X|\mathcal{B})$ ist $\mathcal{B}$-messbar (i.e. $\mathbb{E}(X|\mathcal{B})^{-1}(E) \in \mathcal{B}$ für jede Borel-Menge $E \in \mathcal{B}(\mathbb{R})$).*
2. *$\mathbb{E}(X|\mathcal{B})$ ist integrierbar und erfüllt*

$$\int_C \mathbb{E}(X|\mathcal{B})d\,\mathbb{P} = \int_C X d\,\mathbb{P} \tag{1.13}$$

für jedes $C \in \mathcal{B}$.

Beweis: Übungsaufgabe

Der finale Schritt von der abzählbar erzeugten σ-Algebra $\mathcal{B}$ auf eine beliebige sub-σ-Algebra $\mathcal{C} \subseteq \mathcal{A}$ ist nun wenig überraschend:

Definition 1.14 *Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$. Dann heißt eine Zufallsvariable X_0 (eine Version der)* bedingte Erwartung von X unter $\mathcal{C}$ *(wir schreiben $X_0 = \mathbb{E}(X|\mathcal{C})$) genau dann, wenn sie die folgenden zwei Eigenschaften hat:*

(CE1) X_0 ist $\mathcal{C}$-messbar.

(CE2) Für jedes $C \in \mathcal{C}$ gilt die Gleichheit

$$\int_C X_0\, d\,\mathbb{P} = \int_C X d\,\mathbb{P}.$$

Satz 1.15 *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$ und X eine integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann existiert $\mathbb{E}(X|\mathcal{C})$ und ist eindeutig $\mathbb{P}$-fast überall.*

Beweis: Zum Beweise der Existenz folgen wir, wie gewohnt, dem Integralaufbau und starten mit einer nicht-negativen Zufallsvariable X. Das Maß μ, definiert durch $\mu(C) = \int_C X d\mathbb{P}$ ist (wegen der vorausgesetzten Integrierbarkeit von X) endlich, und wird offensichtlich von $\mathbb{P}$ dominiert. Nach dem Satz von Radon-Nikodym existiert daher eine $\mathcal{C}$-messbare, integrierbare ($\mathbb{P}$-fast überall eindeutig bestimmte) Zufallsvariable X_0 mit

$$\int_C X_0 d\mathbb{P} = \mu(C) = \int_C X d\mathbb{P}$$

für alle $C \in \mathcal{C}$. X_0 hat also die gewünschten Eigenschaften und der Satz ist für nicht-negatives X bewiesen.
Für allgemeines (integrierbares X) verwenden wir die Zerlegung $X = X^+ - X^-$, bestimmen die entsprechenden bedingten Erwartungen X_0^+, X_0^- und setzen einfach $X_0 = X_0^+ - X_0^-$.
Zum Beweis der behaupteten Eindeutigkeit nehmen wir an, X_0 und X_0' erfüllen (CE1) und (CE2), und setzen $E := \{w \in \Omega : X_0(\omega) > X_0'(\omega)\} \in \mathcal{C}$. Eigenschaft (CE2) liefert

$$0 = \int_E X_0 - X_0' \, d\mathbb{P} = \int_\Omega \underbrace{\mathbf{1}_E \, (X_0 - X_0')}_{\geq 0} \, d\mathbb{P},$$

woraus sofort $\mathbb{P}(E) = 0$, und damit $X_0 \leq X_0'$ $[\mathbb{P}]$ folgt. Nachdem sich $X_0' \leq X_0$ $[\mathbb{P}]$ analog ergibt, folgt insgesamt $\mathbb{P}(\{\omega \in \Omega : X_0(\omega) = X_0'(\omega)\}) = 1$. ■

Einige der wichtigsten Eigenschaften bedingter Erwartungen sind im folgenden Satz zusammengefasst.

Satz 1.16 *Seien X, Y integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{D} \subseteq \mathcal{C}$ sub-σ-Algebren von $\mathcal{A}$. Dann gilt (alle Aussagen sind $[\mathbb{P}]$ zu interpretieren):*

1. $\mathbb{E}(aX + bY|\mathcal{C}) = a\mathbb{E}(X|\mathcal{C}) + b\mathbb{E}(Y|\mathcal{C})$ *(Linearität)*
2. $Y \leq X$ *impliziert* $\mathbb{E}(Y|\mathcal{C}) \leq \mathbb{E}(X|\mathcal{C})$ *(Monotonie)*
3. *Ist X sogar $\mathcal{C}$-messbar, dann gilt* $\mathbb{E}(X|\mathcal{C}) = X$
4. *Falls* $\mathbb{E}(|XY|) < \infty$ *und Y sogar $\mathcal{C}$-messbar ist, dann folgt* $\mathbb{E}(XY|\mathcal{C}) = Y\mathbb{E}(X|\mathcal{C})$.
5. $\mathbb{E}\big(\mathbb{E}(X|\mathcal{C})|\mathcal{D}\big) = \mathbb{E}\big(\mathbb{E}(X|\mathcal{D})|\mathcal{C}\big) = \mathbb{E}(X|\mathcal{D})$ *(Turmeigeschaft)*
6. $\mathbb{E}(|X|\big|\mathcal{C}) \geq \big|\mathbb{E}(X|\mathcal{C})\big|$ *(Dreiecksungleichung)*
7. *Falls $\mathcal{A}_\sigma(X)$ und $\mathcal{C}$ unabhängig sind, gilt* $\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X)$
8. *Fall $\mathcal{C}$ nur aus 0- oder 1-Mengen besteht folgt* $\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X)$
9. *Für eine Folge $(X_n)_{n\in\mathbb{N}}$ von Zufallsvariablen, die $\mathbb{P}$-fast sicher gegen X konvergiert und $|X_n| \leq Y$ $[\mathbb{P}]$ erfüllt, folgt*
$$\lim_{n\to\infty} \mathbb{E}(X_n|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}) \, [\mathbb{P}]$$
und
$$\lim_{n\to\infty} \int_\Omega \big|\mathbb{E}(X_n|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\big| d\mathbb{P} = 0.$$

Beweis: Linearität folgt sofort aus der Tatsache, dass $a\mathbb{E}(X|\mathcal{C}) + b\mathbb{E}(Y|\mathcal{C})$ $\mathcal{C}$-messbar ist und dass für jedes $C \in \mathcal{C}$

$$\begin{aligned}\int_C a\mathbb{E}(X|\mathcal{C}) + b\mathbb{E}(Y|\mathcal{C}) \, d\mathbb{P} &= a\int_C \mathbb{E}(X|\mathcal{C}) d\mathbb{P} + b\int_C \mathbb{E}(Y|\mathcal{C}) \, d\mathbb{P} = a\int_C X d\mathbb{P} + b\int_C Y \, d\mathbb{P} \\ &= \int_C aX + bY \, d\mathbb{P}\end{aligned}$$

gilt.
Zum Beweis der Monotonie setzen wir $C := \{\omega \in \Omega : \mathbb{E}(Y|\mathcal{C}) > \mathbb{E}(X|\mathcal{C})\} \in \mathcal{C}$ und erhalten (unter Verwendung der Linearität und (CE2))

$$\int_\Omega \underbrace{\mathbf{1}_C\, \mathbb{E}(Y|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})}_{\geq 0}\, d\mathbb{P} \quad = \quad \int_C \mathbb{E}(Y|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})\, d\mathbb{P} = \int_C Y - X d\mathbb{P} \leq 0,$$

und damit $\mathbb{P}(C) = 0$. Der Beweis der restlichen Aussagen ist eine Übungsaufgabe.

Satz 1.16 besagt insbesondere, dass für jede integrierbare Zufallsvariable X die Eigenschaft $\mathbb{E}\big(\mathbb{E}(X|\mathcal{C})|\mathcal{C}\big) = \mathbb{E}(X|\mathcal{C})$ gilt, die Abbildung $X \mapsto \mathbb{E}(X|\mathcal{C})$ also idempotent ist. Idempotenz kennen wir insbesondere von orthogonalen Projektionen (i.e. Bestapproximierende in Unterräumen). Zusätzlich wurde vorhin schon erwähnt wurde, dass $\mathbb{E}(X|\mathcal{C})$ als geglättete Version von X gesehen werden kann. Mit dieser Interpretation im Hinterkopf ist das folgende Resultat nicht mehr allzu überraschend:

Satz 1.17 *Seien X quadratisch integrierbar und $\mathcal{C}$ eine sub-σ-Algebren von $\mathcal{A}$. Dann gilt für jede quadratisch integrierbare, $\mathcal{C}$-messbare Zufallsvariable Y die Ungleichung*

$$\mathbb{E}\big((X-Y)^2\big) \geq \mathbb{E}\big((X - \mathbb{E}(X|\mathcal{C}))^2\big) \tag{1.14}$$

mit Gleichheit, dann und nur dann, wenn $Y = \mathbb{E}(X|\mathcal{C})\,[\mathbb{P}]$.
Mit anderen Worten: $\mathbb{E}(X|\mathcal{C})$ ist die orthogonale Projektion von $X \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ auf den (abgeschlossenen) Unterraum $L^2(\Omega, \mathcal{C}, \mathbb{P})$.

Beweis: Nachdem aus der quadratischen Integrierbarkeit von X die quadratische Integrierbarkeit von $\mathbb{E}(X|\mathcal{C})^2$ folgt (Übungsaufgabe), reicht es, Ungleichung (1.14) zu beweisen.
Sei nun Y $\mathcal{C}$-messbar und quadratisch integrierbar. Anwendung der Cauchy-Schwarz Ungleichung liefert sofort $\mathbb{E}(|XY|) < \infty$, weiters können lt. Punkt 4. in Satz 1.16 $\mathbb{E}(XY)$ und $\mathbb{E}(X\mathbb{E}(X|\mathcal{C}))$ umgeschrieben werden zu

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}\Big(\mathbb{E}(XY|\mathcal{C})\Big) \overset{4.}{=} \mathbb{E}\big(Y\mathbb{E}(X|\mathcal{C})\big) \\
\mathbb{E}\big(X\mathbb{E}(X|\mathcal{C})\big) &= \mathbb{E}\Big(\mathbb{E}\big(X\mathbb{E}(X|\mathcal{C})\big|\mathcal{C}\big)\Big) \overset{4.}{=} \mathbb{E}\big(\mathbb{E}(X|\mathcal{C})^2\big).
\end{aligned}$$

Unter Verwendung dieser Gleichheiten folgt die gewünschte Ungleichung direkt aus

$$\begin{aligned}
\mathbb{E}(X-Y)^2 - \mathbb{E}\big((X - \mathbb{E}(X|\mathcal{C}))^2\big) &= \mathbb{E}\Big(X^2 - 2XY + Y^2 - X^2 + 2X\mathbb{E}(X|\mathcal{C}) - \mathbb{E}(X|\mathcal{C})^2\Big) \\
&= \mathbb{E}(Y^2) - 2\mathbb{E}(XY) + 2\underbrace{\mathbb{E}\big(X\mathbb{E}(X|\mathcal{C})\big)}_{=\mathbb{E}\left(\mathbb{E}\left(X\mathbb{E}(X|\mathcal{C})\middle|\mathcal{C}\right)\right)} - \mathbb{E}\big(\mathbb{E}(X|\mathcal{C})^2\big) \\
&= \mathbb{E}(Y^2) - 2\mathbb{E}\big(Y\mathbb{E}(X|\mathcal{C})\big) + \mathbb{E}\big(\mathbb{E}(X|\mathcal{C})^2\big) \\
&= \mathbb{E}\big(Y - \mathbb{E}(X|\mathcal{C})\big)^2 \geq 0. \quad \blacksquare
\end{aligned}$$

Die von der Zufallsvariable X erzeugte σ-Algebra $\mathcal{C} = \mathcal{A}_\sigma(X) = X^{-1}(\mathcal{B}(\mathbb{R}))$ ist (aufgrund der Messbarkeit von X) offensichtlich eine sub-σ-Algebra von $\mathcal{A}$. Für eine integrierbare Zufallsvariable Y auf $(\Omega, \mathcal{A}, \mathbb{P})$ und eine weitere (nicht notwendigerweise integrierbare) Zufallsvariable X nennen wir daher

$$\mathbb{E}(Y|X) := \mathbb{E}(Y|\mathcal{A}_\sigma(X)), \tag{1.15}$$

bedingte Erwartung von Y unter X.

Das folgende Hilfsresultat impliziert, dass $\mathbb{E}(Y|X)$ sich immer in der Form $\mathbb{E}(Y|X) = h\circ X$ mit einer (Borel-) messbaren Funktion $h : \mathbb{R} \to \mathbb{R}$ darstellen lässt.

Lemma 1.18 (Faktorisierungslemma) *Seien X und Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann sind die folgenden beiden Bedingungen äquivalent:*

1. *Y ist $\mathcal{A}_\sigma(X)$ messbar.*
2. *Es existiert eine (Borel) messbare Transformation $h : \mathbb{R} \to \mathbb{R}$ mit $Y = h \circ X$.*

Zusatz: Ist Y nicht-negativ, dann kann auch h nicht-negativ gewählt werden. Weiters ist h eindeutig bestimmt $\mathbb{P}^X$-fast überall.

Beweis: Dem Integralaufbau folgend eine einfach Übungsaufgabe.

1.3 Disintegration und Markov Kerne

Wir wissen aus dem vorigen Abschnitt, dass sich für (reelle) Zufallsvariable X, Y auf $(\Omega, \mathcal{A}, \mathbb{P})$ die bedingte Erwartung $\mathbb{E}(Y|X)$ von Y unter X in der Form $\mathbb{E}(Y|X) = h \circ X$ darstellen lässt, wobei $h : \mathbb{R} \to \mathbb{R}$ eine messbare (und i.A. nicht eindeutige) Funktion ist. Wie am Beginn von Abschnitt 1.2 skizziert, ist unser Hauptziel, allgemeine bedingte Verteilungen (i.e. die Verteilung von Y gegeben $X = x$ für $x \in \mathbb{R}$ und beliebige, nicht notwendigerweise diskrete Zufallsvariable X) zu konstruieren. Die folgenden Definition ist wenig überraschend:

Definition 1.19 *Seien X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$, zusätzlich sei Y integrierbar. Dann ist EINE Version der* bedingte Erwartung $\mathbb{E}(Y|X = x)$ von Y gegeben $X = x$ *gegeben durch*

$$\mathbb{E}(Y|X = x) = h(x), \tag{1.16}$$

wobei h EINE lt. Faktorisierungslemma existierende Funktion mit $\mathbb{E}(Y|X) = h \circ X$ ist.

Frage 1.20 Stimmt die Gleichheit $\mathbb{E}(Y|X = X(\omega)) = \mathbb{E}(Y|X)(\omega)$ für jedes ω, für $\mathbb{P}$-fast-jedes ω, oder i.A. nicht für $\mathbb{P}$-fast-jedes ω?

Haben wir mit Definition 1.19 unser Ziel der Konstruktion bedingter Verteilungen schon erreicht? Naheliegenderweise würden wir nun für jedes $A \in \mathcal{A}$ die bedingte Wahrscheinlichkeit $\mathbb{P}(A|X = x)$ von A gegeben $X = x$ definieren durch (h_A erfülle $\mathbb{E}(\mathbf{1}_A|X) = h_A \circ X$)

$$\mathbb{P}(A|X = x) := \mathbb{E}(\mathbf{1}_A|X = x) = h_A(x). \tag{1.17}$$

Ist die dadurch definierte Abbildung $A \mapsto \mathbb{P}(A|X = x)$ für jedes x ein Wahrscheinlichkeitsmaß auf $\mathcal{A}$? Die Frage ist nicht trivial zu beantworten, da die involvierten Funktionen h_A nur $\mathbb{P}^X$-überall eindeutig bestimmt sind (für absolut stetiges X können wir h_A also auf beliebigen Lebesgue-Nullmengen abändern), und die Nullmengen von A abhängen können. Beispielsweise gilt

$$\mathbb{P}(\emptyset|X = x) = 0$$

nur für $\mathbb{P}^X$-fast-alle $x \in \mathbb{R}$. Im Allgemeinen kann und wird $\mathcal{A}$ überabzählbar viele Mengen enthalten, wir können also nicht einfach auf den Nullmengen modifizieren, die entsprechenden Nullmengen vereinigen und argumentieren, dass die resultierende Menge wieder eine

Nullmenge ist. Wenn die σ-Algebra sich 'gut' durch abzählbar viele 'brave' Mengen approximieren lässt, kann aber tatsächlich eine Version von $\mathbb{P}(\cdot|X=x)$ bestimmt werden, sodass $A \mapsto \mathbb{P}(A|X=x)$ ein Wahrscheinlichkeitsmaß für jedes feste $x \in \mathbb{R}$, und die Abbildung $x \mapsto \mathbb{P}(A|X=x)$ messbar in x für jedes feste $A \in \mathcal{A}$ ist. $(x,A) \mapsto \mathbb{P}(A|X=x)$ ist dann ein sogenannter Markov Kern.

Definition 1.21 *Seien $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ Messräume. Eine Abbildung $K : \Omega_1 \times \mathcal{A}_2 \to [0,1]$ heißt* Markov Kern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$ *(oft auch 'stochastischer Kern' oder 'Übergangswahrscheinlichkeit') genau dann, wenn die folgenden zwei Bedingungen erfüllt sind:*

- *Für jedes feste $A_2 \in \mathcal{A}_2$ gilt: Die Abbildung $\omega_1 \mapsto K(\omega_1, A_2)$ ist $\mathcal{A}_1 - \mathcal{B}(\mathbb{R})$-messbar.*
- *Für jedes feste $\omega_1 \in \Omega_1$ ist $A_2 \mapsto K(\omega_1, A_2)$ ein Wahrscheinlichkeitsmaß auf $\mathcal{A}_2$.*

Beispiel 1.22 Wir betrachten $\Omega_1 = \Omega_2 = \Omega = \{1,2,\ldots,N\}$ mit $N \geq 2$ und $\mathcal{A}_1 = \mathcal{A}_2 = \mathcal{A} = \mathfrak{p}(\Omega)$. Weiters sei $M = (m_{i,j}) \in [0,1]^{N \times N}$ eine sogenannte stochastische Matrix, i.e. eine Matrix mit nicht-negativen Einträgen, die

$$\sum_{j=1}^{N} m_{i,j} = 1, \quad i \in \{1,2,\ldots,N\}$$

erfüllt. Definieren wir $K : \Omega \times \mathcal{A} \longrightarrow [0,1]$ durch

$$K(i,A) = \sum_{j \in A} m_{i,j},$$

dann ist offensichtlich K ein Markov Kern von Ω nach Ω (beachten Sie, dass die Messbarkeit wegen $\mathcal{A} = \mathfrak{p}(\Omega)$ trivial ist). Mit anderen Worten: Jede stochastische Matrix induziert einen Markov Kern. Umgekehrt (einfache Übungsaufgabe) induziert jeder Markov Kern eine stochastische Matrix.

Beispiel 1.23 Wir betrachten $\Omega_1 = (0,\infty), \mathcal{A}_1 = \mathcal{B}((0,\infty))$, $\Omega_2 = \mathbb{N}, \mathcal{A}_2 = \mathfrak{p}(\Omega_2)$ und setzen (für $B = \emptyset$ sei die Summe per definitionem gleich 0)

$$K(\theta, B) = \sum_{k \in B} \frac{\theta^k}{k!} e^{-\theta}$$

für jedes $\theta \in \Omega$ und jedes $B \in \mathcal{A}$. Dann ist K ein Markov Kern von $(\Omega_1, \mathcal{A}_1)$ nach $(\Omega_2, \mathcal{A}_2)$ (wie lässt sich die Messbarkeit zeigen?), und das Maß $K(\theta, \cdot)$ entspricht der Poisson-Verteilung mit Parameter θ.

Beispiel 1.24 (Fortsetzung von Beispiel 1.11) Sei (X,Y) stetig gleichverteilt auf dem Einheitskreis $B(0,1)$, i.e. die Dichte von (X,Y) ist gegeben durch $f = \frac{1}{\pi} \mathbf{1}_{B(0,1)}$. Wir setzen $\Omega := \Omega_1 = \Omega_2 = (-1,1)$, $\mathcal{A} = \mathcal{A}_1 = \mathcal{A}_2 = \mathcal{B}((-1,1))$, sowie $I_x := (-\sqrt{1-x^2}, \sqrt{1-x^2})$ für jedes $x \in (-1,1)$. Definieren wir $K : (-1,1) \times \mathcal{B}((-1,1)) \longrightarrow [0,1]$ durch

$$K(x,F) = \frac{1}{2\sqrt{1-x^2}} \int_E \mathbf{1}_{I_x}(y)\, d\lambda(y) = \frac{1}{2\sqrt{1-x^2}} \lambda\big(F \cap I_x\big), \tag{1.18}$$

dann lässt sich unschwer zeigen, dass K ein Markov Kern von $(\Omega, \mathcal{A})$ nach $(\Omega, \mathcal{A})$ ist (Übungsaufgabe!). Woher kommt die Idee, den Kern wie in Gleichung (1.18) zu wählen?

Definition 1.25 *Seien X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$. Dann heißt ein Markov Kern K von $(\Omega, \mathcal{C})^\dagger$ nach $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine* reguläre bedingte Verteilung von Y unter $\mathcal{C}$ *genau dann wenn*

$$K(\omega, B) = \mathbb{E}(\mathbf{1}_B \circ Y|\mathcal{C})(\omega)\ [\mathbb{P}]$$

für jedes $B \in \mathcal{B}(\mathbb{R})$ gilt.
Für den Spezialfall $\mathcal{C} = \mathcal{A}_\sigma(X)$ heißt ein Markov Kern K von $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ nach $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ eine reguläre bedingte Verteilung von Y unter X *genau dann, wenn für jedes $B \in \mathcal{B}(\mathbb{R})$ die Gleichheit*

$$K(X(\omega), B) = \mathbb{E}(\mathbf{1}_B \circ Y|X)(\omega)$$

für $\mathbb{P}$-fast alle ω gilt (mit anderen Worten: wenn $(\omega, B) \mapsto K(X(\omega), B)$ eine Version der bedingten Erwartung von $\mathbf{1}_B \circ Y$ unter $\mathcal{A}_\sigma(X)$ ist.)

Bevor wir die Existenz regulärer bedingter Verteilungen beweisen, unterstreichen wir deren Wichtigkeit für die Zerlegung von Maßen:

Bemerkung 1.26 Falls K eine reguläre bedingte Verteilung von Y unter X ist, dann gilt also per definitionem für alle $E, B \in \mathcal{B}(\mathbb{R})$ die folgenden Gleichheit

$$\int_{X^{-1}(E)} K(X(\omega), B) d\mathbb{P}(\omega) = \int_{X^{-1}(E)} \mathbf{1}_B \circ Y d\mathbb{P} = \mathbb{P}(X^{-1}(E) \cap Y^{-1}(B)) = \mathbb{P}^{(X,Y)}(E \times B).$$

Nachdem sich der erste Ausdruck mittels Change of Coordinates umschreiben lässt zu

$$\int_{X^{-1}(E)} K(X(\omega), B) d\mathbb{P}(\omega) = \int_\Omega \mathbf{1}_E \circ X(\omega)\, K(X(\omega), B) d\mathbb{P}(\omega) = \int_E K(x, B) d\mathbb{P}^X(x)$$

erhalten wir insgesamt

$$\mathbb{P}^{(X,Y)}(E \times B) = \int_E K(x, B) d\mathbb{P}^X(x), \tag{1.19}$$

das Wahrscheinlichkeitsmaß $\mathbb{P}^{(X,Y)}$ wird also zerlegt (disintegriert) in eine Familie eindimensionaler Maße $K(x, \cdot)$ und die Randverteilung $\mathbb{P}^X$ von X.

Gleichung (1.19) charakterisiert sogar reguläre bedingte Verteilungen von Y unter X - es gilt das folgende Lemma:

Lemma 1.27 *Ein Markov Kern K von $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ nach $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ist eine reguläre bedingte Verteilung von Y unter X genau dann, wenn Gleichung (1.19) für alle $E, B \in \mathcal{B}(\mathbb{R})$ gilt.*

Beweis: Einfache Übungsaufgabe.

Beispiel 1.28 (Fortsetzung von Beispiel 1.24) Wir bestätigen die in Bemerkung 1.26 hergeleitete Gleichung (1.19) und berechnen zuerst die (Rand-)Verteilung $\mathbb{P}^X$ von X: Für $x \in (-1, 1)$ gilt

$$\begin{aligned}
\mathbb{P}(X \le x) &= \mathbb{P}\big(X \le x, Y \in (-1,1)\big) = \int_{(-1,x]\times(-1,1)} f(t,s) d\lambda_2(t,s) \\
&= \int_{(-1,x]} \frac{1}{\pi} \int_{(-1,1)} \mathbf{1}_{I_x}(y) d\lambda(y) d\lambda(x) = \frac{1}{\pi} \int_{(-1,x]} 2\sqrt{1-x^2} d\lambda(x).
\end{aligned}$$

[†] Beachten Sie, dass hier die Meßbarkeit bezüglich der sub-σ-Algebra $\mathcal{C}$ gefordert wird

X ist also absolut stetig mit Dichte $\frac{2}{\pi}\sqrt{1-x^2}\,\mathbf{1}_{(-1,1)}(x)$. Für jedes ganz in $(-1,1)^2$ liegende, messbare Rechteck $R = E \times F$ erhalten wir

$$\begin{aligned}
\int_E K(x,F)d\mathbb{P}^X(x) &= \int_E \frac{1}{2\sqrt{1-x^2}}\lambda\big(F\cap I_x\big)d\mathbb{P}^X(x) \\
&= \int_E \frac{1}{2\sqrt{1-x^2}}\lambda\big(F\cap I_x\big)\frac{2}{\pi}\sqrt{1-x^2}\,\mathbf{1}_{(-1,1)}(x)d\lambda(x) \\
&= \frac{1}{\pi}\lambda(F)\lambda(E) = \frac{1}{\pi}\lambda_2(E\times F) = \mathbb{P}^{(X,Y)}(E\times F).
\end{aligned}$$

Satz 1.29 (Existenz regulärer bedingter Verteilungen) *Sei Y eine Zufallsvariable auf $(\Omega,\mathcal{A},\mathbb{P})$ und $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$. Dann existiert eine reguläre bedingte Verteilung von Y unter $\mathcal{C}$.*

Beweis: Die Beweisidee besteht darin, Versionen der bedingten Verteilungsfunktionen geschickt zu wählen, und daraus einen Markov Kern zu basteln.
Für jedes $r \in \mathbb{Q}$ bezeichne $F(r,\cdot)$ eine Version von $\mathbb{E}(\mathbf{1}_{(-\infty,r]}\circ Y|\mathcal{C})$. Nachdem für $r < s, r, s \in \mathbb{Q}$ offensichtlich $\mathbf{1}_{(-\infty,r]}\circ Y(\omega) \le \mathbf{1}_{(-\infty,s]}\circ Y(\omega)$ für jedes $\omega \in \Omega$ gilt, folgt aus Satz 1.16 Punkt 2. sofort die Existenz einer Nullmenge $A_{r,s} \in \mathcal{C}$ mit

$$F(r,\omega) \le F(s,\omega) \qquad \text{für jedes } \omega \in A^c_{r,s}.$$

Weiters existiert lt. Satz 1.16 Punkt 8. eine Nullmenge $B_r \in \mathcal{C}$ mit

$$\lim_{n\to\infty} F\left(r+\frac{1}{n},\omega\right) = F(r,\omega) \qquad \text{für jedes } \omega \in B^c_r$$

sowie eine weitere Nullmenge $D \in \mathcal{C}$ mit

$$\lim_{n\to\infty} F(-n,\omega) = 0, \quad \text{und} \quad \lim_{n\to\infty} F(n,\omega) = 1 \qquad \text{für jedes } \omega \in D^c$$

Setzen wir

$$N := \bigcup_{r,s\in\mathbb{Q},r<s} A_{r,s} \cup \bigcup_{r\in\mathbb{Q}} B_r \cup D$$

dann folgt sofort $N \in \mathcal{C}$ sowie $\mathbb{P}(N) = 0$. Für jedes $\omega \in N^c$ ist $r \mapsto F(r,\omega)$ die Einschränkung einer Verteilungsfunktion auf $\mathbb{Q}$. Diese Verteilungsfunktion $\overline{F}(\cdot,\omega)$ ist offensichtlich gegeben durch

$$\overline{F}(z,\omega) = \inf\{F(r,\omega) : r \in \mathbb{Q}\cap(z,\infty)\}.$$

Wir wählen eine beliebige eindimensionale Verteilungsfunktion F_0 und definieren $\overline{F}$ auf $\mathbb{R}\times N^c$ durch $\overline{F}(z,\omega) = F_0(z)$. Ausgehend von $\overline{F}$ definieren wir K für $\omega \in \Omega$ und $B = (-\infty, r]$ durch

$$K(\omega,B) = \overline{F}(r,\omega). \tag{1.20}$$

und zeigen, dass K zu einem Markov-Kern mit den gewünschten Eigenschaften fortgesetzt werden kann. Für $r \in \mathbb{Q}$ gilt nach Konstruktion

$$K(\omega,(-\infty,r]) = \mathbf{1}_{N^c}(\omega)F(r,\omega) + \mathbf{1}_N(\omega)F_0(r),$$

die Fortsetzung von K auf ganz $\Omega \times \mathcal{B}(\mathbb{R})$ ist trivial[†]. Nachdem alle vier auf der rechten Seite vorkommenden Funktionen $\mathcal{C}$-messbar sind, ist auch $\omega \mapsto K(\omega, (-\infty, r])$ $\mathcal{C}$-messbar für jedes $r \in \mathbb{Q}$. Wir wissen, dass $\{(-\infty, r] : r \in \mathbb{Q}\}$ ein durchschnittsstabiler Erzeuger der Borelschen σ-Algebra $\mathcal{B}(\mathbb{R})$ ist; zusätzlich kann leicht überprüft werden, dass

$$\mathcal{D} = \big\{B \in \mathcal{B}(\mathbb{R}) : \omega \mapsto K(\omega, B) \text{ messbar}\big\}$$

ein Dynkin-System ist. Insgesamt folgt also (π-λ-Theorem, Übungsaufgabe) $\mathcal{D} = \mathcal{B}(\mathbb{R})$. K ist also ein Markov-Kern und der Beweis ist komplett, sobald gezeigt ist, dass für jedes $B \in \mathcal{B}(\mathbb{R})$ die $\mathcal{C}$-messbare Abbildung $\omega \mapsto K(\omega, B)$ eine Version der bedingten Erwartung von Y unter $\mathcal{C}$ ist. Für $B = (-\infty, r]$ und beliebiges $C \in \mathcal{C}$ folgt die gewünschte Gleichheit aus

$$\begin{aligned} L(B) &:= \int_C K(\omega, B) d\mathbb{P}(\omega) = \int_C \mathbf{1}_{N^c}(\omega) F(r, \omega) d\mathbb{P}(\omega) = \int_{C \cap N^c} \underbrace{F(r, \omega)}_{= \mathbb{E}(\mathbf{1}_{(-\infty, r]} \circ Y | \mathcal{C}) \ [\mathbb{P}]} d\mathbb{P}(\omega) \\ &= \int_{C \cap N^c} \mathbf{1}_{(-\infty, r]} \circ Y \, d\mathbb{P} = \mathbb{P}\left(C \cap Y^{-1}(B)\right) =: R(B) \end{aligned}$$

Nachdem $B \mapsto L(B)$ und $B \mapsto R(B)$ endliche Maße sind, die auf $\{(-\infty, r] : r \in \mathbb{Q}\}$ übereinstimmen, folgt $L = R$ auf ganz $\mathcal{B}(\mathbb{R})$. ■

Satz 1.29 gilt nicht nur für reellwertige Zufallsvariable Y, sondern ganz allgemein für Zufallsvariable Y mit Werten in einem sogenannten *Borel Raum*. Nachdem jeder separable, vollständige metrischen Raum (M, d) ein Borel Raum ist, existieren reguläre bedingte Verteilungen also insbesondere für Zufallsvariable Y mit Werten in $(\mathbb{R}^d, \|\cdot\|_2)$, in $(C[a, b], \|\cdot\|_\infty)$, etc.:

Korollar 1.30 (Existenz regulärer bedingter Verteilungen, allgemein) *Sei Y eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem vollständigen, separablen metrischen Raum (M, d) und $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$. Dann existiert eine reguläre bedingte Verteilung von Y unter $\mathcal{C}$.*

Beweis: Mit guten Maßtheoriekenntnissen leicht machbar, siehe [4, 5].

Die Kenntnis einer reguläre bedingte Verteilung $K(\cdot, \cdot)$ von Y unter $\mathcal{C}$ ermöglicht die Berechnung einer bedingten Erwartung $g \circ Y$ unter $\mathcal{C}$ - es gilt folgendes Resultat:

Satz 1.31 *Sei Y eine Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in einem vollständigen, separablen metrischen Raum (M, d), $\mathcal{C}$ eine sub-σ-Algebra von $\mathcal{A}$, und $g : M \to \mathbb{R}$ Borel-messbar mit $\mathbb{E}(|g \circ Y|) < \infty$. Dann gilt*

$$\mathbb{E}\big(g \circ Y | \mathcal{C}\big)(\omega) = \int_M g(y) K(\omega, dy) \quad [\mathbb{P}]$$

Beweis: Übungsaufgabe

Für den Fall $\mathcal{C} = \mathcal{A}_\sigma(X)$ lässt sich noch mehr machen. Der Beweis des folgenden Resultats verläuft vollkommen analog zu dem von Satz 1.29 - anstatt $(\Omega, \mathcal{A}, \mathbb{P})$ arbeitet man gleich auf dem Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}^X)$.

[†] 1:1 Zusammenhang 1-dimensionaler Verteilungsfunktionen und der entsprechenden Wahrscheinlichkeitsmaße

Satz 1.32 (Existenz regulärer bedingter Verteilungen von Y unter X) *Seien X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann existiert eine reguläre bedingte Verteilung $K(\cdot,\cdot)$ von Y unter X. Dieses $K(x,\cdot)$ ist für $\mathbb{P}^X$-fast alle $x \in \mathbb{R}$ eindeutig bestimmt. Falls X und Y unabhängig sind, ist $K(x,B) = \mathbb{P}^Y(B)$ eine reguläre bedingte Verteilung von Y unter X.*

In [4] ist die folgende, wesentlich allgemeinere Version dieses Resultats bewiesen (selbe Beweisstruktur und Verwendung der Isomorphie von Borel Räumen und $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$):

Korollar 1.33 (Existenz reg. bedingter Verteilungen von Y unter X, allgemein) *Sei (M,d) ein separabler, vollständiger metrischer Raum und $(N, \mathcal{N})$ ein Messraum. Weiters sei X eine N-wertige, und Y eine M-wertige Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann existiert eine reguläre bedingte Verteilung $K(\cdot,\cdot)$ von Y unter X. Dieses $K(x,\cdot)$ ist für $\mathbb{P}^X$-fast alle $x \in N$ eindeutig bestimmt. Falls X und Y unabhängig sind, ist $K(x,B) = \mathbb{P}^Y(B)$ eine reguläre bedingte Verteilung von Y unter X.*

Nach mehr oder weniger aufwändigen Vorbereitungen können wir nun endlich das Disintegrationstheorem formulieren und beweisen - wir formulieren das Resultat für das einfache Setting in Satz 1.32, die Übertragung auf den allgemeinen Fall ist problemlos.

Satz 1.34 (Disintegrationstheorem) *Seien X, Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ und $K(\cdot,\cdot)$ eine reguläre bedingte Verteilung von Y unter X. Weiters sei $g : \mathbb{R}^2 \to \mathbb{R}$ messbar mit $\mathbb{E}(|g(X,Y)|) < \infty$. Dann folgt:*

$$\mathbb{E}\big(g(X,Y)|X\big)(\omega) = \int_{\mathbb{R}} g(X(\omega),y)K(X(\omega),dy)\ [\mathbb{P}] \tag{1.21}$$

$$\mathbb{E}\big(g(X,Y)\big) = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)K(x,dy)\, d\mathbb{P}^X(x) \tag{1.22}$$

Beweisskizze: Offensichtlich folgt (1.22) sofort mittels Integration bezüglich $\mathbb{P}$ und Change of Coordinates aus (1.21). Letztere Gleichung folgt, indem man, wie gewohnt, dem Integralaufbau folgt und mit $g = \mathbf{1}_{E\times F}$ startet (Übungsaufgabe).

Nachdem Satz 1.34 das Hauptziel dieses Abschnitts war, einige ergänzende Beobachtungen:

(B1) Der elementare Satz von der totalen Wahrscheinlichkeit ist ein Spezialfall von Gleichung (1.22): Sei X eine diskrete Zufallsvariable mit Wertebereich $\{\alpha_1, \alpha_2, \ldots\}$ und $\mathbb{P}(X = \alpha_i) > 0$ für jedes i. Für $B \in \mathcal{B}(\mathbb{R})$ und $g(x,y) = \mathbf{1}_B(x)$ erhalten wir aus (1.22)

$$\begin{aligned}\mathbb{P}(Y \in B) &= \mathbb{E}(\mathbf{1}_B(Y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_B(y)K(x,dy)\, d\mathbb{P}^X(x) = \int_{\mathbb{R}} K(x,B)\, d\mathbb{P}^X(x)\\ &= \sum_{i=1}^{\infty} K(\alpha_i, B)\mathbb{P}(X=\alpha_i) = \sum_{i=1}^{\infty} \mathbb{P}(Y \in B|X=\alpha_i)\mathbb{P}(X=\alpha_i).\end{aligned}$$

(B2) Für den Fall, dass X und Y unabhängig sind, liefert (1.22) unter Verwendung von Change of Coordinates sofort

$$\begin{aligned}\int_{\mathbb{R}^2} g(x,y)d\mathbb{P}^{(X,Y)}(x,y) &= \mathbb{E}\big(g(X,Y)\big) = \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)K(x,dy)\, d\mathbb{P}^X(x)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} g(x,y)d\mathbb{P}^Y(y)\, d\mathbb{P}^X(x),\end{aligned}$$

also genau den Satz von Fubini (für das Produktmaß aus $\mathbb{P}^X$ und $\mathbb{P}^Y$).

(B3) Sei $G \in \mathcal{B}(\mathbb{R}^2)$ beliebig. Definieren wir den x-Schnitt G_x von G durch $G_x := \{y \in \mathbb{R} : (x,y) \in G\} \in \mathcal{B}(\mathbb{R})$, dann liefert (1.22)

$$\begin{aligned}\mathbb{P}^{(X,Y)}(G) &= \mathbb{E}(\mathbf{1}_G(X,Y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_G(x,y)K(x,dy)\,d\mathbb{P}^X(x)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbf{1}_{G_x}(y)K(x,dy)\,d\mathbb{P}^X(x) = \int_{\mathbb{R}} K(x,G_x)\,d\mathbb{P}^X(x)\end{aligned}$$

Beispiel 1.35 (Fortsetzung von Beispiel 1.24) Wir bestätigen die in Bemerkung (B3) hergeleitete Gleichheit für den Fall, dass (X,Y) stetig gleichverteilt am Einheitskreis $B(0,1)$ ist. Es wurde schon gezeigt, dass X absolut stetig mit Dichte $\frac{2}{\pi}\sqrt{1-x^2}\,\mathbf{1}_{(-1,1)}(x)$ ist.
Wir berechnen abschließend die Wahrscheinlichkeit dafür, dass (X,Y) im Kreis $B(0,r)$ mit Radius $r \in [0,1]$ liegt und setzen $I_x^r := (-\sqrt{r^2-x^2}, \sqrt{r^2-x^2})$ für jedes $x \in (0,1)$:

$$\begin{aligned}\mathbb{P}^{(X,Y)}(B(0,r)) &= \int_{\mathbb{R}} K(x,B(0,r)_x)\,d\mathbb{P}^X(x) = \int_{(-r,r)} K(x,I_x^r)\,d\mathbb{P}^X(x)\\ &= \int_{(-r,r)} \frac{2\sqrt{r^2-x^2}}{2\sqrt{1-x^2}}\,d\mathbb{P}^X(x) = \int_{(-r,r)} \frac{2\sqrt{r^2-x^2}}{2\sqrt{1-x^2}}\frac{2}{\pi}\sqrt{1-x^2}\,d\lambda(x)\\ &= \frac{2}{\pi}\int_{(-r,r)} \sqrt{r^2-x^2}d\lambda(x) = \frac{r^2\pi}{\pi} = r^2\end{aligned}$$

Kapitel 2

Analytische Eigenschaften von Copulas

Wir betrachten ob des geringeren technischen Aufwands zuerst zweidimensionale Copulas, studieren deren analytische Eigenschaften, beweisen den Satz von Sklar, und erweitern dann erst auf den multivariaten Fall.

2.1 Der zweidimensionale Fall

2.1.1 Definition und alternative Darstellungen

Die einfachtmögliche (implizite) Definition einer zweidimensionalen Copula ist wiefolgt: Seien X und Y auf $[0,1]$ stetig gleichverteilte Zufallsvariable. Dann heißt die Verteilungsfunktion von (X,Y) eingeschränkt auf $[0,1]^2$ eine Copula. Setzen wir für $(x,y) \in [0,1]^2$

$$A(x,y) := \mathbb{P}^{(X,Y)}([0,x] \times [0,y]) = \mathbb{P}(X \leq x, Y \leq y)$$

dann hat $A : [0,1]^2 \to [0,1]$ unter anderem folgende Eigenschaften:

- Für alle $x, y \in [0,1]$ gilt $A(x,1) = x, A(1,y) = y, A(x,0) = 0 = A(0,y)$.
  Beweis: Die erste Gleichheit folgt sofort aus

  $$A(x,1) = \mathbb{P}(X \leq x, Y \leq 1) = \mathbb{P}(X \leq x) = x,$$

  die anderen Gleichheiten lassen sich analog herleiten.

- Für $0 \leq x_1 \leq x_2 \leq 1$ und $0 \leq y_1 \leq y_2 \leq 1$ gilt

  $$\begin{aligned} A(x_2,y_2) - A(x_1,y_2) - A(x_2,y_1) + A(x_1,y_1) &= \mathbb{P}^{(X,Y)}\big((x_1,x_2] \times (y_1,y_2]\big) \\ &= \mathbb{P}^{(X,Y)}\big([x_1,x_2] \times [y_1,y_2]\big) \geq 0. \end{aligned}$$

  A ist also 2-monoton auf $[0,1]^2$ wie Sie es von allgemeinen zweidimensionalen Verteilungsfunktionen aus der VO Statistik kennen.

Die zwei soeben erwähnten Eigenschaften liefern die in der Literatur übliche Definition zweidimensionaler Copulas:

Definition 2.1 *Eine* Copula *ist eine Funktion* $A : [0,1]^2 \to [0,1]$ *mit den folgenden Eigenschaften:*

1. $\forall x \in [0,1] : A(x,1) = A(1,x) = x,\ A(x,0) = A(0,x) = 0.$
2. *Für* $0 \le x_1 \le x_2 \le 1$ *und* $0 \le y_1 \le y_2 \le 1$ *gilt*

$$A(x_2,y_2) - A(x_1,y_2) - A(x_2,y_1) + A(x_1,y_1) \ge 0. \tag{2.1}$$

$\mathcal{C}$ *bezeichnet im Folgenden die Familie aller zweidimensionalen Copulas.*

Bemerkung 2.2 Obwohl Copulas nur auf $[0,1]^2$ definiert sind, können Sie mit eindeutigen zweidimensionalen Verteilungsfunktionen H_A identifiziert, bzw. zu zweidimensionalen Verteilungsfunktionen fortgesetzt werden. Wie funktioniert diese Fortsetzung?

Beispiel 2.3 Die folgenden Funktionen sind Copulas (direktes Nachrechnen):

$$M(x,y) = \min\{x,y\}, \quad \Pi(x,y) = xy, \quad W(x,y) = \max\{x+y-1,0\}$$

Lemma 2.4 *Sei* $A \in \mathcal{C}$ *beliebig, dann gilt:*

1. *A ist monoton nicht-fallend in jeder Koordinate.*
2. *Für* $x_1, x_2, y_1, y_2 \in [0,1]$ *gilt*

$$|A(x_2,y_2) - A(x_1,y_1)| \le |x_2 - x_1| + |y_2 - y_1|. \tag{2.2}$$

Jede Copula ist also (Lipschitz-) stetig.

Beweis: Einfache Übungsaufgabe.

Definition 2.5 *Ein Wahrscheinlichkeitsmaß* μ *auf* $([0,1]^2, \mathcal{B}([0,1]^2))$ *heißt* doppelt stochastisch*, genau dann, wenn seine eindimensionalen Randverteilungen mit* λ *zusammenfallen, i.e. wenn für jedes* $E \in \mathcal{B}([0,1])$ *die Gleichheit*

$$\mu(E \times [0,1]) = \lambda(E) = \mu([0,1] \times E) \tag{2.3}$$

gilt. $\mathcal{P}_\mathcal{C}$ *bezeichnet im Folgenden die Menge aller doppelt stochastischen Maße.*

Offensichtlich entspricht jeder Copula A genau ein doppelt stochastisches Maß μ_A und umgekehrt - es gilt folgender Satz:

Lemma 2.6 *Für jede Copula* $A \in \mathcal{C}$ *ist das durch* $\mu_A([0,x] \times [0,y]) = A(x,y)$ *auf* $\mathcal{B}([0,1]^2)$ *definierte Wahrscheinlichkeitsmaß*[†] *doppelt stochastisch. Umgekehrt ist für jedes doppelt stochastische Maß* $\mu \in \mathcal{P}_\mathcal{C}$ *die durch* $A(x,y) = \mu([0,x] \times [0,y])$ *auf* $[0,1]^2$ *definierte Funktion eine Copula.*

[†] die Eindeutigkeit der Fortsetzung ist klar aus dem bijektiven Zusammenhang zweidimensionaler Verteilungsfunktionen und zweidimensionalen Wahrscheinlichkeitsmaßen auf $\mathcal{B}(\mathbb{R}^2)$

Beweis: Sei $A \in \mathcal{C}$ beliebig, μ_A für $x, y \in [0,1]$ definiert durch $\mu_A([0,x] \times [0,y]) = A(x,y)$ und in gewohnter Manier (eindeutig!) fortgesetzt auf ganz $\mathcal{B}([0,1]^2)$. Das Wahrscheinlichkeitsmaß $\nu_1 : \mathcal{B}([0,1]) \to [0,1]$ (a.k.a. erste Randverteilung von μ_A), definiert durch

$$\nu_1(E) := \mu_A(E \times [0,1])$$

erfüllt $\nu_1([0,x]) = \mu_A([0,x] \times [0,1]) = A(x,1) = x$ für jedes $x \in [0,1]$, woraus sofort $\nu_1 = \lambda$ folgt. Die zweite Aussage des Satzes ist einfach nachzurechnen. ■

Bemerkung 2.7 In der Folge bezeichnet μ_A das zur Copula $A \in \mathcal{C}$ gehörige doppelt stochastische Maß und $A_\mu \in \mathcal{C}$ die zum doppelt stochastischen Maß $\mu \in \mathcal{P}_\mathcal{C}$ gehörige Copula.

Copulas lassen sich weiters identifizieren mit speziellen Markov Kernen von $([0,1], \mathcal{B}([0,1]))$ nach $([0,1], \mathcal{B}([0,1]))$ (siehe [8]):

Lemma 2.8 *Sei A eine beliebige zweidimensionale Copula. Dann existiert ein Markov Kern $K(\cdot,\cdot)$ von $([0,1], \mathcal{B}([0,1]))$ nach $([0,1], \mathcal{B}([0,1]))$ mit*

$$\mu_A(E \times F) = \int_E K(x,F)\, d\lambda(x) \tag{2.4}$$

für alle $E, F \in \mathcal{B}([0,1])$. Dieser Kern ist für λ-fast jedes $x \in [0,1]$ eindeutig bestimmt (wir nennen ein solches $K(\cdot,\cdot)$ im Folgenden kurz 'Markov Kern von A').
Erfüllt umgekehrt ein Markov Kern $K(\cdot,\cdot)$ von $([0,1], \mathcal{B}([0,1]))$ nach $([0,1], \mathcal{B}([0,1]))$

$$\int_{[0,1]} K(x,F) d\lambda(x) = \lambda(F) \tag{2.5}$$

für jedes $F \in \mathcal{B}([0,1])$, dann ist $K(\cdot,\cdot)$ Markov Kern der Copula $A \in \mathcal{C}$, definiert durch

$$A(x,y) = \int_{[0,x]} K(t,[0,y]) d\lambda(t) \tag{2.6}$$

für $x, y \in [0,1]$.

Beweis: Seien X, Y ($\mathcal{U}(0,1)$-verteilte) Zufallsvariable mit $(X,Y) \sim A^\dagger$. Direkte Anwendung von Satz 1.32 und dem Disintegrationstheorem mit $g := \mathbf{1}_{E\times F}$ liefert die ersten beiden Aussagen. Die verbleibende dritte Aussage ist eine einfache Rechnenaufgabe, es müssen nur die in Definition 2.1 angegebenen Eigenschaften nachgerechnet werden. ■

Ein vierter alternativer Zugang zu Copulas ist mittels sog. Markov Operatoren und geht zurück auf [1]: Ein linearer operator T auf $L^1([0,1], \mathcal{B}([0,1]), \lambda)$ heißt *Markov Operator* (im engeren Sinne), falls er die folgenden drei Eigenschaften erfüllt (alles im L^1-Sinne zu interpretieren):

1. T ist positive, i.e. $T(g) \geq 0$ falls $g \geq 0$
2. $T(\mathbf{1}_{[0,1]}) = \mathbf{1}_{[0,1]}$
3. $\int_{[0,1]} (Tg)(x) d\lambda(x) = \int_{[0,1]} g(x) d\lambda(x)$

†i.e. (X,Y) hat Verteilungsfunktion A

Die Menge aller Markov Operatoren auf $L^1([0,1],\mathcal{B}([0,1]),\lambda)$ wird im Folgenden mit $\mathcal{M}$ bezeichnet. Markov Operatoren haben offensichtlich Operatornorm 1 (einfache Übungsaufgabe).

Lemma 2.9 *Sei $A \in \mathcal{C}$ beliebig, dann ist T_A, definiert durch ($A_{,2}(x,t) := \frac{\partial A}{\partial y}(x,t)$)*

$$(T_A g)(x) := \frac{d}{dx}\int_{[0,1]} A_{,2}(x,t)g(t)d\lambda(t) \tag{2.7}$$

für jedes $g \in L^1([0,1])$ ein Markov Operator, i.e. $T_A \in \mathcal{M}$.
Umgekehrt ist für jeden Markov Operator $T \in \mathcal{M}$ die Abbildung A_T, definiert durch

$$A_T(x,y) := \int_{[0,x]} (T\mathbf{1}_{[0,y]})(t)d\lambda(t) \tag{2.8}$$

für alle $x,y \in [0,1]$ eine Copula.

Beweis: Übungsaufgabe

Gleichung (2.7) lässt sich mittels bedingter Erwartungen und Markov Kernen vereinfachen zu (siehe [8])

$$(T_A g)(x) = \mathbb{E}(g \circ Y | X = x) = \int_{[0,1]} g(y)K_A(x,dy). \tag{2.9}$$

Der Markov Operator beschreibt also nichts anderes als den bedingten Erwartungswert von $g \circ Y$ gegeben $X = x$, wobei $(X,Y) \sim A$.

Beispiel 2.10 Die Wahrscheinlichkeitsdichte f sei gegeben durch

$$f(x,y) = \big(1 + \theta(1-2x)(1-2y)\big)\mathbf{1}_{[0,1]^2}(x,y)$$

mit $\theta \in [-1,1]$. Dann ist das zu f gehörige Wahrscheinlichkeitsmaß μ doppelt stochastisch: Offensichtlich gilt für jedes $E \in \mathcal{B}([0,1])$

$$\mu([0,1]\times E) = \int_{[0,1]\times E} f(x,y)d\lambda_2(x,y) \quad = \quad \int_E \int_{[0,1]} f(x,y)d\lambda(x)d\lambda(y) = \int_E 1 d\lambda(y) = \lambda(E),$$

und wegen der Symmetrie der Dichte daher auch $\lambda(E) = \mu([0,1]\times E) = \mu(E \times [0,1])$. $A \in \mathcal{C}$ bezeichne die entsprechende Copula. Beachten Sie: Für $\theta = 0$ gilt $A = \Pi$.
Wir berechnen den Markov Kern $K_A(\cdot,\cdot)$ und den Markov Operator T_A und erhalten

$$\begin{aligned} K_A(x,[0,y]) &= \int_{[0,y]} f(x,s)d\lambda(s) = y + \theta(1-2x)\int_{[0,y]}(1-2s)d\lambda(s) \\ &= y + \theta(1-2x)(y-y^2) \end{aligned}$$

und

$$(T_A\, g)(x) \quad = \quad \int_{[0,1]} g(y)K_A(x,dy) = \int_{[0,1]} g(y)f(x,y)d\lambda(y).$$

Beispiel 2.11 Eine messbare Abbildung $h : [0,1] \to [0,1]$ heißt λ-treu (oder λ-erhaltend) genau dann, wenn $\lambda^h = \lambda$ gilt. Für $X \sim \mathcal{U}(0,1)$ setzen wir $Y = h \circ X$. Dann folgt $Y \sim \mathcal{U}(0,1)$, die (auf $[0,1]^2$ betrachtete) Verteilungsfunktion A_h von (X,Y) ist daher eine Copula (die wir später als 'vollständig abhängig' bezeichnen werden). Markov Kern K_{A_h} und Markov Operator T_{A_h} sind in diesem Fall besonders einfach - es gilt

$$K_{A_h}(x,F) = \mathbf{1}_F(h(x)) = \delta_{h(x)}(E)$$

und

$$(T_{A_h} g)(x) = \int_{[0,1]} g(y) K_{A_h}(x, dy) = g \circ h(x).$$

Copulas der Form A_h mögen auf den ersten Blick viel zu speziell/zu exotisch und daher nutzlos erscheinen - wir im Laufe der LV sehen, dass dieser erste Eindruck ganz und gar falsch ist.

2.1.2 Der metrische Raum $(\mathcal{C}, d_\infty)$

Literatur

[1] W.F. Darsow, B. Nguyen, E.T. Olsen: Copulas and Markov processes, *Illinois Journal of Mathematics* 36, 600-642 (1992)

[2] F. Durante, C. Sempi: *Principles of copula theory*, CRC Press Boca Raton (2016)

[3] J. Elstrodt: *Maß- und Integrationstheorie*, Springer Verlag, Berlin Heidelberg New York, 1999

[4] O. Kallenberg: *Foundations of modern probability*, Springer New York (2002)

[5] A. Klenke: *Probability Theory - a comprehensive course*, Springer, 2008

[6] R. B. Nelsen: *An introduction to copulas*, Springer New York (2006)

[7] W. Rudin: *Real and complex analysis*, McGraw-Hill New York (1987)

[8] W. Trutschnig: On a strong metric on the space of copulas and its induced dependence measure, *Journal of Mathematical Analysis and Applications* 384, 690-705 (2011)

Kapitel 3

Übungsblätter

01. Übung am 19. März 2018

[LVA 405.552 UV Abhängigkeitsmodellierung, Ankreuzliste siehe www.trutschnig.net/courses mit * versehene Aufgaben sind freiwillig]

Übungsaufgabe 1 Beweisen Sie Lemma 1.6 (Aufwärmübung!)

Übungsaufgabe 2 Beweisen Sie die in Satz 1.7 behauptete Eindeutigkeit des Paares (ν_a, ν_s) (eine zweite Aufwärmübung!).

Übungsaufgabe 3 Beweisen Sie Satz 1.9.

Übungsaufgabe 4 Wir betrachten $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. X sei exponentialverteilt mit Parameter $\theta = 1$, Y sei stetig gleichverteilt auf $[0, 1]$. $\mu = \mathbb{P}^X$ und $\nu = \mathbb{P}^Y$ bezeichnen die entsprechenden Verteilungen. Bestimmen Sie die Lebesgue Zerlegung $\nu = \nu_a + \nu_s$ von ν bezüglich μ sowie die Lebesgue Zerlegung $\mu = \mu_a + \mu_s$ von μ bezüglich ν.

Übungsaufgabe 5 Die Definition der Absolut-Stetigkeit eines Maß ν bezüglich eines Maßes μ funktioniert auch für unendliche Maße. Wir betrachten die folgenden zwei Maße ν, μ auf $(\mathbb{R}, \mathcal{B}([0, 1]))$: μ sei das Zählmaß und ν sei das Lebesgue-Maß, i.e.

$$\mu(E) = \#E, \quad \nu(E) = \lambda(E).$$

für jedes $E \in \mathcal{B}([0, 1])$. Gilt dann $\nu \ll \mu$ (oder umgekehrt)? Existiert eine Radon-Nikodym Ableitung von ν bezüglich μ (oder umgekehrt)?

Übungsaufgabe 6 (just for fun*) † Die Folge $(a_i)_{i\in\mathbb{N}}$ konvergiere gegen $a \in \mathbb{R}$, $k \in \mathbb{N}$ sei beliebig. Berechnen Sie

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{i}{n}\right)^k a_i$$

† eine meiner ehemaligen Prüfungsaufgaben in Maßtheorie, hat nichts mit Radon-Nikodym oder Lebesgue Zerlegung zu tun, ist aber eine gute Übung.

02. Übung am 09. April 2018

[LVA 405.552 UV Abhängigkeitsmodellierung, Ankreuzliste siehe www.trutschnig.net/courses mit * versehene Aufgaben sind freiwillig]

Übungsaufgabe 7 Beweisen Sie Satz 1.10.

Übungsaufgabe 8 Berechnen Sie für $(\Omega, \mathcal{A}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$, $X(\omega) = \omega^2$ und $\mathcal{C} = \mathcal{A}_\sigma(\{[0,1/4], (1/4,3/4], (3/4,1]\})$ die bedingte Erwartung $\mathbb{E}(X|\mathcal{C})$ von X unter $\mathcal{C}$.

Übungsaufgabe 9 Beweisen Sie mindestens zwei der Aussagen 4-9 in Satz 1.16.

Übungsaufgabe 10 Wir betrachten den Wahrscheinlichkeitsraum $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathbb{P})$, wobei $\mathbb{P} = \mu \otimes \mu^\dagger$ und μ ein beliebiges Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbb{R})$ bezeichne. Die Familie $\mathcal{B}_s(\mathbb{R}^2)$ aller symmetrischen Borelmengen ist definiert durch†

$$\mathcal{B}_s(\mathbb{R}^2) = \big\{B \in \mathcal{B}(\mathbb{R}^2) : B^t = B\big\}.$$

Zeigen Sie, dass $\mathcal{B}_s(\mathbb{R}^2)$ eine σ-Algebra ist, und berechnen Sie für eine beliebige (integrierbare) Zufallsvariable $X : \mathbb{R}^2 \to \mathbb{R}$ die bedingte Erwartung $\mathbb{E}(X|\mathcal{B}_s(\mathbb{R}^2))$ von X unter $\mathcal{B}_s(\mathbb{R}^2)$.

Übungsaufgabe 11 Seien $X_1, X_2, \ldots X_n$ i.i.d. und integrierbar und S_n definiert durch $S_n = X_1 + X_2 + \ldots + X_n$. Berechnen Sie $\mathbb{E}(X_1|S_n) := \mathbb{E}(X_1|\mathcal{A}_\sigma(S_n))$.†

Übungsaufgabe 12 (elementar aber wichtig) Seien X und Y Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Dann sind die folgenden beiden Bedingungen äquivalent:

1. Y ist $\mathcal{A}_\sigma(X)$ messbar.
2. Es existiert eine (Borel) messbare Transformation $h : \mathbb{R} \to \mathbb{R}$ mit $Y = h \circ X$.

Zusatz: Ist Y nicht-negativ, dann kann auch h nicht-negativ gewählt werden. Warum ist dieses einfache Resultat beispielsweise nützlich für Aufgabe 11?

---

†$\mathbb{P}$ ist also das Produktmaß von μ mit sich selbst

†$B^t = \{(x,y) \in \mathbb{R}^2 : (y,x) \in B\}$

†Wie gewohnt, bezeichnet $\mathcal{A}_\sigma(Z) = Z^{-1}(\mathcal{B}(\mathbb{R}))$ die von Z erzeugte σ-Algebra.

03. Übung am 16. April 2018

[LVA 405.552 UV Abhängigkeitsmodellierung, Ankreuzliste siehe www.trutschnig.net/courses mit * versehene Aufgaben sind freiwillig]

Übungsaufgabe 13 Zeigen Sie den in Beispiel 1.22 behaupteten bijektiven Zusammenhang zwischen Markov Kernen und stochastischen Matrizen (Aufwärmübung).

Übungsaufgabe 14 Beweisen Sie Lemma 1.27 (Aufwärmübung).

Übungsaufgabe 15 Sei (X, Y) absolut stetig mit Dichte

$$f(x,y) = \big(1 + \theta(1-2x)(1-2y)\big)\mathbf{1}_{[0,1]^2}(x,y),$$

wobei $\theta \in [-1,1]$. Berechnen Sie eine reguläre bedingte Verteilung $K_{Y,X}$ von Y unter X.

Übungsaufgabe 16 Sei $X \sim \mathcal{U}(0,1)$ und $Y = T \circ X$ mit

$$T(x) = 2x\mathbf{1}_{[0,1/2]}(x) + 2(1-x)\mathbf{1}_{(1/2,1]}(x).$$

Berechnen Sie eine reguläre bedingte Verteilung $K_{Y,X}$ von Y unter X.

Übungsaufgabe 17 Seien X, Y integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Eine (Borel-) messbare Funktion $g : \mathbb{R} \to \mathbb{R}$ heißt Version der bedingten Erwartung von Y gegeben X, genau dann wenn $\mathbb{E}(Y|X)(\omega) = g \circ X(\omega)$ für $\mathbb{P}$-fast-jedes $\omega \in \Omega$ gilt.
Beweisen Sie, dass g genau dann eine Version der bedingten Erwartung von Y gegeben X ist, wenn die folgenden Gleichheit für jedes $B \in \mathcal{B}(\mathbb{R})$ gilt:

$$\int_B g(x)\, d\mathbb{P}^X(x) = \int_{X^{-1}(B)} Y d\mathbb{P}$$

Übungsaufgabe 18 Seien die Voraussetzungen des zweiten Teils von Definition 1.25 erfüllt und K eine reguläre bedingte Erwartung von Y unter $\mathcal{C} = \mathcal{A}_\sigma(X)$. Wie kann aus K eine reguläre bedingte Erwartung $K_{Y,X}$ von Y unter X konstruiert werden?

04. Übung am 23. April 2018

[LVA 405.552 UV Abhängigkeitsmodellierung, Ankreuzliste siehe www.trutschnig.net/courses mit * versehene Aufgaben sind freiwillig]

Übungsaufgabe 19 (Fortsetzung von Aufgabe 16) Sei $X \sim \mathcal{U}(0,1)$ und $Y = T \circ X$ mit

$$T(x) = 2x\mathbf{1}_{[0,1/2]}(x) + 2(1-x)\mathbf{1}_{(1/2,1]}(x).$$

Wir setzen $\mu = \mathbb{P}^{(X,Y)}$ sowie $\Gamma(T) := \{(x,y) \in [0,1]^2 : y = T(x)\}$. Welche Verteilung hat Y? Berechnen Sie weiters (mittels Disintegration oder direkt) $\mu(\Gamma(T))$ und $\mu([0,x] \times [0,y])$.

Übungsaufgabe 20 Sei (X,Y) stetig gleichverteilt am Einheitskreis $B(0,1)$. Die Zufallsvariablen R und Φ seien definiert durch

$$R = \sqrt{X^2+Y^2}, \quad \Phi = \arctan\left(\frac{Y}{X}\right).$$

Berechnen Sie eine reguläre bedingte Verteilung von Φ unter R sowie eine reguäre bedingte Verteilung von R unter Φ.

Übungsaufgabe 21 Beweisen Sie Satz 1.31.

Übungsaufgabe 22 Sei (X,Y) absolut stetig mit Dichte f. Konstruieren Sie eine reguläre bedingte Verteilung $K(\cdot,\cdot)$ von Y unter X, die ausschließlich von f und der Randdichte f_1 von X abhängt.

Übungsaufgabe 23 X und Y seien stetig gleichverteilt auf $[0,1]$, A bezeichne die Verteilungsfunktion von (X,Y), $K(\cdot,\cdot)$ bezeichne eine reguläre bedingte Verteilung von Y unter X. Überlegen Sie sich, wie $K(\cdot,\cdot)$ aus A bestimmt werden kann.
Hinweis: Nehmen Sie vereinfachend an, dass die Abbildung $(x,y) \mapsto \frac{\partial A}{\partial x}(x,y)$ stetig ist.

05. Übung am 30. April 2018

[LVA 405.552 UV Abhängigkeitsmodellierung, Ankreuzliste siehe www.trutschnig.net/courses mit * versehene Aufgaben sind freiwillig]

Übungsaufgabe 24 Bestimmen Sie den Markov Kern und den Markov Operator für jede der in Beispiel 2.3 erwähnten Copulas (Aufwärmübung 1).

Übungsaufgabe 25 Beweisen Sie Lemma 2.4 (Aufwärmübung 2).

Übungsaufgabe 26 Sei $h : [0,1] \to [0,1]$ λ-treu. Zeigen Sie, dass es genau eine Copula $A \in \mathcal{C}$ mit $\mu_A(\Gamma(h)) = 1$ gibt, wobei $\Gamma(h)$, wie gewohnt, den Graphen der Funktion h bezeichnet (Aufwärmübung 3).

Übungsaufgabe 27 Sei $(h_n)_{n\in\mathbb{N}}$ eine Folge λ-treuer Transformationen auf $[0,1]$, die in $L^1([0,1], \mathcal{B}([0,1]), \lambda)$ gegen h konvergiert, $(A_{h_n})_{n\in\mathbb{N}}$ bezeichne die entsprechende Folge von Copulas. Zeigen Sie, dass h ebenfalls λ-treu ist und dass für alle $x, y \in [0,1]$

$$\lim_{n\to\infty} A_{h_n}(x,y) = A_h(x,y)$$

gilt. L^1-Konvergenz der Transformationen impliziert also Konvergenz der Copulas.

Übungsaufgabe 28 Zeigen Sie, dass eine Folge $(A_n)_{n\in\mathbb{N}}$ von Copulas genau dann punktweise gegen eine Copula A konvergiert, wenn sie gleichmäßig gegen A konvergiert.

Übungsaufgabe 29 Wir betrachten die absolut stetige Copula aus Beispiel 2.10 und den Fall $\theta = 1$. Überlegen Sie sich, wie mit Hilfe der bedingten Verteilungsfunktionen $y \mapsto K_A(x, [0,y])$ Stichproben von $(X,Y) \sim A$ erzeugt werden können. Implementieren Sie Ihre Methode in R.