Reinhard Höpfner

Vorlesung Stochastik II

Kapitel X: Reguläre bedingte Wahrscheinlichkeitsverteilungen

Wintersemester 2019/20

Institut für Mathematik, Johannes Gutenberg Universität Mainz

November 4, 2019

Übersicht zu Kapitel X :

A. Bedingte Erwartungen

Definition der bedingten Erwartung, Existenz, Eigenschaften 10.1–10.4

Bedingte Erwartung für nichtnegative messbare numerische Funktionen 10.5

Weitere Eigenschaften 10.6–10.9

Jensen-Ungleichung und L^p-Eigenschaften 10.10–10.11

Interpretation der bedingten Erwartung als Projektion in L^2 10.12

Beispiele und Übungsaufgaben 10.13

B. Faktorisierung der bedingten Erwartung

Faktorisierung der bedingten Erwartung 10.14–10.17

'Bedingte Wahrscheinlichkeit von A gegeben $X = x$': vorläufige Bemerkungen 10.18

C. Reguläre Versionen

Definition: Übergangswahrscheinlichkeit 10.19

Definition: reguläre Versionen der bedingten Wahrscheinlichkeit 10.20–10.21

Beispiel: Bedingen nach einer diskreten ZV 10.22

Beispiel: Paardichten, Randdichten, bedingte Dichten 10.23

Polnische Räume 10.24–10.26

Hauptsatz: reguläre Versionen für bedingte Verteilungen $P^{Y|\mathcal{C}}(\cdot,\cdot)$ 10.27

Korollar: reguläre Versionen der bedingten Wahrscheinlichkeit 10.28

Beweis des Hauptsatzes 10.29

Reguläre Versionen auf der Ebene der Faktorisierungen $P^{Y|X=\cdot}(\cdot)$ 10.30–10.31

Bedingte Erwartung als Integral über die bedingte Verteilung 10.32–10.33

Desintegration von Paarverteilungen 10.34

Bedingte Dichten 10.35

A. Bedingte Erwartungen

10.1 Motivation: Wir illustrieren das Thema dieses Kapitels durch ein Beispiel. Betrachte einen stochastischen Prozess $X := (X_n)_{n \in I\!N_0}$ definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ (d.h. eine Kollektion $\{X_n : n \in I\!N_0\}$ $\mathcal{A}$-messbarer ZV, wobei $n \in I\!N_0$ als 'Zeit' interpretiert wird). Stochastische Prozesse modellieren Zufallsphänomene mit zeitlicher Dynamik (etwa: der Wechselkurs zwischen zwei Währungen am Tag $n \in I\!N_0$ um 12:00 h).

Der Informationsstand eines Beobachters, der den Prozess X in idealer Weise bis zur Zeit n (einschliesslich) beobachtet, wird durch die Sub-σ-Algebra $\mathcal{F}_n := \sigma(X_j : 0 \le j \le n)$ von $\mathcal{A}$ beschrieben: von jedem Ereignis $F \in \mathcal{F}_n$, etwa

$$F := \{\max_{0 \le j \le n} X_j > a\}, \text{ oder } F := \{\sum_{j=1}^{n} f(X_j, X_{j-1}) \le b\}, \text{ oder } F := \{X_n > c\} \; ...$$

kann dieser ideale Beobachter sagen, ob F eingetreten ist oder nicht. So ergibt sich eine wachsende Folge von Sub-σ-Algebren $(\mathcal{F}_n)_{n=0,1,2,...}$ von $\mathcal{A}$, die die zeitliche Entwicklung des Informationsstandes eines idealen Beobachters des Prozesses X modelliert.

Bezeichne $N \in I\!N$ einen 'fernen' Zeithorizont, zu dem gewisse $\mathcal{F}_N$-messbare Zufallsvariable Y, etwa

$$Y := \max_{0 \le j \le N} X_j, \text{ oder } Y := \sum_{n=1}^{N} f(X_j, X_{j-1}), \text{ oder } Y := (X_N - c)^+ \; ...$$

beobachtbar sein werden. Für den Beobachter, der zum 'heutigen' Zeitpunkt $n << N$ nur aufgrund seiner Kenntnis des Prozessverlaufs bis zur Zeit n urteilen kann, sind gute *Prognosen* an die noch ausstehende $\mathcal{F}_N$-messbare Variable Y wichtig. Die im folgenden Satz eingeführte 'bedingte Erwartung von Y gegeben $\mathcal{F}_n$' ist eine solche Prognose.

10.2 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$.
Zu jeder Abbildung $f : \Omega \to I\!R$ mit der Eigenschaft

$$f \;\text{ ist } \mathcal{A}\text{-messbar und in } L^1(P)$$

gibt es eine Abbildung $g = g_f : \Omega \to I\!R$, P-fast sicher eindeutig bestimmt, so dass gilt:
i) g ist $\mathcal{C}$-messbar und in $L^1(P)$,
ii) für jedes $C \in \mathcal{C}$ gilt: $E_P(1_C\, g) = E_P(1_C\, f)$.

Man nennt jede Funktion $g : \Omega \to I\!R$ mit den Eigenschaften i) und ii) eine Festlegung der *bedingten Erwartung von f gegeben $\mathcal{C}$*; man schreibt $E(f|\mathcal{C})$ für die Äquivalenzklasse aller Festlegungen, und nennt diese kurz bedingte Erwartung von f gegeben $\mathcal{C}$.

Beweis: Vor dem Beweis eine Vorbemerkung: Betrachtet man σ-endliche Masse μ, ν auf einem messbaren Raum $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$, so heisst nach Kapitel 3 ν dominiert durch μ $(\nu \ll \mu)$ falls für jedes $A \in \widetilde{\mathcal{A}}$ die Implikation $\mu(A) = 0 \implies \nu(A) = 0$ gilt. Nach dem Satz von Radon-Nikodym 3.5 ist die Aussage $\nu \ll \mu$ gleichwertig mit

es existiert eine $\widetilde{\mathcal{A}}$-messbare Funktion $\widetilde{g} : \Omega \to [0, \infty)$ so dass $\nu(A) = \int_A \widetilde{g}\, d\mu$, $A \in \widetilde{\mathcal{A}}$;

ein solches $\widetilde{g}$ heisst eine Festlegung der μ-Dichte von ν und ist μ-fast sicher eindeutig bestimmt. Die Existenz bedingter Erwartungen wird aus 3.5 hergeleitet werden.

1) Wir zeigen: ist $f \in L^1(\Omega, \mathcal{A}, P)$ und ist $g : \Omega \to I\!R$ eine $\mathcal{C}$-messbare Funktion mit Eigenschaft ii), so gilt notwendig $g \in L^1(P)$. Nach Zerlegen $g = g^+ - g^-$ in Positiv- und Negativteil hat man

$$
\begin{aligned}
0 \;\le\; E_P(g^+) \;\le\; E_P(1_{\{g>0\}} g) \;&\stackrel{ii)}{=}\; E_P(1_{\{g>0\}} f) \;\le\; E_P(1_{\{g>0\}} |f|) \;<\; \infty \\
0 \;\le\; E_P(g^-) \;\le\; E_P(1_{\{g<0\}} (-g)) \;&\stackrel{ii)}{=}\; E_P(1_{\{g<0\}} (-f)) \;\le\; E_P(1_{\{g<0\}} |f|) \;<\; \infty
\end{aligned}
$$

und aus $|g| = g^+ + g^-$ folgt $g \in L^1(P)$.

2) Wir zeigen: Aus $f \ge 0$ folgt $g \ge 0$ P-fast sicher, für jedes g mit den Eigenschaften i) und ii). Wegen $\{g < 0\} \in \mathcal{C}$ und wegen ii) hat man zugleich

$$
1_{\{g<0\}}\, g \;\le\; 0 \,, \quad E(\, 1_{\{g<0\}}\, g\,) \stackrel{ii)}{=} E(\, 1_{\{g<0\}}\, f\,) \,, \quad f \;\ge\; 0 \,,
$$

folglich muss $\{g < 0\}$ eine P-Nullmenge sein.

3) Wir zeigen: ist $f \in L^1(\Omega, \mathcal{A}, P)$, so stimmen verschiedene Funktionen $g : \Omega \to I\!R$ mit den Eigenschaften i) und ii) notwendig P-fast sicher überein:
Seien g, $\widetilde{g}$ Funktionen $\Omega \to I\!R$ mit den Eigenschaften i) und ii). Angenommen, $P(g \ne \widetilde{g}) > 0$. Dann wäre etwa $P(\widetilde{g} > g) > 0$, und man könnte ein $r > 0$ finden so dass

$$
C_r := \{\widetilde{g} > g{+}r\} \in \mathcal{C} \,, \quad P(C_r) > 0 \,, \quad E_P(\, 1_{C_r}\, \widetilde{g}\,) \stackrel{ii)}{=} E_P(\, 1_{C_r}\, f\,) \stackrel{ii)}{=} E_P(\, 1_{C_r}\, g\,)
$$

zugleich mit

$$
E_P(\, 1_{C_r}\, \widetilde{g}\,) \;>\; E_P(\, 1_{C_r}\, (g + r)\,) \;=\; E_P(\, 1_{C_r}\, g\,) + r\, P(C_r) \;>\; E_P(\, 1_{C_r}\, g\,)
$$

gilt, ein Widerspruch. Also gilt $g = \widetilde{g}$ P-fast sicher.

4) Es bleibt die Existenz eines g mit den Eigenschaften i) und ii) zu zeigen. Nach Zerlegung von $f \in L^1(\Omega, \mathcal{A}, P)$ in Positiv- und Negativteil genügt ein Existenzbeweis für nichtnegatives $f \in L^1(\Omega, \mathcal{A}, P)$. Dieses definiert (vgl. 3.1) ein endliches Mass Q_f auf $(\Omega, \mathcal{A})$

$$\mathcal{A} \ni A \longrightarrow Q_f(A) := \int_A f \, dP \in [0, \infty)$$

mit $Q_f \ll P$. Bezeichne $\widetilde{Q}_f, \widetilde{P}$ die Restriktionen von Q_f, P auf die Sub-σ-Algebra $\mathcal{C}$. In Einschränkung auf $\mathcal{C}$ gilt wieder $\widetilde{Q}_f \ll \widetilde{P}$, also gibt es nach dem Satz von Radon-Nikodym 3.5 eine $\mathcal{C}$-messbare Festlegung $g : \Omega \to [0, \infty)$ der Dichte $\frac{d\widetilde{Q}_f}{d\widetilde{P}}$. Damit gilt für beliebige Ereignisse $C \in \mathcal{C}$

$$\int_C g \, dP = \int_C g d\widetilde{P} = \widetilde{Q}_f(C) = Q_f(C) = \int_C f \, dP \, .$$

Also ist zu nichtnegativem $f \in L^1(\Omega, \mathcal{A}, P)$ ein $g \in L^1(\Omega, \mathcal{C}, P)$ mit ii) gefunden. $\square$

10.3 Hilfssatz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$.

a) Betrachte Funktionen $f, f_1, f_2, \ldots \in L^1(\Omega, \mathcal{A}, P)$, seien $g, g_1, g_2, \ldots$ Festlegungen der bedingten Erwartungen $E(f|\mathcal{C}), E(f_1|\mathcal{C}), E(f_2|\mathcal{C}), \ldots$, dann gilt

i) (Linearität) für $\alpha_1, \alpha_2 \in I\!R$ ist $\alpha_1 g_1 + \alpha_2 g_2$ eine Festlegung von $E(\alpha_1 f_1 + \alpha_2 f_2|\mathcal{C})$;

ii) (Monotonie) aus $f_1 \leq f_2$ P-fast sicher folgt $g_1 \leq g_2$ P-fast sicher;

iii) (aufsteigende Stetigkeit) für $(f_n)_n$ mit $f_n \geq 0$ und $f_n \uparrow f$ für $n \to \infty$ gilt

$$g_n \uparrow g \quad P\text{-fast sicher für } n \to \infty \, .$$

b) Ist f schon $\mathcal{C}$-messbar, so ist f eine Festlegung von $E(f|\mathcal{C})$.

c) Ist $f = \alpha$ P-fast sicher konstant, so ist $g \equiv \alpha$ eine Festlegung von $E(f|\mathcal{C})$.

Beweis: 1) Die Aussagen a)i) + b) erhält man sofort aus der Definition der bedingten Erwartung, und c) folgt aus b).

2) Aus $f_1 \leq f_2$ P-fast sicher folgt mit a)i)

$$\forall \, C \in \mathcal{C} : \quad 0 \leq E_P(1_C(f_2 - f_1)) = E_P(1_C(g_2 - g_1)) \, .$$

Jetzt betrachtet man Ereignisse $C_m := \{g_2 - g_1 \leq -\frac{1}{m}\} \in \mathcal{C}$: die letzte Gleichung erzwingt $P(C_m) = 0$ für jedes m, also ist $N := \{g_2 - g_1 < 0\}$ eine P-Nullmenge. Dies zeigt a)ii).

Beachte: Wegen $N \in \mathcal{C}$ sind mit g_i auch $\widetilde{g}_i := g_i \, 1_{N^c}$ Festlegungen von $E(f_i|\mathcal{C})$, $i = 1, 2$, und für diese gilt $\widetilde{g}_1 \leq \widetilde{g}_2$ auf ganz Ω.

3) Betrachte $f_n \geq 0$, $f_n \uparrow f$, alle in $L^1(\Omega, \mathcal{A}, P)$. Seien g, g_n, $n \geq 1$ Festlegungen von $E(f|\mathcal{C})$, $E(f_n|\mathcal{C})$, $n \geq 1$. Wie in der letzten Bemerkung in Beweisschritt 2) gibt es dann auch Festlegungen $\widetilde{g}_n$ von

$E(f_n|\mathcal{C})$ in aufsteigender Anordnung $0 \le \widetilde{g}_1 \le \widetilde{g}_2 \le \ldots$: Mit g_1 ist auch $\widetilde{g}_1 := g_1 \vee 0$ eine Festlegung von $E(f_1|\mathcal{C})$, wegen $g_1 \ge 0$ P-fast sicher; mit g_2 ist auch $\widetilde{g}_2 := g_2 \vee \widetilde{g}_1$ eine Festlegung von $E(f_2|\mathcal{C})$, wegen $g_2 \ge \widetilde{g}_1$ P-fast sicher, usw ..., schliesslich definiert man

$$\widetilde{g} \;:=\; \sup_n \, \widetilde{g}_n \;:\; \Omega \to [0,\infty] \quad \mathcal{C}\text{-messbar.}$$

Monotone Konvergenz zeigt dann für jedes $C \in \mathcal{C}$:

$$E_P(1_C\widetilde{g}) \;\;=\;\; \lim_{n\to\infty} E_P(1_C\widetilde{g}_n) \;=\; \lim_{n\to\infty} E_P(1_C f_n) \;=\; E_P(1_C f) \;=\; E_P(1_C g)\;.$$

Da $C \in \mathcal{C}$ beliebig war, folgt $\widetilde{g} = g$ P-fast sicher. Bis auf die P-Nullmenge

$$N := \{g \neq \widetilde{g}\} \cup \bigcup_{n\ge 1} \{g_i \neq \widetilde{g}_i\} \in \mathcal{C}$$

stimmen die verschiedenen Festlegungen überein, und es gilt $g_n(\omega) \uparrow g(\omega)$ für $\omega \notin N$. □

10.4 Bemerkung: Die Aussagen a)i)–iii) von 10.3 zeigen, dass die Zuordnung $f \;\to\; E(f|\mathcal{C})$ die Eigenschaften eines Integrals besitzt (der Grund wird später klar werden, siehe 10.32–10.33). Es ist aber wichtig zu betonen, dass *bedingte Erwartungen* $E(\ldots|\mathcal{C})$ *$\mathcal{C}$-messbare Funktionen* sind (bzw: P-Äquivalenzklassen von $\mathcal{C}$-messbaren Funktionen), und nicht etwa reelle Zahlen, wie es die sprachliche Nähe zu 'Erwartungswert' suggerieren könnte.

10.5 Bemerkung: Ähnlich wie in Kapitel 2 kann man mit Hilfe von 10.3 a)iii) bedingte Erwartungen auch für *nichtnegative* $\mathcal{A}$-messbare Abbildungen f definieren, unter Verzicht auf die Eigenschaft $L^1(P)$:

a) Wir betrachten zuerst nichtnegative MNF $f \in \mathcal{F}^+$: zu jedem $\mathcal{A}$-messbaren $f : \Omega \to [0,\infty]$ gibt es ein $\mathcal{C}$-messbares $g = g_f : \Omega \to [0,\infty]$ so dass gilt

$$(*) \qquad\qquad \text{für jedes } C \in \mathcal{C}\text{:}\;\; E_P(\,1_C\, g\,) \;=\; E_P(\,1_C\, f\,) \;\le\; +\infty\;.$$

Zum Beweis reicht es, für $n \ge 1$ Festlegungen g_n der bedingten Erwartung $E(f\wedge n|\mathcal{C})$ nach Satz 10.2 zu wählen; diese bilden nach Abänderung auf P-Nullmengen wie im Beweis von 10.3 a)iii) eine aufsteigende Folge $(g_n)_n$. Man setzt $g := \sup_{n\ge 1} g_n$ und erhält die Behauptung wegen $f\wedge n \uparrow f$, $g_n \uparrow g$ aus dem Satz von der monotonen Konvergenz 2.7.

b) Für $\mathcal{A}$-messbares $f : \Omega \to [0,\infty)$ kann die bedingte Erwartung $E_P(f|\mathcal{C})$ als $\mathcal{C}$-messbare Funktion $g : \Omega \to [0,\infty)$ festgelegt werden:

Zum Beweis betrachte f, g wie in a); sei zusätzlich $\{f = +\infty\}$ eine P-Nullmenge. Dann definiert

$$C \longrightarrow \mu(C) := E_P(1_C f) \overset{(*)}{=} E_P(1_C g) \quad , \quad C \in \mathcal{C}$$

nach 3.1 c) ein σ-endliches Mass auf der Sub-σ-Algebra $\mathcal{C}$, folglich ist $\{g = +\infty\}$ eine P-Nullmenge.

c) Für $f : \Omega \to [0, \infty)$ $\mathcal{A}$-messbar und $g : \Omega \to [0, \infty)$ $\mathcal{C}$-messbar zeigt der Eindeutigkeitssatz 1.13: es genügt, die Eigenschaft $(*)$ für Mengen C aus einem $\cap$-stabilen Erzeuger $\mathcal{S}$ von $\mathcal{C}$ nachzuprüfen, in dem es eine Mengenfolge $(C_n)_n \subset S$ mit $C_n \uparrow \Omega$ und $E_P(1_{C_n} f) < \infty$ gibt. $\square$

10.6 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$, $f \in L^1(\Omega, \mathcal{A}, P)$, sei $g : \Omega \to I\!R$ eine $\mathcal{C}$-messbare Funktion. Folgende Aussagen sind gleichwertig:

i) g ist eine Festlegung von $E(f|\mathcal{C})$;

ii) für alle $\mathcal{C}$-messbaren beschränkten Funktionen $h : \Omega \to I\!R$ gilt $E(hg) = E(hf)$.

Beweis: ii) $\Longrightarrow$ i): Zunächst ist nach Schritt 1) des Beweises von 10.2 jedes $\mathcal{C}$-messbare g mit der Eigenschaft ii) notwendig in $L^1(P)$. Danach ist Aussage i) derjenige Spezialfall von ii), in dem man nur Indikatorfunktionen $h = 1_C$ betrachtet, $C \in \mathcal{C}$.
i) $\Longrightarrow$ ii): Es genügt ein Nachweis für $f \geq 0$ (sonst zerlege $f = f^+ - f^-$ in Positiv- und Negativteil); wie in Beweisschritt 2) von 10.2 hat man dann auch $g \geq 0$, ggfls. nach Abänderung von g auf P-Nullmengen. Dann aber reicht es, auch nur den Fall $h \geq 0$ zu betrachten. Approximiere $h \geq 0$ monoton durch $\mathcal{C}$-messbare Elementarfunktionen

$$h_n := \sum_{j=1}^{n2^n} \frac{j-1}{2^n} 1_{\{\frac{j-1}{2^n} \leq h < \frac{j}{2^n}\}} + n\, 1_{\{h \geq n\}} \quad , \quad n \geq 1 ,$$

dann gilt unter i) nach Definition der bedingten Erwartung

$$E_P(h_n g) = E_P(h_n f) \quad \forall\, n \geq 1$$

woraus mit monotoner Konvergenz die Aussage ii) folgt. $\square$

Ab jetzt unterscheiden wir in den Schreibweisen häufig nicht mehr zwischen bedingten Erwartungen (als Äquivalenzklasse von Festlegungen) und den Festlegungen selbst, wir sparen oft auch das zwischen verschiedenen Festlegungen derselben bedingten Erwartung geltende 'P-fast sicher', und schreiben kurz $g \in E_p(f|\mathcal{C})$ falls g eine Festlegung der bedingten Erwartung von f gegeben $\mathcal{C}$ ist. Als nächstes formulieren wir einige wichtige – und intuitiv einleuchtende – Eigenschaften dieser 'besten Prognose' für $f \in L^1(\Omega, \mathcal{A}, P)$ aufgrund des partiellen Kenntnisstandes $\mathcal{C}$.

10.7 Bemerkung: Bedingte Erwartungen können sukzessiv berechnet werden: es gilt

$$E(\, f | \, \mathcal{C}_1 \,) \;=\; E(\; E(f|\mathcal{C}_2) \,|\, \mathcal{C}_1 \,)$$

für Sub-σ-Algebren $\mathcal{C}_1 \subset \mathcal{C}_2$ von $\mathcal{A}$ und für $f \in L^1(\Omega, \mathcal{A}, P)$. □

10.8 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $f \in L^1(\Omega, \mathcal{A}, P)$, $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$. Dann gilt für alle $\mathcal{C}$-messbaren und beschränkten Funktionen $h : \Omega \to I\!R$

$$E(\, hf \mid \mathcal{C}) \;=\; h \cdot E(f \mid \mathcal{C}) \;.$$

Beweis: Sei g eine Festlegung von $E_P(f|\mathcal{C})$. Insbesondere ist $g \in L^1(P)$. Damit ist $h \cdot g$ $\mathcal{C}$-messbar und – da h als beschränkt vorausgesetzt wurde– in $L^1(P)$, und wegen 10.6 gilt für jedes feste $C \in \mathcal{C}$

$$E_P(1_C \cdot (h \cdot f)) \;=\; E_P((1_C \cdot h) \cdot f) \;=\; E_P((1_C \cdot h) \cdot g) \;=\; E_P(1_C \cdot (h \cdot g)) \;;$$

also ist $h \cdot g$ eine Festlegung von $E_P(hf|\mathcal{C})$. □

10.8' Bemerkung: Die Aussagen von 10.6, 10.7 und 10.8 gelten entsprechend, wenn anstelle von $f \in L^1(P)$ bedingte Erwartungen für $f \in \mathcal{F}^+$ wie in 10.5 betrachtet werden. □

10.9 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$.
Ist $f \in L^1(\Omega, \mathcal{A}, P)$ unabhängig von $\mathcal{C}$, so ist die konstante Funktion

$$g \equiv E_P(f) \;:\; \Omega \to I\!R$$

eine Festlegung von $E(f|\mathcal{C})$.

Beweis: Schreibe $\alpha := E_P(f)$. Wenn f unabhängig von $\mathcal{C}$ ist, gilt für jedes $C \in \mathcal{C}$

$$E_P(1_C f) \;=\; E_P(1_C) \cdot E_P(f) \;=\; E_P(1_C \cdot \alpha) \;.$$ □

10.10 Jensen-Ungleichung: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{C} \subset \mathcal{A}$ eine Sub-σ-Algebra. Sei $I \subset I\!R$ ein Interval und $G : I \to I\!R$ eine konvexe Funktion. Ist $f \in L^1(\Omega, \mathcal{A}, P)$ P-fast sicher I-wertig, so ist auch $E_P(f|\mathcal{C})$ P-fast sicher I-wertig, und es gilt

$$-\infty \;<\; G\,(\,E(\,f\,|\,\mathcal{C}\,)\,) \;\leq\; E\,(\,G \circ f \,|\, \mathcal{C}\,) \;\;\leq\; +\infty \;.$$

Bemerkung: Für $\mathcal{C} = \{\Omega, \emptyset\}$ reduziert sich die bedingte Erwartung auf eine auf Ω konstante Funktion, und man findet die klassische Jensen-Ungleichung $G(E(f)) \le E(G \circ f))$ (vgl. 2.13') wieder.

Beweis: Da $L^1(\Omega, \mathcal{A}, P)$ aus Äquivalenzklassen von P-fast sicher übereinstimmenden Funktionen besteht, können wir –gegebenenfalls nach Abänderung von f auf einer P-Nullmenge– voraussetzen, dass $f(\omega) \in I$ für alle $\omega \in \Omega$ gilt (ersetze sonst f z.B. durch $f\,1_{N^c} + x_0 1_N$, mit $N := \{f \notin I\} \in \mathcal{A}$ und 'default value' $x_0 \in I$).

1) Wir zeigen zuerst: es gibt eine Festlegung g von $E(f|\mathcal{C})$ mit $g(\omega) \in I$ für alle $\omega \in \Omega$:

Betrachte zuerst den linken Rand $\alpha := \inf\{v : v \in I\}$ des Intervalls I. Falls $\alpha = -\infty$, ist hier nichts zu beweisen. Sei also $\alpha > -\infty$. Für $f \ge \alpha$ gibt es eine Festlegung $\widetilde{g}$ von $E(f|\mathcal{C})$ mit $\widetilde{g} \ge \alpha$ auf ganz Ω. Falls $\alpha \in I$, reicht das. Falls $\alpha \notin I$, hat man $f > \alpha$ auf Ω. Betrachte $C := \{\widetilde{g} = \alpha\} \in \mathcal{C}$, dann muss gelten $E(1_C(\widetilde{g} - \alpha)) = E(1_C(f - \alpha))$, also ist C eine P-Nullmenge. Mit $g := \widetilde{g}1_{C^c} + x_0 1_C$ für ein festes $x_0 \in I$ ist eine Festlegung g von $E(f|\mathcal{C})$ mit $g(\omega) > \alpha$ für alle $\omega \in \Omega$ gefunden. In ähnlicher Weise arbeitet man am rechten Rand $\beta := \sup\{v : v \in I\}$ des Intervalls I.

2) Weiter reicht es, nichtnegative konvexe Funktionen $G : I \to [0, \infty)$ zu betrachten:

Sei $G : I \to I\!R$ konvex, sei $v \in I$, sei $L(x) = ax + b$ eine Stützgerade für G durch den Punkt $(v, G(v))$. Definiere $\widetilde{G}(y) := G(y) - L(y)$, $y \in I$. Dann ist $\widetilde{G}$ konvex und nichtnegativ, also $\widetilde{G} \circ f \in \mathcal{F}^+$. Nach 10.5 sind Festlegungen von $E(\widetilde{G}\circ f|\mathcal{C})$ als $[0, \infty]$-wertige Zufallsvariable wohldefiniert. Mit $f \in L^1(P)$ gilt $L \circ f \in L^1(P)$, und Festlegungen von $E(L\circ f|\mathcal{C})$ nach 10.2 sind reellwertig. Als Summe beider Terme ist

$$-\infty \;<\; E\,(\,G \circ f\,|\,\mathcal{C}\,) \;:=\; E\left(\widetilde{G}\circ f|\mathcal{C}\right) + E(L\circ f|\mathcal{C}) \;\le\; +\infty$$

wohldefiniert, und die Behauptung des Satzes ist gleichwertig zur Aussage

$$0 \;\le\; \widetilde{G}\,(\,E(\,f\,|\,\mathcal{C}\,)\,) \;\le\; E\left(\,\widetilde{G} \circ f\,|\,\mathcal{C}\,\right) \;\le\; +\infty\,.$$

Also reicht es, die Behauptung für $\widetilde{G}$ anstelle von G zu beweisen. Betrachte also ab jetzt nichtnegative konvexe Funktionen G.

3) Jedes konvexe nichtnegative $G : I \to [0, \infty)$ ist aufsteigender Limes einer Folge geeigneter Polygonzüge, und damit obere Einhüllende einer *abzählbaren* Familie von Stützgeraden

$$L_n(x) = a_n x + b_n\,,\; n = 1, 2, \ldots\;; \quad G(x) = \sup\{L_n(x) : n \ge 1\}\,,\; x \in I\,.$$

Mit dem Argument aus Schritt 2) (und wegen 10.5 a) für $(G - L_n) \in \mathcal{F}^+$) ist für jedes n

$$+\infty \;\ge\; E(G \circ f \mid \mathcal{C}) \;:=\; E((G - L_n)\circ f \mid \mathcal{C}) + E(L_n\circ f|\mathcal{C}) \;\ge\; E(L_n \circ f|\mathcal{C}) \;=\; a_n\, g + b_n$$

wohldefiniert, damit gilt für alle $\omega \in \Omega$ (gegebenenfalls nach Abänderung der gewählten Festlegung von $E\,(G \circ f \mid \mathcal{C})$ auf einer abzählbaren Familie von P-Nullmengen in $\mathcal{C}$)

$$+\infty \;\geq\; E\,(G \circ f \mid \mathcal{C})\,(\omega) \;\geq\; \sup_{n=1,2,\dots} \,(a_n\, g(\omega) + b_n) \;=\; G\,(\,g(\omega)\,)\ .$$

Damit ist die Behauptung bewiesen. □

10.11 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C} \subset \mathcal{A}$ eine Sub-σ-Algebra, $p \geq 1$. Dann

$$f \in L^p(\Omega, \mathcal{A}, P) \implies E(f|\mathcal{C}) \in L^p(\Omega, \mathcal{C}, P)\ .$$

Beweis: Da $p \geq 1$, ist das die Jensen-Ungleichung 10.9 mit $G(x) = |x|^p$, $x \in I\!R$. □

Satz 10.11 mit $p = 2$ erlaubt unter der Voraussetzung $f \in L^2(\Omega, \mathcal{A}, P)$, die als 'bedingte Erwartung gegeben eine Sub-σ-Algebra $\mathcal{C}$' eingeführte Prognose $E_P(f|\mathcal{C})$ ganz klassisch als Orthogonalprojektion von f auf den linearen Teilraum $L^2(\Omega, \mathcal{C}, P)$ (vgl. 2.18) aufzufassen.

10.12 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C} \subset \mathcal{A}$ eine Sub-σ-Algebra, betrachte Funktionen $f \in L^2(\Omega, \mathcal{A}, P)$.
Dann sind 'Prognose' $E(f|\mathcal{C})$ und 'Prognosefehler' $f - E(f|\mathcal{C})$ orthogonal in $L^2(P)$.

Beweis: Wähle eine Festlegung g von $E(f|\mathcal{C})$. Nach 10.11 gilt $g \in L^2(\Omega, \mathcal{C}, P)$. Für $\mathcal{C}$-messbare beschränkte Funktionen $h : \Omega \to I\!R$ zeigt 10.6

$$E\,((f-g)h) \;=\; E\,(fh) - E\,(gh) \;=\; 0\ ,$$

daraus folgt für beliebiges $h \in L^2(\Omega, \mathcal{C}, P)$ mit dominierter Konvergenz

$$\infty \;>\; E\,((f-g)h) \;=\; \lim_{n\to\infty} E\,(\,(f-g)\,[(-n) \vee h \wedge (+n)]\,) \;=\; 0\ .$$

Damit steht $(f - g)$ senkrecht auf dem ganzen $L^2(\Omega, \mathcal{C}, P)$, insbesondere also senkrecht auf $g \in L^2(\Omega, \mathcal{C}, P)$, und es gibt in $L^2(\Omega, \mathcal{C}, P)$ keine bessere Prognose an f als g. □

10.13 Übungsaufgaben und Beispiele: a) Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $f \in L^1(\Omega, \mathcal{A}, P)$. Für jedes Ereignis $A \in \mathcal{A}$ mit $P(A) > 0$ kann man einen *elementar bedingten Erwartungswert von f gegeben A* als

$$E(f|A) := \frac{E(1_A f)}{P(A)}$$

definieren; dies entspricht per Aufbau messbarer Funktionen der in der 'Einführung in die Stochastik' betrachteten elementar bedingten Wahrscheinlichkeit

$$P(B|A) := \frac{P(B \cap A)}{P(A)} \ , \quad B \in \mathcal{A} \ .$$

Ist nun $(A_n)_n$ eine Zerlegung von Ω in disjunkte $\mathcal{A}$-Mengen von positivem Mass

$$(A_n)_n \subset \mathcal{A} \ , \quad \Omega = \dot{\bigcup_{n \geq 1}} A_n \ , \quad P(A_n) > 0 \quad \forall \, n$$

und betrachtet man als Sub-σ-Algebra von $\mathcal{A}$

$$\mathcal{C} \ = \ \sigma(A_n : n \geq 1) \ ,$$

so ist die nach 10.2 gebildete bedingte Erwartung von f gegeben $\mathcal{C}$ gegeben durch

$$g \ : \quad \Omega \ni \omega \ \to \sum_{n \geq 1} c_n \, 1_{A_n}(\omega) \ \in \mathbb{R} \ , \quad c_n := \frac{E(1_{A_n} f)}{P(A_n)} \ , \ n \geq 1 \ .$$

Diese $\mathcal{C}$-messbare Funktion wählt bei Vorliegen von $\omega \in A_n$ als 'beste Prognose an f' den elementar bedingten Erwartungswert $E(f|A_n)$ von f gegeben A_n.

b) Betrachte $(\Omega, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, bezeichne $\mathcal{S}_d$ die Gruppe aller Permutationen π auf $\{1, \ldots, d\}$; schreibe $\pi x := (x_{\pi_1}, \ldots, x_{\pi_d})$ für $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Die Sub-$\sigma$-Algebra aller permutationsinvarianten Borelmengen in $\mathbb{R}^d$ ist

$$\mathcal{C} := \{ A \in \mathcal{B}(\mathbb{R}^d) \ : \ \pi A = A \quad \forall \, \pi \in \mathcal{S}_d \}$$

Für jedes permutationsinvariante Wahrscheinlichkeitsmass Q auf $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, d.h.:

$$\text{für alle } A \in \mathcal{B}(\mathbb{R}^d) \text{ gilt:} \quad Q(A) \ = \ Q(\widetilde{\pi} A) \quad \forall \, \widetilde{\pi} \in \mathcal{S}_d \ ,$$

und für jedes nichtnegative $\mathcal{A}$-messbare $f : \Omega \to \mathbb{R}$ ist dann die Symmetrisierte von f

$$g \ := \ \frac{1}{d\,!} \sum_{\pi \in \mathcal{S}_d} f \circ \pi$$

eine Festlegung der bedingten Erwartung $E_Q(f|\mathcal{C})$ im Sinne von 10.5. Beachte: permutationsinvariant auf $\mathbb{R}^d$ sind insbesondere alle Produktmasse $\bigotimes_{j=1}^{d} \nu$, ν ein Wahrscheinlichkeitsmass auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. □

B. Faktorisierung der bedingten Erwartung

10.14 Definition: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$. Wird $\mathcal{C}$ von einer Zufallsvariable X erzeugt

$$X : \ (\Omega, \mathcal{A}) \to (T, \mathcal{T}) \ \text{ messbar} \ , \quad \mathcal{C} := \sigma(X) = X^{-1}(\mathcal{T}) \ ,$$

so schreibt man kurz $E_P(f|X)$ für die in 10.2 oder in 10.5 definierte Äquivalenzklasse $E_P(f|\sigma(X))$ $\sigma(X)$-messbarer Funktionen, für $f \in L^1(\Omega, \mathcal{A}, P)$ oder für $f \in \mathcal{F}^+$, und spricht von (einer Festlegung) der *bedingten Erwartung von f gegeben X*.

Beachte: für $f \in L^1(\Omega, \mathcal{A}, P)$ ist $E(f|X)$ eine $\sigma(X)$*-messbare Funktion*. Das Faktorisierungslemma 1.43 erlaubt eine griffige Formulierung:

10.15 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $(T, \mathcal{T})$ ein messbarer Raum, sei $X : (\Omega, \mathcal{A}) \to (T, \mathcal{T})$ messbar. Sei $f \in L^1(\Omega, \mathcal{A}, P)$. Für jede Festlegung g von $E_P(f|X)$ gibt es eine messbare Funktion $h_g : (T, \mathcal{T}) \to (I\!R, \mathcal{B}(I\!R))$ mit

$$g \;=\; h_g \circ X \;.$$

Man nennt jede messbare Funktion $h : (T, \mathcal{T}) \to (I\!R, \mathcal{B}(I\!R))$ mit der Eigenschaft

$$h \circ X \quad \text{ist eine Festlegung von} \quad E(f|X)$$

eine *Faktorisierung der bedingten Erwartung von f gegeben X* bzw. eine Festlegung der *bedingten Erwartung von f gegeben $X = \cdot$* . Man schreibt $E(f|X = \cdot)$ für jede solche Funktion $h(\cdot)$; verschiedene Festlegungen $h, \widetilde{h}$ von $E(f|X = \cdot)$ stimmen P^X-fast sicher überein.

Beweis: Sofort aus dem Faktorisierungslemma 1.43: genau dann ist $g : \Omega \to I\!R$ $\sigma(X)$-messbar, wenn es eine $\mathcal{T}$-messbare Funktion $h_g : T \to I\!R$ gibt mit $g = h_g \circ X$. $\square$

10.16 Definition: Mit den Voraussetzungen und Bezeichnungen aus 10.15 sei

$$h : (T, \mathcal{T}) \to (I\!R, \mathcal{B}(I\!R)) \;\; \text{eine Festlegung von} \;\; E(f|X = \cdot).$$

Für $x \in T$ nennt man den *Funktionswert* $h(x) \in I\!R$ *bedingte Erwartung von f gegeben $X = x$*, und schreibt $E(f|X = x)$ für $h(x)$.

10.17 Bemerkung: a) An einer festen Stelle x ist – aufgrund der Tatsache, dass verschiedene Festlegungen $h, \widetilde{h}$ von $E(f|X = \cdot)$ nur P^X-fast sicher übereinstimmen – hiermit i.a. überhaupt nichts festgelegt, es sei denn, man habe $\{x\} \in \mathcal{T}$ und $P^X(\{x\}) > 0$: die zweite Bedingung aber ist in den meisten interessanten Situationen (mit Ausnahme diskreter ZV X, siehe b) unten, oder 10.22) nicht erfüllt. Man muss also h stets als Funktion betrachten und kann von einer bedingten Erwartung von

f gegeben $X = x$ nur für P^X-fast alle x sinnvoll reden.

b) Sei X eine diskrete reellwertige ZV auf $(\Omega, \mathcal{A}, P)$, sei S (endlich oder abzählbar) die Menge der mit positiver Wahrscheinlichkeit eintretenden Werte von X. Dann liefert mit 10.13 a)

$$h(x) \;:=\; E(f \mid \{X = x\}) \;=\; \frac{E(f\, 1_{\{X=x\}})}{P(\{X = x\})} \quad \text{falls } x \in S\,, \quad h(x) := 0 \;\text{ sonst}$$

eine Festlegung $h(\cdot)$ von $E(f|X = \cdot)$ gemäss 10.15. □

10.18 Bemerkung: Wir betrachten den Spezialfall von Indikatorfunktionen f. Sei $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$, z.B. von Form $\mathcal{C} = \sigma(X)$ wie in 10.14. Für $f = 1_A$, $A \in \mathcal{A}$ fest, hat man

$$\begin{aligned} P(A|\mathcal{C}) \;&:=\; E(1_A|\mathcal{C}) \quad \textit{bedingte Wahrscheinlichkeit von } A \textit{ gegeben } \mathcal{C} \\ P(A|X) \;&:=\; E(1_A|X) \quad \textit{bedingte Wahrscheinlichkeit von } A \textit{ gegeben } X \end{aligned}$$

als $\mathcal{C}$- bzw. $\sigma(X)$-messbare Funktionen $\Omega \to I\!R$, sowie

$$P(A|X = \cdot) \;:=\; E(1_A|X = \cdot) \quad \textit{bedingte Wahrscheinlichkeit von } A \textit{ gegeben } X = \cdot$$

auf dem Faktorisierungslevel als $\mathcal{T}$-messbare Funktion, und für festes $x \in T$ ist

$$P(A|X = x) := E(1_A|X = x) \quad \textit{bedingte Wahrscheinlichkeit von } A \textit{ gegeben } X = x$$

eine reelle Zahl. Aber Vorsicht: die Abbildungen

$$\begin{aligned} \mathcal{A} \;\ni\; A \;&\to\; P(A|\mathcal{C})(\omega) \;\in I\!R\;, \quad \omega \in \Omega \text{ fest} \\ \mathcal{A} \;\ni\; A \;&\to\; P(A|X = x) \;\in I\!R\;, \quad x \in T \text{ fest} \end{aligned}$$

sind im allgemeinen – d.h. ohne Zusatzvoraussetzungen an die σ-Algebra $\mathcal{A}$ – noch *keine Wahrscheinlichkeitsmasse* auf $\mathcal{A}$, aus dem folgenden Grund:

zwar gilt nach 10.3 für jeden gegebenen Satz von Mengen A, $A_1, A_2, \ldots$ in $\mathcal{A}$

$$(*) \quad \left\{ \begin{array}{lll} i) & 0 \le P(A|\mathcal{C}) \le 1 \quad P\text{-fast sicher} & \\ ii) & P(\emptyset|\mathcal{C}) = 0 \quad P\text{-fast sicher}, \qquad P(\Omega|\mathcal{C}) = 1 \quad P\text{-fast sicher} & \\ iii) & A_i,\ i \ge 1 \text{ paarweise disjunkt}: \quad P(\dot{\bigcup}_{n=1}^{\infty} A_n \mid \mathcal{C}) = \sum_{n=1}^{\infty} P(A_n|\mathcal{C}) \quad P\text{-fast sicher} & \end{array} \right. ,$$

aber die Ausnahmenullmengen in den Aussagen *i)* und *iii)* von $(*)$ hängen jeweils ab von den betrachteten Mengen A, $A_1, A_2, \ldots$, und können *i.a. nicht* zu einer einzigen grossen P-Nullmenge

$N \in \mathcal{A}$ zusammengefasst werden, bezüglich der $(*)$ für beliebige Wahl von Mengen A und Mengenfolgen $A_1, A_2, \ldots$ in $\mathcal{A}$ gültig wäre. Erst der Begriff einer 'regulären Version' einer bedingten Wahrscheinlichkeit oder bedingten Verteilung wird hier Abhilfe schaffen. $\square$

C. Reguläre Versionen

10.19 Definition: Seien $(\Omega_1, \mathcal{A}_1)$ und $(\Omega_2, \mathcal{A}_2)$ messbare Räume. *Übergangswahrscheinlichkeit* (oder *Markov-Kern*, oder einfach *Kern*) *von* $(\Omega_1, \mathcal{A}_1)$ *nach* $(\Omega_2, \mathcal{A}_2)$ nennt man eine Abbildung

$$K(\cdot,\cdot) \; : \quad \Omega_1 \times \mathcal{A}_2 \ni \; (\omega, A) \; \to \; K(\omega, A) \; \in [0,1]$$

mit den Eigenschaften

i) $\forall\, A \in \mathcal{A}_2$ fest : $\quad \omega \to K(\omega, A)$ ist $\mathcal{A}_1$-messbar,

ii) $\forall\, \omega \in \Omega_1$ fest : $\quad A \to K(\omega, A)$ ist ein Wahrscheinlichkeitsmass auf $(\Omega_2, \mathcal{A}_2)$.

10.20 Definition: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$. Jede Übergangswahrscheinlichkeit $K(\cdot,\cdot)$ von $(\Omega, \mathcal{C})$ nach $(\Omega, \mathcal{A})$ mit der Eigenschaft

für jedes $A \in \mathcal{A}$ liefert $K(\cdot, A)$ eine Festlegung von $P(A|\mathcal{C})$

nennt man eine *reguläre Version der bedingten Wahrscheinlichkeit gegeben* $\mathcal{C}$.

10.21 Fragestellung: Existiert eine solche Übergangswahrscheinlichkeit ? Falls ja, hätte man mit $K(\cdot,\cdot)$ eine *simultane Festlegung aller* $P(A|\mathcal{C})$, $A \in \mathcal{A}$, mit den Eigenschaften

i) für jedes $A \in \mathcal{A}$: die $\mathcal{C}$-messbare Funktion $K(\cdot, A)$ ist eine Festlegung von $P(A|\mathcal{C})$

ii) $\forall\, \omega \in \Omega \; : \; K(\omega, \cdot)$ ist ein Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{A})$

gefunden: die in 10.18 $(*)$ aufgezeigten Schwierigkeiten wären ausgeräumt, und die Sprechweise 'bedingte Wahrscheinlichkeit von ... gegeben $\mathcal{C}$' wäre gerechtfertigt.

Die gestellte Frage wird in 10.27+10.28 beantwortet werden. Wir starten mit zwei einfachen Beispielen, in denen man mit ad-hoc Argumenten reguläre Versionen findet.

10.22 Beispiel: Sei X auf $(\Omega, \mathcal{A}, P)$ eine Zufallsvariable mit *höchstens abzählbar vielen Werten*, sei $\mathcal{C} := \sigma(X)$. Den Zielraum $(T, \mathcal{T})$ der ZV X schreiben wir o.E. als $T := I\!N_0$, $\mathcal{T} := \mathcal{P}(I\!N_0)$.

1) Ganz elementar definieren wir für alle $A \in \mathcal{A}$ und alle $k \in I\!N_0$

$$\widetilde{K}(k, A) := \begin{cases} \frac{P(A \cap \{X=k\})}{P(X=k)} & \text{falls } P(X=k) > 0 \,, \\ \nu(A) & \text{sonst} \end{cases}$$

wobei ν (als 'default value') irgendein festes Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{A})$ ist. Dies liefert eine Übergangswahrscheinlichkeit von $(I\!N_0, \mathcal{P}(I\!N_0))$ nach $(\Omega, \mathcal{A})$: für festes $A \in \mathcal{A}$ ist $\widetilde{K}(\cdot, A)$ trivialerweise $\mathcal{P}(I\!N_0)$-messbar, und für festes $k \in I\!N_0$ ist $\widetilde{K}(k, \cdot)$ ein Wahrscheinlichkeitsmass auf $(\Omega, \mathcal{A})$.

2) Betrachte die Sub-σ-Algebra $\mathcal{C} := \sigma(X)$ von $\mathcal{A}$: dann ist X eine messbare Abbildung von $(\Omega, \mathcal{C})$ nach $(I\!N_0, \mathcal{P}(I\!N_0))$, daher ist

$$(+) \qquad K(\omega, A) \;:=\; \widetilde{K}(X(\omega), A), \quad \omega \in \Omega,\; A \in \mathcal{A}$$

eine Übergangswahrscheinlichkeit von $(\Omega, \mathcal{C})$ nach $(\Omega, \mathcal{A})$. Wir zeigen

$$(++) \qquad \forall\, A \in \mathcal{A} \text{ gilt: } \quad K(\cdot, A) \text{ ist eine Festlegung von } \; P(A|\mathcal{C}) \,.$$

Dazu ist die Eigenschaft

$$E_P(1_C K(\cdot, A)) \;=\; E_P(1_C 1_A) \quad \text{für alle } \; C \in \mathcal{C} = X^{-1}(\mathcal{P}(I\!N_0))$$

nachzuprüfen. Mengen in $\mathcal{C}$ sind aber höchstens abzählbare Vereinigungen von Ereignissen $\{X{=}k\}$ für geeignete $k \in I\!N_0$. Die Definition von $\widetilde{K}(\cdot, \cdot)$ aber liefert

$$\int_{\{X=k\}} K(\omega, A) P(d\omega) \;=\; \widetilde{K}(k, A) \cdot P(X=k) \;=\; P(A \cap \{X=k\}) \;=\; \int_{\{X=k\}} 1_A(\omega) P(d\omega) \,.$$

sowohl im Fall $P(X=k) > 0$ als auch im Fall $P(X=k) = 0$. Damit ist (++) nachgewiesen. $\square$

10.23 Beispiel: Sei $(\Omega, \mathcal{A})$ ein Produktraum $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$, seien für $i = 1, 2$ σ-endliche Masse μ_i auf $(\Omega_i, \mathcal{A}_i)$ gegeben. Sei P ein Wahrscheinlichkeitsmass auf dem Produktraum $(\Omega, \mathcal{A})$ mit der Eigenschaft

$$P \;\ll\; \mu_1 \otimes \mu_2 \,.$$

Nach 3.5 gibt es eine produktmessbare Dichte von P bezüglich $\mu_1 \otimes \mu_2$

$$f(\cdot, \cdot) : \; \Omega_1 \times \Omega_2 \to [0, \infty) \quad \mathcal{A}_1 \otimes \mathcal{A}_2\text{-messbar so dass} \quad P(A) = \int_A f \, d(\mu_1 \otimes \mu_2) \,, \; A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \,.$$

Durch Ausintegrieren gewinnt man Randdichten: ist $\mathcal{C}_i := \sigma(\pi_i)$ die von der Projektion $\pi_i : \Omega \to \Omega_i$ auf die i-te Komponente erzeugte Sub-σ-Algebra von $\mathcal{A}$, ergibt sich

$$f_i := \frac{d\,\mathcal{L}(\pi_i|P)}{d\,\mu_i} \quad , \quad f_i(\omega_i) \;=\; \int_{\Omega_j} f(\omega_1,\omega_2)\,\mu_j(d\omega_j)\ , \quad j \neq i\ ,\ \omega_i \in \Omega_i\ .$$

Mit $\mathcal{C} := \mathcal{C}_2$ bestimmen wir nun eine reguläre Version der bedingten Wahrscheinlichkeit gegeben $\mathcal{C}$. Mit einem beliebigen Wahrscheinlichkeitsmass Q auf $(\Omega,\mathcal{A})$ als 'default value' definiere zuerst

$$\widetilde{K}(\omega_2, A) \;:=\; \begin{cases} \int_{\Omega_1} \frac{(1_A\,f)(\omega_1,\omega_2)}{f_2(\omega_2)}\;\mu_1(d\omega_1) & \text{falls } f_2(\omega_2) > 0 \\ Q(A) & \text{sonst} \end{cases}$$

für $\omega_2 \in \Omega_2$ und $A \in \mathcal{A}$. Mit 4.7 und monotoner Konvergenz ist leicht zu sehen, dass die Abbildung $(\omega_2, A) \to \widetilde{K}(\omega_2, A)$ eine Übergangswahrscheinlichkeit von $(\Omega_2, \mathcal{A}_2)$ nach $(\Omega,\mathcal{A})$ liefert; nach Einsetzen der Zufallsvariable π_2 wird

$$(+) \qquad K(\omega, A) \;:=\; \widetilde{K}(\pi_2(\omega), A)\ , \quad \omega \in \Omega\ ,\ A \in \mathcal{A}$$

eine Übergangswahrscheinlichkeit von $(\Omega,\mathcal{C})$ nach $(\Omega,\mathcal{A})$. Zeige nun:

$$(++) \qquad \text{für jedes feste } A \in \mathcal{A} \text{ gilt}:\ \ K(\cdot, A) \text{ ist eine Festlegung von } E_P(1_A \mid \mathcal{C})\ .$$

Dies sieht man so: für jedes $C \in \mathcal{C}$ gibt es ein $A_2 \in \mathcal{A}_2$ mit der Eigenschaft $C = \Omega_1 \times A_2$, also

$$\begin{aligned}
\int_C K(.,A)\,dP \quad &= \quad \int_C K(.,A)\,f\,d(\mu_1{\otimes}d\mu_2) \;=\; \int_{\Omega_1\times A_2} K(.,A)\,f\,d(\mu_1{\otimes}d\mu_2) \\
&= \quad \int_{A_2} \mu_2(d\omega_2) \int_{\Omega_1} \mu_1(d\omega_1)\,f(\omega_1,\omega_2)\,K((\omega_1,\omega_2),A) \\
&= \quad \int_{A_2} \mu_2(d\omega_2)\,[\int_{\Omega_1} \mu_1(d\omega_1)\,f(\omega_1,\omega_2)]\;\widetilde{K}(\omega_2, A) \\
&= \quad \int_{A_2} \mu_2(d\omega_2)\,f_2(\omega_2)\,\widetilde{K}(\omega_2, A) \\
&= \quad \int_{A_2} \mu_2(d\omega_2)\,\left[1_{\{f_2>0\}}(\omega_2) \int_{\Omega_1} (1_A f)(\omega_1,\omega_2)\,\mu_1(d\omega_1) + 1_{\{f_2=0\}}(\omega_2) f_2(\omega_2)\,Q(A)\right]
\end{aligned}$$

mit Fubini; für dasselbe $C = \Omega_1 \times A_2$ hat man aber wegen $1_A f \leq f$

$$\begin{aligned}
E_P(1_C\,1_A) \quad &= \quad \int_C 1_A\,f\,d(\mu_1{\otimes}d\mu_2) \;=\; \int_{A_2} \mu_2(d\omega_2) \int_{\Omega_1} \mu_1(d\omega_1)\,(1_A f)(\omega_1,\omega_2) \\
&= \quad \int_{A_2} \mu_2(d\omega_2)\,\left[1_{\{f_2>0\}}(\omega_2) \int_{\Omega_1} (1_A f)(\omega_1,\omega_2)\,\mu_1(d\omega_1) + 1_{\{f_2=0\}}(\omega_2)\,f_2(\omega_2)\cdot 0\right]\ .
\end{aligned}$$

Beide Ausdrücke reduzieren sich also auf

$$\int_{A_2} \mu_2(d\omega_2)\,1_{\{f_2>0\}}(\omega_2) \int_{\Omega_1} (1_A f)(\omega_1,\omega_2)\,\mu_1(d\omega_1)$$

womit

$$\int_C K(.,A)\,dP \;=\; E_P(1_C\,1_A) \quad , \quad C \in \mathcal{C}$$

und somit (++) bewiesen ist. Also liefert (+) eine reguläre Festlegung der bedingten Wahrscheinlichkeit gegeben $\mathcal{C}$. $\square$

Im allgemeinen aber ist die Existenz einer regulären Version der bedingten Wahrscheinlichkeit gegeben eine beliebige Sub-σ-Algebra $\mathcal{C}$ nur dann garantiert, wenn Wahrscheinlichkeitsmasse auf *polnischen Räumen* (definiert in 10.24 unten) betrachtet werden. In einem beliebigen messbaren Raum $(\Omega, \mathcal{A})$ kann man sich gegebenenfalls auf *Teilstrukturen* zurückziehen, die einen polnischen Raum *imitieren*. Dazu braucht man eine Zufallsvariable Y auf $(\Omega, \mathcal{A})$, die ihre Werte in einem polnischen Raum $(E, \mathcal{E})$ annimmt, zieht sich zurück auf die Urbild-σ-Algebra

$$\mathcal{A}' \;:=\; Y^{-1}(\mathcal{E}) \;=\; \{A \in \mathcal{A} : A = \{Y \in B\}, B \in \mathcal{E}\}$$

und betrachtet nur noch die *Restriktion von P auf $\mathcal{A}'$. Für die Restriktion $P_{|\mathcal{A}'}$* findet man dann reguläre Versionen der bedingten Wahrscheinlichkeit gegeben eine beliebige Sub-σ-Algebra $\mathcal{C}$ (siehe unten 10.27). Damit aber redet man über die 'bedingte Verteilung von Y gegeben $\mathcal{C}$'.

10.24 Definition: Ein messbarer Raum $(E, \mathcal{E})$ heisst *polnisch* falls gilt

i) E ist versehen mit einer Topologie $\mathcal{O}$; es existiert eine (geeignete) $\mathcal{O}$ erzeugende Metrik $d(\cdot,\cdot)$ so dass (E, d) ein vollständiger separabler metrischer Raum ist;

ii) $\mathcal{E} = \sigma(\mathcal{O}) = \mathcal{B}(E)$ ist die Borel-σ-Algebra.

10.25 Bemerkungen: a) Beispiele für polnische Räume sind $\mathbb{R}^d$ mit der Euklidischen Metrik, $d \geq 1$, der Raum $C([0,1])$ der stetigen Funktionen $[0,1] \to \mathbb{R}^d$ mit der Metrik der gleichmässigen Konvergenz $d(f,g) = \max_{0 \leq t \leq 1} |f(t) - g(t)|$, der Raum $C([0,\infty))$ der stetigen Funktionen $[0,\infty) \to \mathbb{R}^d$ mit der Metrik der lokal gleichmässigen Konvergenz $d(f,g) \;=\; \sum_n 2^{-n} \left(\max_{0 \leq t \leq n} |f(t) - g(t)| \wedge 1\right)$, und viele andere mehr. Siehe Billingsley (1968).

b) Produkte polnischer Räume sind wieder polnisch (vgl. 4.14, bzw. Billingsley 1968 S. 225).

c) Wichtig ist in 10.24 die Topologie $\mathcal{O}$: es ist möglich, dass zwei verschiedene Metriken – von denen nur eine die gewünschten Eigenschaften hat – zu derselben Topologie $\mathcal{O}$ auf E führen. $\square$

In Kapitel 6 (teils ohne Beweis wie in 6.C und mit Verweis auf Billingsley 1968) hatte man bereits wichtige Resultate für polnische Räume gesehen; auch in manchen Beweisen wurden typische Schlussweisen für polnische Räume bereits in Spezialfällen benutzt (6.1', 6.18). Die für uns wichtigen Eigenschaften polnischer Räume sind die folgenden.

10.26 Eigenschaften: Ist $(E, \mathcal{E})$ polnisch, so

a) ist die σ-Algebra $\mathcal{E}$ *abzählbar erzeugt*,

d.h. es existiert ein abzählbares Teilsystem $\mathcal{S} \subset \mathcal{E}$ so dass $\sigma(\mathcal{S}) = \mathcal{E}$;

b) gibt es eine abzählbare Algebra $\mathcal{G} \subset \mathcal{E}$ mit $\sigma(\mathcal{G}) = \mathcal{E}$,

c) ist jedes endliche Mass μ auf $(E, \mathcal{E})$ kompakt approximierbar:

$$\mu(A) \;=\; \sup\{\, \mu(K) \;:\; K \subset A\,,\; K \text{ kompakt} \,\}\,, \quad A \in \mathcal{A}\,.$$

Beweis: c) Wurde bereits in 6.1' bewiesen.

a) Ausgehend von einer abzählbar dichten Teilmenge S von E, betrachte das System aller offenen Kugeln mit Zentrum in S und rationalem Radius $\mathcal{S} := \{B_\varepsilon(x) : x \in S,\, \varepsilon \in \mathbb{Q}^+\}$, dann kann jede offene Menge $O \subset E$ als (abzählbare) Vereinigung geeigneter Mengen aus $\mathcal{S}$ geschrieben werden. Damit ist $\mathcal{E} = \sigma(\mathcal{O})$ abzählbar erzeugt.

b) Von $\mathcal{S}$ geht man schrittweise weiter zu

$$\begin{aligned}
\widetilde{\mathcal{S}} \;&:=\; \{\, A \subset E : A \in \mathcal{S} \text{ oder } A^c \in \mathcal{S} \,\} \;\cup\; \{\, \Omega\,,\, \emptyset \,\}\,; \\
\widetilde{\widetilde{\mathcal{S}}} \;&:=\; \text{Klasse aller endlichen Durchschnitte von Mengen aus } \widetilde{\mathcal{S}}\,; \\
\mathcal{G} \;&:=\; \text{Klasse aller endlichen Vereinigungen von Mengen aus } \widetilde{\widetilde{\mathcal{S}}}\,.
\end{aligned}$$

Dann ist $\mathcal{G}$ eine abzählbare Algebra in $\mathcal{E}$ mit $\sigma(\mathcal{G}) = \mathcal{E}$: Klar enthält $\mathcal{G}$ E und $\emptyset$, und ist abgeschlossen unter endlichen Vereinigungen. Betrachtet man

$$[\bigcup_{i=1}^{\ell} \bigcap_{j=1}^{k} F_{i,j}\,]^c \;=\; \bigcap_{i=1}^{\ell} \bigcup_{j=1}^{k} F_{i,j}^c$$

für $F_{i,j} \in \widetilde{\mathcal{S}}$ und $H_{i,j} := F_{i,j}^c \in \widetilde{\mathcal{S}}$ und schreibt die rechte Seite um gemäss

$$[\bigcup_{j=1}^{k} H_{1,j}] \cap \ldots \cap [\bigcup_{j=1}^{k} H_{\ell,j}] \;=\; \bigcup_{(j_1,\ldots,j_\ell) \in \{1,\ldots,k\}^\ell} [H_{1,j_1} \cap \ldots \cap H_{\ell,j_\ell}]\,,$$

ergibt sich, dass $\mathcal{G}$ auch stabil unter Komplementbildung ist. $\square$

10.27 Hauptsatz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, sei $\mathcal{C}$ eine beliebige Sub-σ-Algebra von $\mathcal{A}$. Sei $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ eine Zufallsvariable mit Werten in einem polnischen Raum $(E, \mathcal{E})$.

a) Dann existiert eine reguläre Version der *bedingten Verteilung von Y gegeben $\mathcal{C}$*, d.h. eine Übergangswahrscheinlichkeit $\widehat{K}(.,.)$ von $(\Omega, \mathcal{C})$ nach $(E, \mathcal{E})$

$$\begin{cases} \forall\, \omega \in \Omega \text{ fest:} \quad F \to \widehat{K}(\omega, F) \quad \text{ist ein Wahrscheinlichkeitsmass auf } \mathcal{E} \\ \forall\, F \in \mathcal{E} \text{ fest:} \quad \omega \to \widehat{K}(\omega, F) \quad \text{ist eine } \mathcal{C}\text{-messbare Funktion} \end{cases}$$

mit der Eigenschaft:

$$\forall\, F \in \mathcal{E} \;:\quad \widehat{K}(\cdot, F) \;\text{ ist eine Festlegung von }\; P(\{Y \in F\} \mid \mathcal{C})\;.$$

Die übliche Notation für eine reguläre Version der bedingten Verteilung von Y gegeben $\mathcal{C}$ ist

$$P^{Y|\mathcal{C}}(\omega, F) \;:=\; \widehat{K}(\omega, F)\;, \quad \omega \in \Omega\;,\; F \in \mathcal{E}\;.$$

b) Für zwei verschiedene Festlegungen $\widehat{K}_i(\cdot,\cdot)$, $i = 1, 2$, einer regulären Version der bedingten Verteilung von Y gegeben $\mathcal{C}$ existiert eine P-Nullmenge $N \in \mathcal{C}$ mit der Eigenschaft

$$\widehat{K}_1(\omega, F) \;=\; \widehat{K}_2(\omega, F) \quad \text{für alle } F \in \mathcal{E} \text{ und alle } \omega \in \Omega \setminus N\;.$$

Vor dem Beweis dieses Satzes zuerst die Antwort auf die Frage aus 10.21: aus 10.27 ergibt sich, falls $(\Omega, \mathcal{A})$ selbst ein polnischer Raum ist, die wichtige Aussage

10.28 Korollar: Sei $(\Omega, \mathcal{A})$ polnisch. Dann gilt: für *jedes* Wahrscheinlichkeitsmass P auf $(\Omega, \mathcal{A})$ und für *jede* Sub-σ-Algebra $\mathcal{C}$ von $\mathcal{A}$ existiert eine reguläre Festlegung der bedingten Wahrscheinlichkeit gegeben $\mathcal{C}$.

Beweis: Wende 10.27 an auf den Fall $(E, \mathcal{E}) := (\Omega, \mathcal{A})$ und $Y := \mathrm{id}|_\Omega$. □

10.29 Beweis von 10.27: Der Beweis wird in mehreren Schritten gegeben.

1) Für jedes $F \in \mathcal{E}$ fixiere eine Folge $(K_n^F)_n$ kompakter Teilmengen von F mit der Eigenschaft

$$P(\{Y \in F\}) \;=\; \sup_n P(\{Y \in K_n^F\})\;.$$

Dies ist die kompakte Approximierbarkeit des Wahrscheinlichkeitsmasses $\mathcal{L}(Y|P)$ auf dem polnischen Raum $(E, \mathcal{E})$ nach 6.1'; der Beweis von 6.1' erfordert Vollständigkeit von E.

2) Betrachte zu F und $(K_n^F)_n$ wie in 1) die bedingten Erwartungen gegeben $\mathcal{C}$. Wir zeigen: für jede Wahl von Festlegungen

$$g \;=\; P(\{Y \in F\}|\mathcal{C})\;, \quad g_n \;=\; P(\{Y \in K_n^F\}|\mathcal{C})\;, \quad n \geq 1,$$

gibt es eine P-Nullmenge $N \in \mathcal{C}$ so dass

$$(*) \qquad g(\omega) \;=\; \sup_n g_n(\omega) \quad \forall\, \omega \notin N\,.$$

Zum Beweis geht man aus von Monotonie 10.3 a)ii) der bedingten Erwartung

$$K_n^F \subset F \quad \Longrightarrow \quad g_n \leq g \quad P\text{-fast sicher}\;,\; n \geq 1$$

und erhält nach Vereinigung von abzählbar vielen P-Nullmengen aus $\mathcal{C}$ zunächst einmal

$$(\circ) \qquad \sup_n g_n \;\leq\; g \quad P\text{-fast sicher}\,.$$

Damit hat man eine Ungleichungskette

$$\begin{aligned} P(Y \in F) \;&\overset{1)}{=}\; \sup_n P(Y \in K_n^F) \;=\; \sup_n \int_\Omega 1_{\{Y \in K_n^F\}}\, dP \;=\; \sup_n \int_\Omega g_n\, dP \\ &\leq\; \int_\Omega \sup_n g_n\, dP \;\overset{(\circ)}{\leq}\; \int_\Omega g\, dP \;=\; \int_\Omega 1_{\{Y \in F\}}\, dP \;=\; P(Y \in F) \end{aligned}$$

in der notwendig überall '=' steht. Wegen $(\circ)$ muss dann insbesondere $(*)$ gelten.

3) Da $(E, \mathcal{E})$ nach Voraussetzung ein polnischer Raum ist, existiert nach 10.26 b) eine abzählbare Algebra $\mathcal{G} \subset \mathcal{E}$ mit $\sigma(\mathcal{G}) = \mathcal{E}$. Zu jedem $F \in \mathcal{G}$ sei $(K_n^F)_{n\geq 1}$ die in Schritt 1) fixierte Folge. Auch

$$\widetilde{\mathcal{G}} \;:=\; \text{die von dem erweiterten System } \{\, F\,,\, K_n^F \;:\; n \geq 1\,,\, F \in \mathcal{G}\,\} \text{ erzeugte Algebra}$$

ist dann abzählbar, in Wiederholung des Arguments aus 10.26 a), und erzeugt $\mathcal{E}$:

$$\mathcal{E} \;=\; \sigma(\widetilde{\mathcal{G}})\,.$$

Für jedes $\widetilde{F} \in \widetilde{\mathcal{G}}$ wählt man nun eine Festlegung von $P(\{Y \in \widetilde{F}\} \mid \mathcal{C})$ mit Werten in $[0,1]$ (und benutzt konstante Festlegungen $\equiv 1$, $\equiv 0$ für $\widetilde{F} = E$, $\widetilde{F} = \emptyset$).

4) Da $\widetilde{\mathcal{G}}$ abzählbar, findet man *eine* P-Nullmenge $N_{\widetilde{\mathcal{G}}} \in \mathcal{C}$, so dass für jedes $\omega \notin N_{\widetilde{\mathcal{G}}}$ gilt:

i) $\widetilde{F} \to P(\{Y \in \widetilde{F}\} \mid \mathcal{C})(\omega)$ ist ein Inhalt auf $\widetilde{\mathcal{G}}$;

ii) $P(\{Y \in F\} \mid \mathcal{C})(\omega) \;=\; \sup_n P(\{Y \in K_n^F\} \mid \mathcal{C})(\omega)$ für jedes $F \in \mathcal{G}$.

Beachte hierbei: mit $\widetilde{\mathcal{G}}$ ist auch das System aller endlichen Teilfamilien $A_1, ..., A_l$ von $\widetilde{\mathcal{G}}$, $l \geq 0$, abzählbar; für jede Teilfamilie von paarweise disjunkten $A_1, ..., A_l$ entsteht wie in 10.18 $(*)$ in jeder der Gleichungen

$$P(\{Y \in \bigcup_{j=1}^{l} A_j\} \mid \mathcal{C}) \;=\; \sum_{j=1}^{l} P(\{Y \in A_j\} \mid \mathcal{C}) \quad P\text{-fast sicher}$$

als Ausnahmemenge eine P-Nullmenge in $\mathcal{C}$; durch Vereinigung dieser abzählbar vielen Ausnahmemengen mit den abzählbar vielen Ausnahmemengen, die für $F \in \mathcal{G}$ in Schritt 2) entstanden sind, erhält man die gewünschte P-Nullmenge $N_{\widetilde{\mathcal{G}}} \in \mathcal{C}$.

5) Wir zeigen: in Einschränkung auf die Algebra $\mathcal{G}$ aus Schritt 3) ist der in 4) i) erzielte Inhalt

$$(**) \qquad \mathcal{G} \ni F \longrightarrow P(\{Y \in F\} \mid \mathcal{C})(\omega) \quad \in \quad [0,1]$$

für festes $\omega \notin N_{\widetilde{\mathcal{G}}}$ σ-stetig in $\emptyset$: für jedes $\omega \notin N_{\widetilde{\mathcal{G}}}$ ist $(**)$ dann ein Prämass auf $\mathcal{G}$.
Zum Beweis fixiere $\omega \notin N_{\widetilde{\mathcal{G}}}$ und betrachte eine absteigende Mengenfolge $(F_j)_j \subset \mathcal{G}$ mit $F_j \downarrow \emptyset$. Nach ii) in Schritt 4) gibt es zu jedem j ein n_j mit

$$P(\{\, Y \in F_j \setminus K^{F_j}_{n_j} \,\} \mid \mathcal{C})(\omega) \;<\; \varepsilon\, 2^{-j} \;, \quad j \geq 1 \;;$$

wegen $F_j \downarrow \emptyset$ und $K^{F_j}_{n_j} \subset F_j$ gilt dann

$$\bigcap_{j=1}^{\infty} K^{F_j}_{n_j} \;=\; \emptyset \;.$$

Nun ist $\bigcap_{j=1}^{\ell} K^{F_j}_{n_j}$, $\ell = 1, 2, \ldots$, eine fallende Folge von Kompakta, also existiert nach dem Cantorschen Durchschnittssatz (z.B. Rudin 1998, S. 44) ein j_0 so dass schon

$$\bigcap_{j=1}^{j_0} K^{F_j}_{n_j} \;=\; \emptyset \;.$$

Damit argumentiert man für die fallende Folge $(F_j)_j$ so: für beliebiges $m \geq j_0$ gilt zuerst

$$F_m \;\subset\; F_{j_0} \;=\; F_{j_0} \setminus \emptyset \;=\; F_{j_0} \setminus \bigcap_{j=1}^{j_0} K^{F_j}_{n_j} \;=\; \bigcup_{j=1}^{j_0} (F_{j_0} \setminus K^{F_j}_{n_j}) \;\subset\; \bigcup_{j=1}^{j_0} (F_j \setminus K^{F_j}_{n_j}) \;.$$

Weil für jedes feste $\omega \notin N_{\widetilde{\mathcal{G}}}$ $P(\{Y \in \cdot\} \mid \mathcal{C})(\omega)$ ein Inhalt auf $\widetilde{\mathcal{G}}$ ist, nach i) in Schritt 4), darf man für $\omega \notin N_{\widetilde{\mathcal{G}}}$ und $m \geq j_0$ schreiben

$$P(\{Y \in F_m\} \mid \mathcal{C})(\omega) \;\leq\; P(\{Y \in F_{j_0}\} \mid \mathcal{C})(\omega) \;\leq\; \sum_{j=1}^{j_0} P(\{Y \in F_j \setminus K^{F_j}_{n_j}\} \mid \mathcal{C})(\omega) \;<\; \varepsilon \;.$$

Da $\varepsilon > 0$ beliebig war, ist damit für $\omega \notin N_{\widetilde{\mathcal{G}}}$ gezeigt

$$(F_j)_j \subset \mathcal{G} \;,\; F_j \downarrow \emptyset \quad\Longrightarrow\quad P(\{Y \in F_j\} \mid \mathcal{C})(\omega) \;\downarrow\; 0 \quad \text{für } \; j \to \infty.$$

Für $\omega \notin N_{\widetilde{\mathcal{G}}}$ ist damit $P(\{Y \in \cdot\} \mid \mathcal{C})(\omega)$ als Inhalt auf $\mathcal{G}$ σ-stetig in der leeren Menge. Damit ist bewiesen, dass für $\omega \notin N_{\widetilde{\mathcal{G}}}$ die Abbildung $(**)$ ein Prämass auf $\mathcal{G}$ liefert.

6) Jedes endliche Prämass auf $\mathcal{G}$ kann auf genau eine Weise zu einem (endlichen) Mass auf $\mathcal{E} = \sigma(\mathcal{G})$ fortgesetzt werden, nach 1.31. Also gibt es nach Schritt 5) für jedes feste $\omega \notin N_{\widetilde{\mathcal{G}}}$ genau ein Wahrscheinlichkeitsmass Q_ω auf $\mathcal{E}$ mit

$$(+) \qquad Q_\omega(F) \;=\; P(\{Y \in F\}|\mathcal{C})(\omega) \quad \text{falls } F \in \mathcal{G} \, .$$

Mit einem beliebigen Wahrscheinlichkeitsmass Q auf $\mathcal{E}$ als default value definiert man nun

$$(++) \qquad \widehat{K}(\omega, F) \;:=\; \begin{cases} Q_\omega(F) \, , & F \in \mathcal{E} \, , \quad \text{falls} \quad \omega \notin N_{\widetilde{\mathcal{G}}} \\ Q(F) \, , & F \in \mathcal{E} \, , \quad \text{falls} \quad \omega \in N_{\widetilde{\mathcal{G}}} \, . \end{cases}$$

Wir zeigen: $\widehat{K}(\cdot,\cdot)$ ist eine wohldefinierte Übergangswahrscheinlichkeit von $(\Omega, \mathcal{C})$ nach $(E, \mathcal{E})$. Nach Konstruktion ist $\widehat{K}(\omega, \cdot)$ für festes $\omega \in \Omega$ ein Wahrscheinlichkeitsmass auf $\mathcal{E}$. Den Nachweis von

$$\forall \, F \in \mathcal{E} \; : \quad \omega \;\to\; \widehat{K}(\omega, F) \;=\; 1_{N_{\widetilde{\mathcal{G}}}}(\omega) Q(F) + 1_{N^c_{\widetilde{\mathcal{G}}}}(\omega) Q_\omega(F) \quad \text{ist } \mathcal{C}\text{-messbar}$$

führt man mit einem Dynkin-Schluss: betrachte das System

$$\mathcal{H} \;:=\; \{ \, F \in \mathcal{E} : \omega \to \widehat{K}(\omega, F) \text{ ist } \mathcal{C}\text{-messbar} \, \}.$$

Zunächst gilt $E \in \mathcal{H}$ (wegen $Q_\omega(E) = 1$ für alle $\omega \in \Omega$ und $N_{\widetilde{\mathcal{G}}} \in \mathcal{C}$) sowie $\emptyset \in \mathcal{H}$ (wegen $Q_\omega(\emptyset) = 0$ für alle $\omega \in \Omega$ und $N_{\widetilde{\mathcal{G}}} \in \mathcal{C}$). Weil $Q_\omega(\cdot)$ für festes $\omega \in \Omega$ als Wahrscheinlichkeitsmass gegeben ist, muss $\mathcal{H}$ notwendig stabil unter Bildung von Komplementen und unter Bildung von abzählbaren disjunkten Vereinigungen sein. Damit ist $\mathcal{H}$ Dynkin. Weiter ist $\mathcal{G}$ ein Teilsystem von $\mathcal{H}$: für $F \in \mathcal{G}$ ist die rechte Seite von (+) eine $\mathcal{C}$-messbare Funktion, es gilt $N_{\widetilde{\mathcal{G}}} \in \mathcal{C}$, also ist $\omega \to \widehat{K}(\omega, F)$ eine $\mathcal{C}$-messbare Funktion für jedes $F \in \mathcal{G}$. Als Algebra ist $\mathcal{G}$ durchschnittsstabil. Enthält aber $\mathcal{H}$ den $\cap$-stabilen Erzeuger $\mathcal{G}$ von $\mathcal{E}$, dann als Dynkin-System auch $\delta(\mathcal{G}) = \sigma(\mathcal{G}) = \mathcal{E}$ nach 1.11. Gezeigt ist $\mathcal{H} = \mathcal{E}$. Also ist $\widehat{K}(\cdot,\cdot)$ eine wohldefinierte Übergangswahrscheinlichkeit von $(\Omega, \mathcal{C})$ nach $(E, \mathcal{E})$.

7) Wir zeigen

$$(+++) \qquad \forall \, F \in \mathcal{E} \;\text{ ist }\; \widehat{K}(\cdot, F) \;\text{ eine Festlegung von }\; P(\{Y \in F\} \mid \mathcal{C}) \, .$$

Zum Beweis betrachte $C \in \mathcal{C}$ beliebig und fest. Durch

$$F \;\longrightarrow\; \int_C \widehat{K}(\omega, F) \, P(d\omega) \, , \quad F \;\longrightarrow\; \int_C 1_{\{Y \in F\}}(\omega) \, P(d\omega)$$

sind zwei endliche Masse auf $\mathcal{E}$ gegeben (beachte: Mischen der Wahrscheinlichkeitsmasse $\widehat{K}(\omega, \cdot)$ über $P(d\omega)$, $\omega \in C$, liefert wieder ein Wahrscheinlichkeitsmass). Wegen (+) stimmen diese überein auf dem $\cap$-stabilen Erzeuger $\mathcal{G}$ von $\mathcal{E}$. Nach Eindeutigkeitssatz 1.13 stimmen beide als endliche Masse dann auf ganz $\mathcal{E}$ überein. Gezeigt ist

$$\text{für alle } F \in \mathcal{E}, \text{ für alle } C \in \mathcal{C} : \qquad \int_C \widehat{K}(\cdot, F) \, dP \;=\; \int_C 1_{\{Y \in F\}} \, dP$$

und damit (+++). Damit ist die Existenzaussage 10.27 a) bewiesen.

8) Es bleibt die Eindeutigkeitsaussage b) zu zeigen. Seien dazu $\widehat{K}_1, \widehat{K}_2$ Übergangswahrscheinlichkeiten von $(\Omega, \mathcal{C})$ nach $(E, \mathcal{E})$, so dass für jedes $F \in \mathcal{E}$ gilt

$$\widehat{K}_1(\cdot, F) \text{ und } \widehat{K}_2(\cdot, F) \text{ sind Festlegungen von } E(\{Y \in F\} \mid \mathcal{C}) \ .$$

Sei $\mathcal{G}$ die abzählbare Algebra aus Schritt 3) mit $\mathcal{E} = \sigma(\mathcal{G})$; wieder existiert *eine* P-Nullmenge $N \in \mathcal{C}$ mit der Eigenschaft

$$\widehat{K}_1(\omega, F) \ = \ \widehat{K}_2(\omega, F) \quad \text{für alle } \ F \in \mathcal{G} \quad \text{falls } \omega \notin N \ .$$

Für $\omega \notin N$ müssen dann – wieder nach 1.13 – die Wahrscheinlichkeitsmasse $\widehat{K}_1(\omega, \cdot)$ und $\widehat{K}_2(\omega, \cdot)$ auf ganz $\mathcal{E}$ übereinstimmen. Damit ist der Beweis von 10.27 abgeschlossen. □

Ist $\mathcal{C} = \sigma(X)$, so erhält man für $(\omega, F) \to P(\{Y \in F\}|\mathcal{C})(\omega)$ auch auf der Ebene der Faktorisierungen reguläre Versionen, falls Y Werte in einem polnischen Raum annimmt:

10.30 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, betrachte eine Sub-σ-Algebra der Form $\mathcal{C} := \sigma(X)$, erzeugt von einer messbaren Abbildung $X : (\Omega, \mathcal{A}) \to (T, \mathcal{T})$. Sei $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ eine Zufallsvariable mit Werten in einem polnischen Raum $(E, \mathcal{E})$.
a) Dann existiert eine *reguläre Version der bedingten Verteilung von Y gegeben $X = \cdot$*, d.h. eine Übergangswahrscheinlichkeit $\widehat{\widehat{K}}$ von $(T, \mathcal{T})$ nach $(E, \mathcal{E})$ mit der Eigenschaft

$$\text{für jedes } F \in \mathcal{E} \text{ ist } \ \widehat{\widehat{K}}(\cdot, F) \ \text{ eine Festlegung von } \ P(\{Y \in F\} \mid X = \cdot) \ .$$

b) Für verschiedene Festlegungen $\widehat{\widehat{K}}_1$, $\widehat{\widehat{K}}_2$ einer regulären Version der bedingten Verteilung von Y gegeben $X = \cdot$ gilt: es existiert eine P^X-Nullmenge $N \in \mathcal{T}$ mit

$$\widehat{\widehat{K}}_1(x, F) \ = \ \widehat{\widehat{K}}_2(x, F) \quad \text{für alle } \ F \in \mathcal{E} \ , \ x \in T \setminus N \ .$$

Man schreibt $P^{Y|X=\cdot}(\cdot)$ für die Übergangswahrscheinlichkeit $\widehat{\widehat{K}}(\cdot, \cdot)$ von $(T, \mathcal{T})$ nach $(E, \mathcal{E})$, und $P^{Y|X=x}$ für die Wahrscheinlichkeitsmasse $\widehat{\widehat{K}}(x, \cdot)$, $x \in T$.

Beweis: Der Beweis verläuft genau analog zum Beweis von 10.27, nur arbeitet man im Falle $\mathcal{C} = \sigma(X)$ mit Faktorisierungen $P(\{Y \in F\}|X = \cdot)$ und P^X-Nullmengen anstelle von $P(\{Y \in F\}|\mathcal{C})$ und P-Nullmengen. □

10.31 Übungsaufgabe: Unter den Voraussetzungen aus 10.27 und 10.30 schreibe analog zu 10.14 $P^{Y|X}(\cdot,\cdot)$ statt $P^{Y|\sigma(X)}(\cdot,\cdot)$. Für die in 10.27 und 10.30 gewonnenen regulären Versionen der bedingten Verteilung von Y

$$\begin{array}{lcll} \widehat{K}(\cdot,\cdot) & = & P^{Y|X}(\cdot,\cdot) & \text{(Übergangswahrscheinlichkeit von } (\Omega,\sigma(X)) \text{ nach } (E,\mathcal{E})) \\ \widehat{\widehat{K}}(\cdot,\cdot) & = & P^{Y|X=\cdot}(\cdot) & \text{(Übergangswahrscheinlichkeit von } (T,\mathcal{T}) \text{ nach } (E,\mathcal{E})) \end{array}$$

gilt das folgende: mit $\widehat{K}(\cdot,\cdot)$ ist auch

$$(*) \qquad (\omega,F) \longrightarrow \widehat{\widehat{K}}(X(\omega),F)$$

eine reguläre Version der bedingten Verteilung von Y gegeben $\mathcal{C} := \sigma(X)$; insbesondere gibt es eine P-Nullmenge N in $\mathcal{C}$ so dass für jedes $\omega \in \Omega \setminus N$

$$\widehat{K}(\,\omega,\cdot\,) \quad \text{und} \quad \widehat{\widehat{K}}(\,X(\omega),\cdot\,)$$

als Wahrscheinlichkeitsmasse auf $(E,\mathcal{E})$ übereinstimmen.
Dies beweist man so: wegen der Messbarkeit von $X : (\Omega,\mathcal{C}) \to (T,\mathcal{T})$ ist die Abbildung in $(*)$ zunächst eine Übergangswahrscheinlichkeit von $(\Omega,\mathcal{C})$ nach $(E,\mathcal{E})$. Ist für ein Ereignis $F \in \mathcal{E}$ (beliebig aber fest) $h_F := \widehat{\widehat{K}}(\cdot,F)$ eine Festlegung von $P(\{Y \in F\}|X = \cdot)$, so ist $\widehat{\widehat{K}}(X(\cdot),F) = h_F \circ X$ eine Festlegung von $P(\{Y \in F\}|\mathcal{C})$. Also liefert $(*)$ eine reguäre Version der bedingten Verteilung von Y gegeben $\mathcal{C}$. Damit sind zwei Festlegungen einer regulären Version der bedingten Verteilung von Y gegeben $\mathcal{C}$ gefunden, der Rest ist 1.27 b). $\square$

Ist $(\Omega,\mathcal{A})$ selbst ein polnischer Raum, so erhält man bedingte Erwartungen $E_P(f\,|\,\mathcal{C})$ durch *Ausintegrieren von f bezüglich einer regulären Version $K(\cdot,\cdot)$ der bedingten Wahrscheinlichkeit*

$$(\omega,A) \longrightarrow K(\omega,A) = E_P(1_A|\mathcal{C})(\omega)\ ;$$

eine suggestive Schreibweise für die Übergangswahrscheinlichkeit ist dann

$$K(\omega,dy) \;=\; P(dy\,|\,\mathcal{C})(\omega)\ .$$

Den Grundraum als polnischen Raum anzunehmen ist eine eher starke Voraussetzung. Jedoch kann man oft wie in 10.27 die Struktur eines polnischen Raumes über eine Zufallsvariable Y 'importieren', die Werte in einem polnischen Raum annimmt: in diesem Fall kann man bedingte Erwartungen $E_P(\varphi(Y)\,|\,\mathcal{C})$ durch Ausintegrieren von φ bezüglich einer regulären Version der bedingten Verteilung von Y gegeben $\mathcal{C}$ berechnen. Der folgende Satz 10.32 und sein Korollar 10.32' präzisieren dies.

10.32 Satz: Sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum, $\mathcal{C}$ Sub-σ-Algebra von $\mathcal{A}$, $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ eine Zufallsvariable, sei $\varphi : E \to I\!\!R$ eine $\mathcal{E}$-messbare Funktion mit $\varphi(Y) \in L^1(P)$.

a) Sei eine reguläre Version $\widehat{K}(\cdot,\cdot)$ der bedingten Verteilung $P^{Y|\mathcal{C}}(\cdot,\cdot)$ gegeben. Dann ist

$$N := \left\{ \omega \in \Omega \ : \ \int_E |\varphi|(y) \widehat{K}(\omega, dy) = +\infty \right\}$$

eine P-Nullmenge in $\mathcal{C}$, und durch

$$\Omega \ni \omega \quad \longrightarrow \quad \begin{cases} \int_E \varphi(y)\, \widehat{K}(\omega, dy)\ , & \omega \notin N \\ 0\ , & \omega \in N \end{cases}$$

erhält man eine Festlegung der bedingten Erwartung $E_P(\varphi \circ Y \mid \mathcal{C})$.

b) Sei $\mathcal{C} = \sigma(X)$ für eine Zufallsvariable $X : (\Omega, \mathcal{A}) \to (T, \mathcal{T})$ mit Werten einem beliebigen messbaren Raum. Sei eine reguläre Version $\widehat{\widehat{K}}(\cdot,\cdot)$ der bedingten Verteilung $P^{Y|X=\cdot}(\cdot)$ gegeben. Dann ist

$$N' := \left\{ x \in T \ : \ \int_E |\varphi|(y)\, \widehat{\widehat{K}}(x, dy) \ = \ +\infty \right\}$$

eine P^X-Nullmenge in $\mathcal{T}$, und durch

$$T \ni x \quad \longrightarrow \quad \begin{cases} \int_E \varphi(y)\, \widehat{\widehat{K}}(x, dy)\ , & x \notin N' \\ 0\ , & x \in N' \end{cases}$$

erhält man eine Festlegung der bedingten Erwartung $E(\varphi \circ Y | X = \cdot)$.

Beweis: Wir zeigen a); b) beweist man analog auf der Ebene der Faktorisierungen.

1) Im Fall einer Indikatorfunktion $\varphi = 1_F$, $F \in \mathcal{E}$, ist N die leere Menge, und nach Definition einer regulären Version der bedingten Verteilung von Y gegeben $\mathcal{C}$ in 10.27 ist

$$\int_E \varphi(y) \widehat{K}(\cdot, dy) \ = \ \widehat{K}(\cdot, F)$$

eine Festlegung von

$$E_P(\varphi(Y)|\mathcal{C}) \ = \ P(\{Y \in F\}|\mathcal{C})\ .$$

2) Betrachte nun eine *nichtnegative* $\mathcal{E}$-messbare Funktion $\varphi : E \to [0, \infty)$. Dann wird φ monoton approximiert durch $\mathcal{E}$-Elementarfunktionen

$$\varphi_n \ := \ \sum_{j=1}^{n2^n} \frac{j-1}{2^n}\, 1_{\{\frac{j-1}{2^n} \le \varphi < \frac{j}{2^n}\}} \ + \ n\, 1_{\{\varphi \ge n\}}\ , \quad n \to \infty\ .$$

Schreibe $F_{n,j} := \{\frac{j-1}{2^n} \le \varphi < \frac{j}{2^n}\}$, $j = 1, \ldots, n2^n$, und $F_{n,n2^n+1} := \{\varphi \ge n\}$. Nach 1) ist

$$(+) \qquad \int_E \varphi_n(y)\, \widehat{K}(\cdot, dy) \;=\; \sum_{j=1}^{n2^n+1} \frac{j-1}{2^n}\, \widehat{K}(\cdot, F_{n,j})$$

für festes n eine Festlegung von

$$E_P(\varphi_n \circ Y \mid \mathcal{C}) \;=\; \sum_{j=1}^{n2^n+1} \frac{j-1}{2^n}\, P(\{Y \in F_{n,j}\} \mid \mathcal{C}) \;.$$

Monotone Konvergenz unter jedem der Wahrscheinlichkeitsmasse $\widehat{K}(\omega, \cdot)$ zeigt

$$(++) \qquad \text{für jedes feste } \omega \in \Omega : \quad \int_E \varphi_n(y)\, \widehat{K}(\omega, dy) \;\longrightarrow\; \int_E \varphi(y)\, \widehat{K}(\omega, dy) \;, \quad n \to \infty$$

und die Monotonieeigenschaft 10.3 a)iii) angewandt auf $\varphi_n \uparrow \varphi$ zeigt

$$(+++) \qquad E_P(\,\varphi \circ Y \mid \mathcal{C}) \;=\; \sup_n E_P(\,\varphi_n \circ Y \mid \mathcal{C}) \quad P\text{-fast sicher} \;.$$

Diese drei Behauptungen zusammen ergeben

$$\int_E \varphi(y)\, \widehat{K}(\cdot, dy) \quad \text{ist eine Festlegung von} \quad E_P(\,\varphi \circ Y \mid \mathcal{C})$$

im Sinne von 10.5 (beachte: da $\widehat{K}(\cdot,\cdot)$ eine Übergangswahrscheinlichkeit von $(\Omega, \mathcal{C})$ nach $(E, \mathcal{E})$ ist, sind die Abbildungen

$$\omega \;\to\; \int_E \varphi_n(y)\, \widehat{K}(\omega, dy) \;, \quad n \ge 1 \;, \quad \omega \;\to\; \int_E \varphi(y)\, \widehat{K}(\omega, dy)$$

$\mathcal{C}$-messbare Funktionen $\Omega \to [0, \infty]$).

3) Setzt man nun $\varphi \ge 0$ und $\varphi \circ Y \in L^1(P)$ voraus, so ist die bedingte Erwartung $E_P(\,\varphi \circ Y \mid \mathcal{C})$ eine Funktion in $L^1(\Omega, \mathcal{C}, P)$, folglich muss für die in (++) gewählte Festlegung

$$N \;:=\; \left\{ \omega \in \Omega \;:\; \int_E \varphi(y)\, \widehat{K}(\omega, dy) \;=\; +\infty \right\}$$

eine P-Nullmenge in $\mathcal{C}$ sein.

4) Nach Zerlegung $\varphi = \varphi^+ - \varphi^-$ ist wegen 3) die Aussage a) für allgemeines $\varphi \in L^1(E, \mathcal{E}, P^Y)$ bewiesen. $\square$

Auf polnischen Räumen $(\Omega, \mathcal{A})$ gilt als Spezialfall (betrachte $\Omega = E$, $\mathcal{A} = \mathcal{E}$, $Y = id_{|\Omega}$ in 10.32):

10.32' Korollar: Sei P ein Wahrscheinlichkeitsmass auf einem polnischen Raum $(\Omega, \mathcal{A})$, $\mathcal{C}$ eine Sub-σ-Algebra von $\mathcal{A}$, sei $f : E \to I\!R$ eine $\mathcal{A}$-messbare Funktion mit $f \in L^1(P)$.

Wähle eine reguläre Version $K(\cdot,\cdot)$ der bedingten Wahrscheinlichkeit $(\omega, A) \to K(\omega, A) = P(A|\mathcal{C})(\omega)$. Dann ist $N := \{\omega \in \Omega : \int |f|(\omega')\, K(\omega, d\omega') = \infty\}$ eine P-Nullmenge in $\mathcal{C}$, und

$$\omega \;\longrightarrow\; g(\omega) \;:=\; 1_{N^c}(\omega) \int f(\omega')\, K(\omega, d\omega')$$

ist eine Festlegung von $E_P(f|\mathcal{C})$.

Satz 10.32 und sein Korollar 10.32' erklären, warum – wie in 10.4 erwähnt – die Abbildung $f \to E_P(f|\mathcal{C})$ die typischen Eigenschaften eines Integrals (Linearität, Monotonie, aufsteigende Konvergenz) besitzt: die rechten Seiten sind auf 'guten' Räumen Integrale bezüglich einer Übergangswahrscheinlichkeit. Eine nützliche Variante von 10.32 auf dem Faktorisierungslevel ist:

10.33 Satz: Unter den Voraussetzungen und Bezeichnungen aus 10.32 b) sei $\widehat{\widehat{K}}(\cdot,\cdot)$ eine reguläre Version der bedingten Verteilung $P^{Y|X=\cdot}(\cdot)$. Dann gilt:

a) Ist $\varphi : T\times E \to I\!R$ eine $\mathcal{T}\otimes\mathcal{E}$-messbare Funktion und $\varphi(X,Y) \in L^1(P)$, so gilt:

$$x \;\longrightarrow\; \int_E \varphi(x,y)\, \widehat{\widehat{K}}(x,dy) \quad \text{ist eine Festlegung von } \; E(\varphi(X,Y)|X=\cdot)$$

b) Man kann den Erwartungswert der Zufallsvariable $\varphi(X,Y)$ zweistufig berechnen:

$$E_P(\varphi(X,Y)) \;=\; \int_T \int_E \varphi(x,y)\, \widehat{\widehat{K}}(x,dy)\, P^X(dx)\, .$$

c) Die Verteilung $Q := P^{(X,Y)} = \mathcal{L}((X,Y)|P)$ auf $(T\times E, \mathcal{T}\otimes\mathcal{E})$ lässt sich wie folgt 'desintegrieren':

$$Q(dx,dy) \;=\; P^X(dx)\, P^{Y|X=x}(dy) \;=\; P^X(dx)\, \widehat{\widehat{K}}(x,dy)\,, \quad (x,y) \in T\times E\, .$$

Beweis: Die Aussage a) ist eine Variante des Beweises von 10.32. Mit der Definition der bedingten Erwartung bzw. der Definition der Faktorisierung folgt daraus

$$\begin{aligned} \int_\Omega \varphi(X,Y)\, dP \;&=\; \int_\Omega E(\varphi(X,Y)|\sigma(X))\, dP \\ &=\; \int_T E(\varphi(X,Y)|X=\cdot)\, dP^X \;=\; \int_T \int_E \varphi(x,y)\, \widehat{\widehat{K}}(x,dy)\, P^X(dx)\, . \end{aligned}$$

Damit ist b) bewiesen, und c) folgt aus dieser Gleichungskette als Spezialfall (betrachte Indikatorfunktionen $\varphi = 1_G,\; G \in \mathcal{T}\otimes\mathcal{E}$). □

Man beachte, dass in den letzten beiden Sätzen 10.32 und 10.33 nicht gefordert wurde, dass der Raum $(E,\mathcal{E})$, in dem die Zufallsvariable Y Werte annimmt, polnisch sei: diese zwei Sätze sind anwendbar, wann immer – zum Beispiel auch elementar nach 10.22 oder mit Hilfe von bedingten Dichten nach 10.23 – eine reguläre Version der bedingten Verteilung von Y gegeben X gefunden werden kann. Ist jedoch $(E, \mathcal{E})$ polnisch, so ist die Existenz einer regulären Version nach 10.27 gesichert.

10.34 Korollar: Sei $(E, \mathcal{E})$ ein polnischer Raum, sei $(T, \mathcal{T})$ ein beliebiger messbarer Raum. Auf dem Produktraum $(E \times T, \mathcal{E} \otimes \mathcal{T})$ mit kanonischer Variable $(Y, X) := \mathrm{id}\,|_{E \times T}$ kann jedes Wahrscheinlichkeitsmass Q durch

$$Q(dy, dx) \;=\; Q^X(dx)\, Q^{Y|X=x}(dy)$$

desintegriert werden.

Beweis: Auf $(\Omega, \mathcal{A}, Q)$ gegeben durch $\Omega := E \times T$, $\mathcal{A} := \mathcal{E} \otimes \mathcal{T}$ liefert Satz 10.30 die Existenz einer regulären Version der bedingten Verteilung von Y gegeben $X = \cdot$; der Rest ist 10.33 c). $\square$

10.35 Übungsaufgabe: Seien $(E, \mathcal{E}, \mu)$, $(T, \mathcal{T}, \nu)$ messbare Räume versehen mit einem σ-endlichen Mass. Ist Q ein Wahrscheinlichkeitsmass auf dem Produktraum $(E \times T, \mathcal{E} \otimes \mathcal{T})$ mit

$$Q \;\ll\; \mu \otimes \nu$$

und ist $(Y, X) := \mathrm{id}\,|_{E \times T}$ die kanonische Variable auf dem Produktraum, so erhält man wie in 10.23 eine Paardichte $(y, x) \to f(y, x)$ von $Q = Q^{(Y,X)}$ bezüglich $\mu \otimes \nu$, und daraus durch Ausintegrieren eine Randdichte $x \to f_X(x)$ von $Q^X(dx)$ bezüglich $\nu(dx)$ zusammen mit einer bedingten Dichte

$$y \;\longrightarrow\; f^{Y|X=x}(y) \;:=\; 1_{\{f_X > 0\}}(x)\, \frac{f(y, x)}{f_X(x)} + 1_{\{f_X = 0\}}(x)\, g(y)$$

von $Q^{Y|X=x}(dy)$ bezüglich $\mu(dy)$, für jedes $x \in T$ (mit einer beliebigen Wahrscheinlichkeitsdichte g auf $(E, \mathcal{E})$ als 'default value'). Zum Beweis verifiziere man wie in 10.23, dass

$$K(x, dy) \;:=\; f^{Y|X=x}(y)\, \mu(dy)$$

eine Übergangswahrscheinlichkeit von $(T, \mathcal{T})$ nach $(E, \mathcal{E})$ und eine Festlegung der bedingten Verteilung von Y gegeben $X = \cdot$ ist. Satz 10.33 für Funktionen $\varphi = \varphi \circ (Y, X)$ in $L^1(E \times T, \mathcal{E} \otimes \mathcal{T}, Q)$

$$E_Q(\, \varphi(Y, X) \mid X = x\,) \;=\; \int_E \varphi(y, x)\, f^{Y|X=x}(y)\, \mu(dy)\,, \quad x \in T$$

liefert dann Festlegungen der Faktorisierung der bedingten Erwartung durch Ausintegieren mittels der bedingten Dichte $f^{Y|X=x}(\cdot)$. $\square$