

8 Bedingte Erwartungen und Wahrscheinlichkeiten (Version Juni 2017)

8.1 Definition der bedingten Erwartung und grundlegende Eigenschaften

Erinnerung:

Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum und $(A_i)_{i=1,2,\dots}$ eine Partition desselben, d.h. $A_i \in \mathcal{F}, A_i \cap A_j = \emptyset$ für $i, j \in \mathbf{N}, i \neq j$ und $\Omega = \bigcup_{i=1}^{\infty} A_i$. Dann heißt im Fall $\mathbf{P}(A_i) > 0, A \in \mathcal{F}$ $\mathbf{P}(A|A_i) := \frac{\mathbf{P}(A \cap A_i)}{\mathbf{P}(A_i)}$ *bedingte Wahrscheinlichkeit* von A gegeben A_i und das dadurch auf (Ω, \mathcal{F}) definierte W-Maß $\mathbf{P}(\cdot|A_i)$ *bedingtes Wahrscheinlichkeitsmaß* (von \mathbf{P} bezüglich A_i).

Naheliegender ist die folgende Definition eines bedingten Erwartungswertes gegeben A_i :

Definition 8.1. Sei X eine (reelle) Zufallsgröße auf $(\Omega, \mathcal{F}, \mathbf{P})$ mit $\mathbf{E}|X| < \infty$. Ist $A_i \in \mathcal{F}$ und $\mathbf{P}(A_i) > 0$, so heißt

$$\mathbf{E}(X|A_i) := \int X \, d\mathbf{P}(\cdot|A_i)$$

bedingter Erwartungswert von X gegeben A_i .

Bemerkung 8.2. Nach Definition von $\mathbf{P}(\cdot|A_i)$ kann man $\mathbf{E}(X|A_i)$ auch als $\frac{1}{\mathbf{P}(A_i)} \int_{A_i} X \, d\mathbf{P}$ schreiben. Also ist $\mathbf{E}(X|A_i)$ der Mittelwert von X (bzgl. \mathbf{P}) auf der Menge A_i .

Wir fassen nun bedingte Erwartungswerte als Zufallsgrößen auf:

Sei wie vorher $(A_i)_{i \in \mathbf{N}}$ eine Partition von (Ω, \mathcal{F}) und \mathcal{G} die davon erzeugte σ -Algebra, d.h. $\mathcal{G} = \left\{ \bigcup_{i \in I} A_i, I \subseteq \mathbf{N} \right\}$.

Definition 8.3. Sei X eine Zufallsgröße auf $(\Omega, \mathcal{F}, \mathbf{P})$ mit $\mathbf{E}|X| < \infty$.

$\mathbf{E}(X|\mathcal{G})(\omega) := \begin{cases} \int X \, d\mathbf{P}(\cdot|A_i) & \text{wenn } \omega \in A_i \text{ und } \mathbf{P}(A_i) > 0 \\ \text{beliebig aber konstant auf den } A_j & \text{sonst} \end{cases}$
heißt *bedingte Erwartung* von X gegeben \mathcal{G} .

Interpretation:

Gemäß dem Wahrscheinlichkeitsmaß \mathbf{P} wird ein $\omega \in \Omega$ ausgelost. Wir erfahren ω nicht. Es wird uns nur mitgeteilt, in welchem der A_i ω liegt. Daraufhin sollen wir eine Prognose für $X(\omega)$ abgeben. Offenbar ist es vernünftig, als Prognose $\mathbf{E}(X|\mathcal{G})(\omega)$ abzugeben, da dies der \mathbf{P} -Mittelwert über die $\bar{\omega} \in \Omega$ ist, die nach der Information " $\omega \in A_i$ " noch in Frage kommen.

Satz 8.4. *Mit den Voraussetzungen von Definition 8.3 gilt:*

- (i) $\mathbf{E}(X|\mathcal{G})$ ist eine \mathcal{G} -messbare ZG auf $(\Omega, \mathcal{F}, \mathbf{P})$.
- (ii) $\int_A \mathbf{E}(X|\mathcal{G})(\omega) d\mathbf{P}(\omega) = \int_A X(\omega) d\mathbf{P}(\omega)$ für alle $A \in \mathcal{G}$.

Beweis.

- (i) ist klar, da $\mathbf{E}(X|\mathcal{G})$ auf den A_j jeweils konstant ist.
- (ii) Ist A eine \mathbf{P} -Nullmenge, so ist die Gleichheit klar.
Es genügt, die Gleichheit für alle A_i mit $\mathbf{P}(A_i) > 0$ zu zeigen. Sei $\omega_i \in A_i$ beliebig. Mit Bemerkung 8.2 folgt:

$$\begin{aligned} \int_{A_i} \mathbf{E}(X|\mathcal{G})(\omega) d\mathbf{P}(\omega) &= \mathbf{P}(A_i)\mathbf{E}(X|\mathcal{G})(\omega_i) \\ &= \mathbf{P}(A_i)\mathbf{E}(X|A_i) = \int_{A_i} X(\omega) d\mathbf{P}(\omega). \end{aligned}$$

□

Oft interessiert man sich für bedingte Erwartungen bezüglich einer Teil- σ -Algebra \mathcal{G} von \mathcal{F} , die *nicht* durch eine (abzählbare) Partition erzeugt wird (sondern z.B. von einer ZG mit Dichte). Hierauf lässt sich Definition 8.3 nicht problemlos verallgemeinern, Satz 8.4 allerdings schon, wie wir zeigen werden. Wir werden sehen, dass auch allgemein eine Zufallsgröße $\mathbf{E}(X|\mathcal{G})$ existiert, die (i), (ii) in Satz 8.4 erfüllt und diese (bis auf Nullmengen) eindeutig ist.

Für den Beweis des entsprechenden Satzes benötigen wir noch ein Hilfsmittel aus der Maßtheorie, nämlich den Satz von Radon-Nikodym. Dessen Beweis findet sich in jedem Buch zur Maßtheorie. Zunächst führen wir den Begriff der Absolutstetigkeit von Maßen ein.

Definition 8.5. Seien μ und ν Maße auf dem Messraum (Ω, \mathcal{F}) . Dann heißt ν *absolutstetig* bezüglich μ , falls für jede Menge $A \in \mathcal{F}$ mit $\mu(A) = 0$ folgt, dass auch $\nu(A) = 0$ ist. Man schreibt in diesem Fall $\nu \ll \mu$.

Wir erinnern an die Definition 3.5 der Dichte eines Maßes bezüglich eines anderen Maßes.

Satz 8.6. *Seien μ und ν σ -endliche Maße auf dem Messraum (Ω, \mathcal{F}) . Dann hat ν eine Dichte $\frac{d\nu}{d\mu}$ bezüglich μ genau dann wenn $\nu \ll \mu$.*

Die Richtung von links nach rechts ist sehr einfach zu sehen, der Beweis der umgekehrten Richtung ist deutlich schwieriger.

Nun folgt die oben angekündigte Aussage.

Satz und Definition 8.7. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein W-Raum, $\mathcal{G} \subseteq \mathcal{F}$ eine Unter- σ -Algebra und X eine Zufallsgröße (bzgl. \mathcal{F}) mit $\mathbf{E}|X| < \infty$. Dann existiert (bis auf f.s. Gleichheit) genau eine Zufallsgröße X_0 auf $(\Omega, \mathcal{F}, \mathbf{P})$ mit

(i) X_0 ist \mathcal{G} -messbar.

(ii) $\int_A X_0 d\mathbf{P} = \int_A X d\mathbf{P}$ für alle $A \in \mathcal{G}$.

X_0 heißt *bedingte Erwartung* von X gegeben \mathcal{G} und wird mit $\mathbf{E}(X|\mathcal{G})$ bezeichnet.

Beweis. 1. Schritt: Sei zunächst zusätzlich $X \geq 0$. Definiere $Q(A) = \int_A X d\mathbf{P}$ für $A \in \mathcal{G}$. $Q : \mathcal{G} \rightarrow [0, \infty[$ ist dann (wegen $\int_\Omega X d\mathbf{P} < \infty$) ein endliches Maß auf \mathcal{G} . Sei \mathbf{P}_0 die Restriktion von \mathbf{P} auf \mathcal{G} . Dann gilt $Q \ll \mathbf{P}_0$. Also existiert nach dem Satz von Radon-Nikodym eine bis auf f.s. Gleichheit eindeutige \mathcal{G} -messbare nichtnegative Funktion X_0 mit $Q(A) = \int_A X_0 d\mathbf{P}_0 = \int_A X d\mathbf{P}$.

2. Schritt: Ist X eine beliebige Zufallsgröße mit $\mathbf{E}|X| < \infty$, so zerlegt man $X = X^+ - X^-$, wobei $X^+ = X \vee 0$, $X^- = (-X) \vee 0$. Wegen $\mathbf{E}|X^+| < \infty$, $\mathbf{E}|X^-| < \infty$ erhält man mit dem 1. Schritt \mathcal{G} -messbare nichtnegative X'_0 und X''_0 mit

$$\int_A X'_0 d\mathbf{P} = \int_A X^+ d\mathbf{P}, \int_A X''_0 d\mathbf{P} = \int_A X^- d\mathbf{P}, A \in \mathcal{G}.$$

Dann erfüllt $X_0 := X'_0 - X''_0$ (i) und (ii). Ist Y_0 eine weitere Lösung von (i), (ii), so gilt

$$\int_A X_0 d\mathbf{P}_0 = \int_A Y_0 d\mathbf{P}_0, A \in \mathcal{G}.$$

Wählt man speziell $A_1 = \{\omega : X_0(\omega) > Y_0(\omega)\}$ und $A_2 = \{\omega : X_0(\omega) < Y_0(\omega)\}$, so folgt $\mathbf{P}_0(A_1) = \mathbf{P}_0(A_2) = 0$, also $X_0 = Y_0$ f.s. \square

Bemerkung 8.8. Ist $\mathcal{G} = \{\emptyset, \Omega\}$ die triviale σ -Algebra, so folgt offensichtlich $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}X$ für alle $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$.

Satz 8.9. (*Eigenschaften von bedingten Erwartungen*): Sei $(\Omega, \mathcal{F}, \mathbf{P})$ WR, alle X_i, X ZV mit endlichem Erwartungswert, $a, b \in \mathbf{R}$, $\mathcal{G} \subseteq \mathcal{F}$ σ -Algebra.

(i) $\mathbf{E}(aX_1 + bX_2|\mathcal{G}) = a\mathbf{E}(X_1|\mathcal{G}) + b\mathbf{E}(X_2|\mathcal{G})$ f.s.

(ii) $X \geq 0$ f.s. $\Rightarrow \mathbf{E}(X|\mathcal{G}) \geq 0$ f.s.

(iii) X \mathcal{G} -messbar $\Rightarrow \mathbf{E}(X|\mathcal{G}) = X$ f.s.

(iv) $\sigma(X)$ unabhängig von $\mathcal{G} \Rightarrow \mathbf{E}(X|\mathcal{G}) = \mathbf{E}X$ f.s.

(v) $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}X$

(vi) $\mathcal{H} \subseteq \mathcal{G}$ σ -Algebra $\Rightarrow \mathbf{E}(\mathbf{E}(X|\mathcal{H})|\mathcal{G}) = \mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{H}) = \mathbf{E}(X|\mathcal{H})$ f.s.

(vii) $X_1 = X_2$ f.s. $\Rightarrow \mathbf{E}(X_1|\mathcal{G}) = \mathbf{E}(X_2|\mathcal{G})$ f.s.

(viii) $X \equiv a$ f.s. $\Rightarrow \mathbf{E}(X|\mathcal{G}) = a$ f.s.

- (ix) $X_1 \leq X_2$ f.s. $\Rightarrow \mathbf{E}(X_1|\mathcal{G}) \leq \mathbf{E}(X_2|\mathcal{G})$ f.s.
- (x) (Jensensche Ungleichung): Ist $\varphi : \mathbf{R} \rightarrow \mathbf{R}$ konvex und $\varphi(X) \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$, dann gilt $\varphi(\mathbf{E}(X|\mathcal{G})) \leq \mathbf{E}(\varphi(X)|\mathcal{G})$ f.s.
- (xi) (monotone Konvergenz): Wenn $0 \leq X_n \uparrow X$ f.s. $\Rightarrow \mathbf{E}(X_n|\mathcal{G}) \uparrow \mathbf{E}(X|\mathcal{G})$ f.s.
- (xii) (Fatou): $X_n \geq 0, X = \liminf_{n \rightarrow \infty} X_n \Rightarrow \mathbf{E}(X|\mathcal{G}) \leq \liminf_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{G})$ f.s.
- (xiii) (beschr. Konvergenz): Wenn $X_n \rightarrow X$ f.s. und $|X_n| \leq Y \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P}), n \in \mathbf{N} \Rightarrow \lim_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{G}) = \mathbf{E}(X|\mathcal{G})$ f.s. und \mathcal{L}^1 .
- (xiv) $YX \in \mathcal{L}^1, Y$ \mathcal{G} -messbar $\Rightarrow \mathbf{E}(YX|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G})$ f.s.

Beweis.

- (i) zu zeigen: $\int_G (aX_1 + bX_2) d\mathbf{P} = \int_G (a\mathbf{E}(X_1|\mathcal{G}) + b\mathbf{E}(X_2|\mathcal{G})) d\mathbf{P}$ für alle $G \in \mathcal{G}$. Die rechte Seite ist nach Definition gleich $a \int_G X_1 d\mathbf{P} + b \int_G X_2 d\mathbf{P}$, also gleich der linken Seite.
- (ii) wurde im 1. Schritt des Beweises von Satz und Definition 8.7 gezeigt.
- (iii) klar nach Definition.
- (iv) $\int_G X d\mathbf{P} = \mathbf{E}(\mathbf{1}_G X) = (\mathbf{E}\mathbf{1}_G)\mathbf{E}X = \mathbf{P}(G)\mathbf{E}X = \int_G \mathbf{E}X d\mathbf{P}$ für $G \in \mathcal{G}$.
- (v) $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \int_{\Omega} \mathbf{E}(X|\mathcal{G}) d\mathbf{P} = \int_{\Omega} X d\mathbf{P} = \mathbf{E}X$.
- (vi) Alle Ausdrücke sind \mathcal{H} -messbar (der linke nach (iii)). Die Gleichheit von linkem und rechtem Ausdruck folgt aus (iii), die des mittleren und rechten wegen $\int_H \mathbf{E}(X|\mathcal{H}) d\mathbf{P} = \int_H X d\mathbf{P} = \int_H \mathbf{E}(X|\mathcal{G}) d\mathbf{P}$ für alle $H \in \mathcal{H}$ (da $\mathcal{H} \subseteq \mathcal{G}$).
- (vii) klar.
- (viii) klar.
- (ix) folgt aus (i) und (ii).
- (x) Sei \mathcal{A} die Menge der Funktionen $f : \mathbf{R} \rightarrow \mathbf{R}$ mit $f(x) = ax + b$ für $a, b \in \mathbf{Q}$. Dann ist φ konvex gdw. φ affin ist (dann ist die Behauptung klar – sogar mit “=”) oder $\varphi(x) = \sup_{f \in \mathcal{A}, f \leq \varphi} f(x)$ für alle $x \in \mathbf{R}$. Im zweiten Fall gilt für $f \in \mathcal{A}, f \leq \varphi$ mit (i), (viii) und (ix) $f(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(f(X)|\mathcal{G}) \leq \mathbf{E}(\varphi(X)|\mathcal{G})$ f.s. und daher $\varphi(\mathbf{E}(X|\mathcal{G})) = \sup_{f \in \mathcal{A}, f \leq \varphi} f(\mathbf{E}(X|\mathcal{G})) \leq \mathbf{E}(\varphi(X)|\mathcal{G})$ f.s., wobei das letzte “ \leq ” folgt, da \mathcal{A} abzählbar ist.

- (xi) Wegen (ix) existiert $Z = \lim_{n \rightarrow \infty} \mathbf{E}(X_n | \mathcal{G})$ f.s. (evtl. ∞) und Z hat eine \mathcal{G} -messbare Version mit $\mathbf{E}(X_n | \mathcal{G}) \uparrow Z$ f.s. Für $G \in \mathcal{G}$ gilt nach dem Satz über die monotone Konvergenz $\int_G Z d\mathbf{P} = \lim_{n \rightarrow \infty} \int_G \mathbf{E}(X_n | \mathcal{G}) d\mathbf{P} = \lim_{n \rightarrow \infty} \int_G X_n d\mathbf{P} = \int_G X d\mathbf{P}$, also $Z = \mathbf{E}(X | \mathcal{G})$ f.s.
- (xii) Sei $Y_n := \inf_{m \geq n} X_m$. Dann gilt $Y_n \uparrow X$ und mit (xi) und (ix) $\mathbf{E}(X | \mathcal{G}) = \lim_{n \rightarrow \infty} \mathbf{E}(Y_n | \mathcal{G}) \leq \lim_{n \rightarrow \infty} \mathbf{E}(X_n | \mathcal{G})$.
- (xiii) $Y_n := \sup_{m \geq n} |X_m - X|$. Dann gilt $Y_n \downarrow 0$ f.s. und $0 \leq Y_n \leq 2Y$ f.s. und somit $0 \leq (2Y - Y_n) \uparrow 2Y$ f.s. Nach (xi) folgt $\mathbf{E}(2Y - Y_n | \mathcal{G}) \uparrow \mathbf{E}(2Y | \mathcal{G})$ f.s., also mit (i) $\mathbf{E}(Y_n | \mathcal{G}) \downarrow 0$. Mit (i), (ix) und (x) folgt $|\mathbf{E}(X | \mathcal{G}) - \mathbf{E}(X_n | \mathcal{G})| = |\mathbf{E}(X - X_n | \mathcal{G})| \leq \mathbf{E}(|X - X_n| | \mathcal{G}) \leq \mathbf{E}(Y_n | \mathcal{G}) \downarrow 0$ f.s. Daraus folgt die f.s. Konvergenz. Bildet man Erwartungswerte, so folgt mit (v) $\mathbf{E}|\mathbf{E}(X | \mathcal{G}) - \mathbf{E}(X_n | \mathcal{G})| \leq \mathbf{E}Y_n \downarrow 0$; also die \mathcal{L}^1 -Konvergenz.
- (xiv) zu zeigen: $\int_G Y \mathbf{E}(X | \mathcal{G}) d\mathbf{P} = \int_G Y X d\mathbf{P}$ für alle $G \in \mathcal{G}$. Für $Y = \mathbf{1}_{\tilde{G}}$, $\tilde{G} \in \mathcal{G}$ gilt dies und damit auch für alle einfachen \mathcal{G} -messbaren Y . Für \mathcal{G} -messbares Y mit $XY \in \mathcal{L}^1$ wähle man einfache \mathcal{G} -messbare Y_n mit $Y_n^+ \uparrow Y^+$ und $Y_n^- \uparrow Y^-$. Dann gilt $|Y_n X| \leq |Y X| \in \mathcal{L}^1$ und mit (xiii) folgt $\mathbf{E}(Y X | \mathcal{G}) = \lim_{n \rightarrow \infty} \mathbf{E}(Y_n X | \mathcal{G}) = \lim_{n \rightarrow \infty} Y_n \mathbf{E}(X | \mathcal{G}) = Y \mathbf{E}(X | \mathcal{G})$ f.s.

□

Bemerkung 8.10. Man beachte, dass wir in (x) insbesondere die gewöhnliche Jensensche Ungleichung und in (xii) das gewöhnliche Lemma von Fatou (jedenfalls für Wahrscheinlichkeitsräume und reellwertige Funktionen) mit Hilfe des Satzes von der monotonen Konvergenz bewiesen haben (man setze jeweils $\mathcal{G} := \{\emptyset, \Omega\}$).

Ist \mathcal{G} die von einer (Ω', \mathcal{F}') -wertigen Zufallsgröße Z erzeugte σ -Algebra, so schreibt man auch $\mathbf{E}(X | Z)$ statt $\mathbf{E}(X | \sigma(Z))$. Häufig begegnet man Ausdrücken der Form $\mathbf{E}(X | Z = z)$ mit $z \in \Omega'$. Wie kann man diese sauber definieren? Die Antwort liefert der folgende Faktorisierungssatz.

Satz 8.11. *Sei Ω eine nichtleere Menge, (Ω', \mathcal{F}') ein Messraum, $f : \Omega \rightarrow \bar{\mathbf{R}}$ eine numerische Funktion und $g : \Omega \rightarrow \Omega'$ eine Abbildung. Dann ist f genau dann $\sigma(g)$ -messbar, wenn eine \mathcal{F}' -messbare numerische Funktion $h : \Omega' \rightarrow \bar{\mathbf{R}}$ existiert mit $f = h \circ g$.*

Beweis. Die eine Richtung ist klar: wenn ein solches h existiert, dann ist f $\sigma(g)$ -messbar. Die andere Richtung ist nur wenig schwieriger (siehe z.B. Klenke, *Wahrscheinlichkeitstheorie*, Seite 40). □

Nun zurück zu obiger Frage. Wählt man im Faktorisierungssatz $f(\omega) = \mathbf{E}(X | Z)(\omega)$ und $g = Z$, so folgt die Existenz einer \mathcal{F}' -messbaren Funktion $h : \Omega' \rightarrow \mathbf{R}$ mit

$\mathbf{E}(X|Z)(\omega) = (h \circ Z)(\omega)$, und man definiert $\mathbf{E}(X|Z = z) := h(z)$. h ist im allgemeinen *nicht* eindeutig bestimmt, sondern nur \mathbf{P}_Z f.s..

8.2 Bedingte Erwartungen als orthogonale Projektion

Wir zeigen, dass man bedingte Erwartungen als orthogonale Projektionen (oder beste Approximationen) in einem geeigneten Hilbertraum interpretieren kann. Dies funktioniert allerdings nur im Fall $\mathbf{E}(X^2) < \infty$. Daher wird dieser Zugang seltener gewählt als der vorher dargestellte. Wir erwähnen, dass der geometrische Zugang in einigen Fällen zur expliziten Berechnung von bedingten Erwartungen verwendet werden kann, insbesondere im Fall von Gaußprozessen, auf die wir am Ende des Kapitels noch etwas eingehen werden (vgl. Literatur zur Vorhersagetheorie z.B. Doob: Stochastic Processes (1953)). Wir verwenden im folgenden einige grundlegende Eigenschaften von Hilberträumen.

Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein Wahrscheinlichkeitsraum, \mathcal{G} eine Teil- σ -Algebra von \mathcal{F} und $H = L^2(\Omega, \mathcal{F}, \mathbf{P})$, $K = L^2(\Omega, \mathcal{G}, \mathbf{P})$. Dabei ist $L^2(\Omega, \mathcal{D}, \mathbf{P})$ die Menge der Äquivalenzklassen von $\mathcal{L}^2(\Omega, \mathcal{D}, \mathbf{P})$ bezüglich f.s. Gleichheit. $L^2(\Omega, \mathcal{D}, \mathbf{P})$ ist ein Hilbertraum mit Skalarprodukt $\langle X, Y \rangle := \mathbf{E}(XY)$ ($\mathcal{D} \in \{\mathcal{F}, \mathcal{G}\}$).

Sei $\|Z\|_2 := (\langle Z, Z \rangle)^{1/2} = (\int_{\Omega} Z^2 d\mathbf{P})^{1/2}$ die L^2 -Norm auf H . K ist ein abgeschlossener Unterraum von H . Daher existiert zu jedem $X \in H$ genau ein $V \in K$ mit

$$\|X - V\|_2 = \inf_{Y \in K} \|X - Y\|_2.$$

Dieses V heißt *orthogonale Projektion von X auf K* oder *beste Approximation von X in K* .

Wir zeigen, dass V mit $\mathbf{E}(X|\mathcal{G})$ übereinstimmt, genauer: $\mathbf{E}(X|\mathcal{G})$ ist ein Element der Äquivalenzklasse V .

Satz 8.12. *Mit den obigen Bezeichnungen sei $X \in H$ und V die orthogonale Projektion von X auf K . Dann gilt $V = \mathbf{E}(X|\mathcal{G})$ (im obigen Sinn).*

Beweis. Da V orthogonale Projektion von X auf K ist und $\mathbf{1}_G \in K$ für $G \in \mathcal{G}$, gilt

$$\langle X - V, \mathbf{1}_G \rangle = 0.$$

Weiter ist V als Element von K \mathcal{G} -messbar, also

- (i) V ist \mathcal{G} -messbar.
- (ii) $\int_G V d\mathbf{P} = \int_G X d\mathbf{P}$ für alle $G \in \mathcal{G}$.

□

Bemerkung 8.13. Man kann auch für $X \in L^1$ $\mathbf{E}(X|\mathcal{G})$ “konstruktiv” durch Approximation erklären. Idee: Man schneide X ab : $X_n := (X \wedge n) \vee (-n)$. Dann ist $X_n \in L^2$. Ist V_n die nach Satz 8.12 erklärte orthogonale Projektion von X_n , so zeigt man, dass V_n in L^1 (aber im allgemeinen nicht in L^2 !) eine Cauchyfolge bildet und daher gegen ein $V \in L^1$ konvergiert. Man zeigt dann, dass $V = \mathbf{E}(X|\mathcal{G})$ ist.

8.3 (Reguläre) bedingte Wahrscheinlichkeiten

Wie erklärt man allgemein bedingte Wahrscheinlichkeiten?

Definition 8.14. Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein W-Raum und \mathcal{G} eine Teil- σ -Algebra von \mathcal{F} . Dann heißt für $A \in \mathcal{F}$

$$\mathbf{P}(A|\mathcal{G}) := \mathbf{E}(\mathbf{1}_A|\mathcal{G})$$

bedingte Wahrscheinlichkeit von A gegeben \mathcal{G} .

Man sieht unschwer, dass im Fall, dass \mathcal{G} durch eine Partition $(A_i)_{i \in \mathbf{N}}$ erzeugt ist, $\mathbf{P}(A|\mathcal{G})(\omega) = \mathbf{P}(A|A_i)$ gilt, falls $\omega \in A_i$ und $\mathbf{P}(A_i) > 0$. Also ist die jetzige Definition verträglich mit der früheren.

Es stellt sich die Frage, ob $\mathbf{P}(\cdot|\mathcal{G})(\omega)$ ein W-Maß auf (Ω, \mathcal{F}) ist. Diese Frage ist im allgemeinen mit *nein* zu beantworten. Unter zusätzlichen Voraussetzungen (aber nicht immer!) ist es aber möglich, eine solche Version von $\mathbf{P}(\cdot|\mathcal{G})(\omega)$ zu wählen, die für jedes ω ein W-Maß ist. Wir werden diese Aussage in Satz 8.16 formulieren. In jedem Fall gilt aber die folgende Aussage:

Satz 8.15. *Unter den Voraussetzungen von Definition 8.14 gilt*

- (i) $\mathbf{P}(\Omega|\mathcal{G}) = 1$ *f.s.*
- (ii) $\mathbf{P}(\emptyset|\mathcal{G}) = 0$ *f.s.*
- (iii) für alle $A \in \mathcal{F}$ gilt $1 \geq \mathbf{P}(A|\mathcal{G}) \geq 0$ *f.s.*
- (iv) $A_i \in \mathcal{F}, i \in \mathbf{N}, A_i \cap A_j = \emptyset$ für $i \neq j \Rightarrow$
 $\mathbf{P}(\cup_{i=1}^{\infty} A_i|\mathcal{G}) = \sum_{i=1}^{\infty} \mathbf{P}(A_i|\mathcal{G})$ *f.s.*

Beweis.

- (i), (ii) folgen aus Satz 8.9 (viii)
- (iii) folgt aus (i), (ii) und Satz 8.9 (ix) □
- (iv) folgt aus Satz 8.9 (xi).

Satz 8.16. *Sei (Ω, d) ein polnischer Raum, d.h. ein vollständiger, metrischer, separabler Raum (separabel heißt: es existiert eine abzählbare dichte Teilmenge von Ω). Sei \mathcal{F} die Borel- σ -Algebra auf Ω und \mathbf{P} ein W-Maß darauf. Weiter sei \mathcal{G} eine Unter- σ -Algebra von \mathcal{F} . Dann existiert eine Funktion $K : \Omega \times \mathcal{F} \rightarrow \mathbf{R}$ mit*

(i) $K(\omega, \cdot)$ ist ein W -Maß für alle $\omega \in \Omega$

(ii) $\left. \begin{array}{l} K(\cdot, A) = \mathbf{P}(A|\mathcal{G}) \text{ f.s.} \\ K(\cdot, A) \text{ ist } \mathcal{G}\text{-messbar} \end{array} \right\} \text{ für alle } A \in \mathcal{F}$

d.h. ein Markoffkern K von (Ω, \mathcal{G}) nach (Ω, \mathcal{F}) mit $K(\cdot, A) = \mathbf{P}(A|\mathcal{G})$ f.s. für jedes $A \in \mathcal{F}$.

Beweis. Der Satz ist ein Spezialfall des folgenden Satzes 8.19. \square

Definition 8.17. Eine Funktion K mit (i) und (ii) wie im letzten Satz bezeichnet man als *reguläre bedingte Wahrscheinlichkeit* gegeben \mathcal{G} .

Bemerkung 8.18. In der Monographie von Doob, *Stochastic Processes*, S. 624, findet sich ein Beispiel, für das *keine* reguläre bedingte Wahrscheinlichkeit existiert.

Satz und Definition 8.19. Sei X eine ZG auf einem WR $(\Omega, \mathcal{F}, \mathbf{P})$ mit Werten in einem polnischen Raum (E, d) mit $\mathcal{E} = \mathcal{B}_E$ und sei $\mathcal{G} \subseteq \mathcal{F}$ eine Teil- σ -Algebra. Dann existiert ein Markovkern K von (Ω, \mathcal{G}) nach (E, \mathcal{E}) mit

$$K(\omega, A) = \mathbf{P}(X^{-1}(A)|\mathcal{G})(\omega) \text{ f.s. für jedes (feste) } A \in \mathcal{E}.$$

K heißt *reguläre bedingte Verteilung* von X gegeben \mathcal{G} .

Bemerkung 8.20. Satz 8.16 folgt aus Satz 8.19, indem man $E = \Omega, \mathcal{E} = \mathcal{F}$ und X gleich der Identität auf Ω setzt.

Beweis. (nach Durrett (1996): Probability: Theory and Examples, S. 230).

1. Fall: $(E, \mathcal{E}) = (\mathbf{R}, \mathcal{B})$

Für $q \in \mathbf{Q}$ wähle eine Version von

$$\overline{G}(q, \omega) := \mathbf{P}(X \leq q|\mathcal{G})(\omega) = \mathbf{P}(\{\tilde{\omega} : X(\tilde{\omega}) \leq q\}|\mathcal{G})(\omega).$$

Sei

$$\Omega_1 := \{\omega : q \mapsto \overline{G}(q, \omega) \text{ ist monoton nicht fallend}\},$$

$$\Omega_2 := \{\omega : \lim_{q \in \mathbf{Q}, q \rightarrow \infty} \overline{G}(q, \omega) = 1\},$$

$$\Omega_3 := \{\omega : \lim_{q \in \mathbf{Q}, q \rightarrow -\infty} \overline{G}(q, \omega) = 0\}.$$

Dann gilt für $i = 1, 2, 3$ $\Omega_i \in \mathcal{G}$ und $\mathbf{P}(\Omega_i) = 1$.

Sei Φ eine beliebige Verteilungsfunktion auf \mathbf{R} . Definiere

$$G(q, \omega) := \begin{cases} \overline{G}(q, \omega) & \omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3 \\ \Phi(q) & \omega \notin \Omega_1 \cap \Omega_2 \cap \Omega_3. \end{cases}$$

$G(q, \cdot)$ ist \mathcal{G} -messbar für jedes $q \in \mathbf{Q}$. Nun definieren wir

$$F(x, \omega) := \inf\{G(q, \omega) : q > x\}, \quad x \in \mathbf{R}, \omega \in \Omega.$$

Dann gilt

- (i) $F(\cdot, \omega)$ ist eine Verteilungsfunktion für jedes $\omega \in \Omega$.
- (ii) $F(x, \omega) = \mathbf{P}(X \leq x | \mathcal{G})(\omega)$ f.s. für jedes $x \in \mathbf{R}$.
- (iii) $F(x, \cdot)$ ist \mathcal{G} -messbar für jedes $x \in \mathbf{R}$.

Punkt (i) ist klar. Um (ii) und (iii) für ein festes $x \in \mathbf{R}$ zu zeigen, wählen wir eine monoton fallende Folge von $q_n \in \mathbf{Q}$ mit $q_n > x$ und $\lim_{n \rightarrow \infty} q_n = x$. Dann gilt $\lim_{n \rightarrow \infty} G(q_n, \omega) = F(x, \omega)$ (woraus (iii) folgt) und $G(q_n, \omega) = \mathbf{P}(X \leq q_n | \mathcal{G})(\omega) \rightarrow \mathbf{P}(X \leq x | \mathcal{G})(\omega)$ f.s. nach dem Satz von der beschränkten Konvergenz für bedingte Erwartungen. Somit folgt (ii).

Nun definieren wir für jedes $\omega \in \Omega$ ein W-maß $K(\omega, \cdot)$ auf $(\mathbf{R}, \mathcal{B})$ durch $K(\omega, (-\infty, x]) := F(x, \omega)$ und

$$\mathcal{H} := \{B \in \mathcal{B} : K(\omega, B) = \mathbf{P}(X \in B | \mathcal{G})(\omega) \text{ f.s. und } K(\cdot, B) \text{ ist } \mathcal{G}\text{-messbar}\}.$$

Offensichtlich ist \mathcal{H} ein Dynkinsystem, welches das \cap -stabile Erzeugendensystem $\{(-\infty, x] : x \in \mathbf{R}\}$ von \mathcal{B} enthält. Daher ist $\mathcal{H} = \mathcal{B}$ und der erste Fall gezeigt.

2. Fall:

Ohne Beweis benutzen wir die Tatsache, dass zu jedem polnischen Raum E ein $B \in \mathcal{B}$ und eine bijektive, messbare Abbildung $\varphi : E \rightarrow B$ existiert, so dass auch φ^{-1} messbar ist (d.h. (E, \mathcal{E}) und $(B, \mathcal{B}|_B)$ sind isomorph als messbare Räume). Somit folgt der allgemeine aus dem 1. Fall. \square

Das folgende Korollar werden wir in der Vorlesung Wahrscheinlichkeitstheorie 2 benötigen (es ist aber auch von eigenständigem Interesse).

Korollar 8.21. *Seien (E_1, \mathcal{E}_1) und (E_2, \mathcal{E}_2) messbare Räume, wobei (E_2, d) polnisch sei und $\mathcal{E}_2 = \mathcal{B}_{E_2}$. Sei \mathbf{P} ein W-maß auf $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ und $Q := \mathbf{P}\pi_1^{-1}$ ($\pi_1 =$ Projektion auf die 1. Komponente (E_1, \mathcal{E}_1)). Dann existiert ein Markovkern K von (E_1, \mathcal{E}_1) nach (E_2, \mathcal{E}_2) , so dass $\mathbf{P} = Q \otimes K$.*

Beweis. Sei $X := \pi_2 : E_1 \times E_2 \rightarrow E_2$ die Projektion auf die 2. Komponente E_2 und $\mathcal{G} := \{A \times E_2 : A \in \mathcal{E}_1\} = \sigma(\pi_1)$. Nach Satz 8.19 existiert ein Markovkern \overline{K} von $(E_1 \times E_2, \mathcal{G})$ nach (E_2, \mathcal{E}_2) mit $\overline{K}(\omega, B) = \mathbf{P}(X \in B | \mathcal{G})(\omega)$ f.s. ($\omega \in E_1 \times E_2$) für jedes $B \in \mathcal{E}_2$. Definiere $K : E_1 \times \mathcal{E}_2 \rightarrow [0, 1]$ durch

$$K(e_1, B) := \overline{K}((e_1, e_2), B), \text{ wobei } e_2 \in E_2 \text{ beliebig.}$$

Da $\bar{K}(\cdot, B)$ \mathcal{G} -messbar ist, hängt $K(e_1, B)$ nicht von der Wahl von e_2 ab. Für $A \in \mathcal{E}_1$ und $B \in \mathcal{E}_2$ gilt ($\tilde{e}_2 \in E_2$ sei fest gewählt)

$$\begin{aligned} (Q \otimes K)(A \times B) &= \int_A K(e_1, B) dQ(e_1) = \int_A \bar{K}((e_1, \tilde{e}_2), B) dQ(e_1) \\ &= \int_{A \times E_2} \bar{K}((e_1, \tilde{e}_2), B) d\mathbf{P}(e_1, e_2) = \int_{A \times E_2} \bar{K}((e_1, e_2), B) d\mathbf{P}(e_1, e_2) \\ &= \int \mathbf{1}_{A \times E_2}(e_1, e_2) \mathbf{1}_{E_1 \times B}(e_1, e_2) d\mathbf{P}(e_1, e_2) = \mathbf{P}(A \times B), \end{aligned}$$

wobei das dritte Gleichheitszeichen der Formel aus dem Transformationssatz mit $\varphi = \pi_1$ folgt. Also stimmen \mathbf{P} und $Q \otimes K$ auf einem \cap -stabilen Erzeuger überein und es folgt $\mathbf{P} = Q \otimes K$. \square

Die folgende Aussage zeigt, dass sich bedingte Erwartungen als Erwartungswerte bezüglich regulärer bedingter Wahrscheinlichkeiten (im Fall von deren Existenz) darstellen lassen.

Satz 8.22. *Sei $(\Omega, \mathcal{F}, \mathbf{P})$ ein W -Raum, $\mathcal{G} \subseteq \mathcal{F}$ eine Teil- σ -Algebra und $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbf{P})$. Es existiere eine reguläre bedingte Wahrscheinlichkeit K gegeben \mathcal{G} . Dann gilt*

$$\mathbf{E}(X|\mathcal{G})(\omega) = \int_{\Omega} X(\omega') K(\omega, d\omega') \quad f.s.$$

Beweis. Zu zeigen ist die \mathcal{G} -Messbarkeit der rechten Seite und

$$\int_G \int_{\Omega} X(\omega') K(\omega, d\omega') d\mathbf{P}(\omega) = \int_G X(\omega) d\mathbf{P}(\omega)$$

für alle $G \in \mathcal{G}$.

Klar ist beides im Fall $X = \mathbf{1}_F$ mit $F \in \mathcal{F}$, dann ist $\int_{\Omega} X(\omega') K(\omega, d\omega') = K(\omega, F)$ \mathcal{G} -messbar und beide Seiten der Gleichung gleich $\mathbf{P}(F \cap G)$. Für $X \geq 0$ folgt die Aussage mit maßtheoretischer Induktion.

Mittels der Zerlegung $X = X^+ - X^-$ folgt der allgemeine Fall. \square

8.4 Bedingte Erwartungen für Gaußsysteme, Vorhersageprobleme

Wir sahen, dass man bedingte Erwartungen von \mathcal{L}^2 -Zufallsgrößen als orthogonale Projektion in einem Hilbertraum auffassen kann. Leider sind diese Hilberträume

in der Regel ∞ -dimensional und die Berechnung der Projektion damit keineswegs problemlos. Selbst wenn z.B. $\mathcal{L}(X) = \mathcal{N}(0, 1)$ ist, ist die Dimension von $L^2(\Omega, \sigma(X), \mathbf{P})$ unendlich, da man unendlich viele beschränkte linear unabhängige Funktionen von X bilden kann. Die Berechnung von bedingten Erwartungen bei Gaußprozessen erweist sich dagegen (trotz der ∞ -Dimensionalität der zugehörigen L^2 -Räume) als machbar, jedenfalls bei Prozessen mit endlicher Indexmenge. Hier handelt es sich um recht einfache Anwendungen der linearen Algebra.

Definition 8.23. Ein $(\mathbf{R}, \mathcal{B})$ -wertiger stochastischer Prozess d.h. eine Familie $(X_t)_{t \in I}$ von reellen Zufallsvariablen auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ heißt *Gaußprozess* oder *Gaußsystem*, wenn jedes endliche Teilsystem mehrdimensional normalverteilt ist.

Bemerkung 8.24. Kennt man bei einem Gaußprozess $(X_t)_{t \in I}$ die Erwartungswertfunktion $\mu_t := \mathbf{E}X_t, t \in I$ und die Kovarianzfunktion $r_{s,t} := \text{cov}(X_s, X_t), s, t \in I$, so ist die Verteilung von $(X_t)_{t \in I}$ dadurch eindeutig festgelegt. Mit dem Satz von Kolmogorov, den wir in der Vorlesung Wahrscheinlichkeitstheorie 2 kennenlernen werden, sieht man leicht, dass es zu beliebiger Funktion $\mu : I \rightarrow \mathbf{R}$ und positiv semidefiniter Funktion $r : I \times I \rightarrow \mathbf{R}$ genau ein W -Maß \hat{P} auf $(\mathbf{R}^I, \mathcal{B}^I)$ gibt, welches die Verteilung eines Gaußprozesses $(X_t)_{t \in I}$ mit Erwartungswertfunktion μ und Kovarianzfunktion r ist (man definiert sich aus μ und r die endlich-dimensionalen Verteilungen, zeigt, dass sie *konsistent* sind).

Satz 8.25. Sei $(X_t)_{t \in I}$ ein Gaußprozess auf $(\Omega, \mathcal{F}, \mathbf{P})$ und $I = \bigcup_{s \in S} J_s$ für eine Indexmenge S . Dann sind äquivalent:

a) $s_1, s_2 \in S, s_1 \neq s_2, t_1 \in J_{s_1}, t_2 \in J_{s_2} \Rightarrow X_{t_1}$ und X_{t_2} sind unkorreliert.

b) $\sigma(X_t, t \in J_s), s \in S$ sind unabhängig.

Spezialfall:

$S = I$ und $J_s = \{s\}$, dann besagt Satz 8.25, dass bei einem Gaußprozess die (X_t) unabhängig sind genau dann, wenn sie paarweise unkorreliert sind.

Beweis.

$b) \Rightarrow a)$ ist klar, da $X_t \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbf{P})$ für alle $t \in I$ (aus Unabhängigkeit folgt im Fall der quadratischen Integrierbarkeit die Unkorreliertheit).

$a) \Rightarrow b)$. Nach einem Resultat aus Kapitel 4 genügt es, für endliche Teilmengen S_0 von S und endliche Teilmengen von $J_s, s \in S_0$ die Unabhängigkeit zu zeigen. Betrachtet man die gemeinsame Verteilung der X_t zu den endlich vielen gewählten $t \in I$, so sieht man, dass ihre Kovarianzmatrix nach a) eine Blockstruktur

hat, d.h. $\text{cov}(X_t, X_u) = 0$, wenn t und u in disjunkten J_s liegen. Wegen der Eindeutigkeit der gemeinsamen Verteilung bei gegebener charakteristischer Funktion folgt die Behauptung. \square

Sei $(X_t)_{t \in I}$ ein stochastischer Prozess mit $X_t \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, $\tilde{I} \subseteq I$ auf $(\Omega, \mathcal{F}, \mathbf{P})$. Weiter sei $H = \overline{L^2(\Omega, \mathcal{F}, \mathbf{P})}$, $K = L^2(\Omega, \mathcal{G}, \mathbf{P})$ mit $\mathcal{G} = \sigma(X_t, t \in \tilde{I})$ und $\tilde{K} = \text{Spann}\{X_t, t \in \tilde{I}; 1\}$. \tilde{K} ist also der Abschluss (in L^2) der endlichen Linearkombination der $X_t, t \in \tilde{I}$ und Konstanten. \tilde{K} ist ein abgeschlossener Teilraum von K (und H). Im Fall $|\tilde{I}| < \infty$ ist \tilde{K} immer endlich dimensional (Dimension $\leq |\tilde{I}| + 1$), während K auch in diesem Fall meist unendlich-dimensional ist. Wir interpretieren \tilde{I} als die Menge der Indizes (z.B. Zeitpunkte), für die Beobachtungen vorliegen. Will man für $t_0 \notin \tilde{I}$ X_{t_0} prognostizieren, so ist es sinnvoll, die orthogonale Projektion von X_{t_0} auf K zu bilden. Da diese \mathcal{G} -messbar ist, lässt sie sich nach dem Faktorisierungssatz als messbare Funktion der $X_t, t \in \tilde{I}$, also der Beobachtungsdaten schreiben, die man dann nur noch in die Funktion einzusetzen braucht. Da die Berechnung dieser Funktion oft große Schwierigkeiten bereitet, ist man häufig bereit, eine schlechtere, aber besser berechenbare Vorhersagemethode zu wählen. Als solche bietet sich die orthogonale Projektion von X_{t_0} auf den kleineren Raum \tilde{K} an. Ist $|\tilde{I}| < \infty$, so ist dies sehr einfach.

Ein starkes Argument für diese Methode ist, dass im Fall von Gaußprozessen diese sogar optimal ist, d.h. die bedingte Erwartung ist (f.s.) eine Linearkombination der $X_t, t \in \tilde{I}$ und der Konstanten 1, oder jedenfalls (im Fall $|\tilde{I}| = \infty$) ein Grenzwert von solchen. Wir zeigen zuerst letzteres und dann die Berechnung der orthogonalen Projektion auf \tilde{K} im allgemeinen (nicht notwendig Gaußschen) Fall, wenn $|\tilde{I}| < \infty$.

Satz 8.26. *Ist $X_t, t \in I$ ein Gaußprozess, $\tilde{I} \subseteq I, t_0 \in I \setminus \tilde{I}$, so gilt mit obigen Bezeichnungen*

$$\mathbf{E}(X_{t_0} | \mathcal{G}) \in \tilde{K},$$

d.h. $\mathbf{E}(X_{t_0} | \mathcal{G})$ ist die orthogonale Projektion von X_{t_0} auf \tilde{K} .

Beweis. Wir zeigen die Behauptung nur im Fall $|\tilde{I}| = m < \infty$. Sei $\tilde{I} = \{t_1, \dots, t_m\}$ und V die orthogonale Projektion von X_{t_0} auf \tilde{K} . Diese hat die Form $V = \sum_{i=1}^m a_i X_{t_i} + a$, und es gilt

$$X_{t_0} - V \perp \{1, X_{t_1}, \dots, X_{t_m}\} \quad (\perp = \text{orthogonal})$$

und – da $X_{t_0}, \dots, X_{t_m}, V, 1$ ein Gaußsystem ist – folgt aus Satz 8.25

$$X_{t_0} - V \perp\!\!\!\perp \sigma(\{1, X_{t_1}, \dots, X_{t_m}\}) = \mathcal{G}.$$

Hieraus folgt aber

$$X_{t_0} - V \perp L^2(\Omega, \mathcal{G}, \mathbf{P}) = K.$$

Wegen $V \in K$ ist daher V die orthogonale Projektion von X_{t_0} auf K , d.h. $\mathbf{E}(X_{t_0}|\mathcal{G}) = V$ f.s. \square

Satz 8.27. Sei $(X_t)_{t \in I}$ ein stochastischer Prozess, $m \in \mathbf{N}$, $\tilde{I} = \{t_1, \dots, t_m\} \subseteq I$, $t_0 \in I \setminus \tilde{I}$ und $\mathbf{E}X_t^2 < \infty$, $t \in \tilde{I} \cup \{t_0\}$. Sei

$$\begin{aligned}\mu_t &:= \mathbf{E}X_t \\ R_{t,u} &:= \mathbf{E}(X_t X_u) \quad t, u \in \tilde{I} \cup \{t_0\}.\end{aligned}$$

Sei V die orthogonale Projektion von X_{t_0} auf $\text{Spann}\{X_t, t \in \tilde{I}; 1\}$. Dann gilt

$$V = \sum_{i=1}^m a_i X_{t_i} + a,$$

wobei a, a_1, \dots, a_m irgendeine Lösung des linearen Gleichungssystems

$$\begin{cases} \sum_{j=1}^m a_j R_{t_i, t_j} + a \mu_{t_i} = R_{t_0, t_i} & i = 1, \dots, m \\ \sum_{j=1}^m a_j \mu_{t_j} + a = \mu_{t_0} \end{cases} \quad (1)$$

ist.

Beweis. V ist dadurch charakterisiert, dass V eine Linearkombination von X_{t_i} , $i \in \{1, \dots, m\}$ und 1 ist und gilt

$$\begin{aligned}\mathbf{E}((X_{t_0} - V)X_{t_i}) &= 0 \quad i = 1, \dots, m \\ \mathbf{E}(X_{t_0} - V) &= 0.\end{aligned} \quad (2)$$

Setzt man $V = \sum_{j=1}^m a_j X_{t_j} + a$ in (2) ein und nutzt die Linearität des Erwartungswertes aus, so ist das entstehende lineare Gleichungssystem identisch mit (1). \square

Bemerkung 8.28. Der Satz zeigt insbesondere, dass man zur Berechnung der optimalen linearen Vorhersage V für X_{t_0} die gemeinsame Verteilung von X_{t_i} , $i \in \{0, \dots, m\}$ gar nicht benötigt, sondern nur die Erwartungswerte und Kovarianzen. Auch der Vorhersagefehler $\mathbf{E}((X_{t_0} - V)^2)$ hängt nur von diesen Größen ab.

Bemerkung 8.29. Statt *linearer* Vorhersagen kann man auch quadratische, kubische usw. Vorhersagen betrachten. Z.B. lässt sich die beste Approximation von X_{t_0} durch ZG der Form $a + \sum_{i=1}^m a_i X_{t_i} + \sum_{i,j=1}^m a_{ij} X_{t_i} X_{t_j}$ analog zu Satz 8.27 ebenfalls durch Lösen eines *linearen* Gleichungssystems für die Größen a, a_i, a_{ij} berechnen. Die Zahl der Unbekannten ist dabei aber deutlich größer, und man benötigt (gemischte) Momente der X_{t_i} bis zur vierten Ordnung.

Zum Abschluss noch eine Übungsaufgabe.

Aufgabe 8.30. Man beweise oder widerlege folgende Aussage: Sind X und Y integrierbare Zufallsvariable auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$, so dass auch XY integrierbar ist und ist \mathcal{G} eine Unter- σ -Algebra von \mathcal{F} , so ist $\mathbf{E}(X|\mathcal{G})\mathbf{E}(Y|\mathcal{G})$ integrierbar.