Vorlesung über Statistik für Zeitreihen
Wintersemester 2006/2007

Wolfgang Wefelmeyer

Version vom 14. Juni 2014

Inhaltsverzeichnis

1 Einleitung 2
2 Kontiguität 3
3 LAN und Faltungssatz 8
4 LAN für unabhängige Beobachtungen 10
5 Effiziente Schätzer für parametrische Familien 12
6 LAN und Faltungssatz für Funktionale 13
7 Effizienz empirischer Schätzer 19
8 LAN für Markov-Ketten 21
9 Gleichmäßige Ergodizität und Martingal-Approximation 24
10 Effizienz empirischer Schätzer für Markov-Ketten 28
11 ARMA-Prozesse 35
12 Hilberträume und Orthonormalsysteme 41
13 Kovarianz-Funktion und Ordnung linearer Zeitreihen 45
14 Spektralverteilung stationärer Prozesse 47
15 Spektraldarstellung stationärer Prozesse 51
16 Vorhersage stationärer Prozesse 55
17 Schätzen des Mittelwerts und der Autokovarianz-Funktion 56
18 Schätzer für die Parameter von AR(p)-Prozessen 57
19 Schätzen der Spektraldichte stationärer Prozesse 58
20 Effiziente Schätzer im AR(1)-Modell 62

1 Einleitung

Die Kapitel 2–7 haben mit Zeitreihen noch nichts zu tun. Kapitel 2 und 3 geben Resultate zur Kontiguität und (lokalen) asymptotischen Normalität (LAN) für allgemeine Folgen von Modellen, die einen eindimensionalen Parameter haben. Insbesondere werden der Faltungssatz von Hájek und Le Cam und eine Charakterisierung asymptotisch effizienter Schätzer bewiesen. In den Kapiteln 4 und 5 wird gezeigt, daß für parametrische Modelle mit unabhängigen und identisch verteilten Beobachtungen die Eigenschaft LAN aus der Hellinger-Differenzierbarkeit des Modells folgt; es ergibt sich sofort eine Version des Faltungssatzes. Kapitel 6 verallgemeinert die Resultate aus Kapitel 3 auf Modelle mit unendlichdimensionalem Parameterraum und das Schätzen differenzierbarer Funktionale des Parameters. Kapitel 7 gibt eine einfache Anwendung auf nichtparametrische Modelle mit unabhängigen und identisch verteilten Beobachtungen. Kapitel 8–10 übertragen Kapitel 4–7 auf Markov-Ketten. Die Kapitel 11–15 behandeln die klassischen wahrscheinlichkeitstheoretischen Resultate zu stationären Zeitreihen. Die Kapitel 16–19 behandeln die klassischen statistischen Resultate zu stationären Zeitreihen: Vorhersage, Schätzen von Mittelwert und Autokovarianz-Funktion, Schätzen der Parameter von AR(p)-Prozessen, und schließlich das Periodogramm und geglättete Periodogramme als Schätzer der Spektraldichte. Die Beweise der Sätze finden sich z.B. in Brockwell und Davis (1991). In Kapitel 20 führen wir effiziente Schätzer für verschiedene Funktionale von AR(1)-Prozessen ein. Die Beweise zur Ergodizität finden sich z.B. in Meyn und Tweedie (1993). Die Beweise zur Hellinger-Differenzierbarkeit und zum effizienten Schätzen des Autoregressions-Parameters finden sich in Koul und Schick (1997).

2 Kontiguität

Wir wissen schon, daß wir glatte Funktionale einer unbekannten Verteilung mit der Rate $n^{-1/2}$ schätzen können, wenn wir n unabhängige und identisch verteilte Beobachtungen haben. Das gilt für parametrische Modelle (Maximum-Likelihood-Schätzer) und nichtparametrische Modelle (M-Schätzer, empirische Schätzer). Für parametrische Modelle bedeutet das insbesondere, daß wir zwei Parameter mit Sicherheit asymptotisch unterscheiden können, wenn ihr Abstand langsamer als $n^{-1/2}$ schrumpft. Interessant sind also die Abstände der Ordnung $n^{-1/2}$. Wir zeigen im folgenden, daß $n^{-1/2}$ tatsächlich die beste Konvergenzrate für Schätzer glatter Funktionale ist. Dazu brauchen wir nur eine wie $n^{-1/2}$ schrumpfende Umgebung eines Parameters zu betrachten. Das sieht man auch heuristisch, wenn man den Likelihoodquotienten ansieht. Ist f_ϑ die Dichte eines Wahrscheinlichkeitsmaßes P_ϑ, ist Θ eine offene Teilmenge von $\mathbb{R}$, und sind n unabhängige Beobachtungen $X_1, \dots, X_n$ aus einem der P_ϑ gegeben, so ist der Log-Likelihoodquotient

$$\begin{aligned}
\log \frac{dP^n_{\vartheta+n^{-1/2}t}}{dP^n_\vartheta}(X_1, \dots, X_n) &= \sum_{i=1}^n (\ell_{\vartheta+n^{-1/2}t}(X_i) - \ell_\vartheta(X_i)) \\
&\approx tn^{-1/2} \sum_{i=1}^n \dot\ell_\vartheta(X_i) + \frac{1}{2} t^2 \frac{1}{n} \sum_{i=1}^n \ddot\ell(X_i) \\
&\approx tn^{-1/2} \sum_{i=1}^n \dot\ell_\vartheta(X_i) - \frac{1}{2} t^2 I_\vartheta
\end{aligned}$$

mit $\ell_\vartheta = \log f_\vartheta$; $\dot\ell_\vartheta$, $\ddot\ell_\vartheta$ den Ableitungen nach ϑ; und $I_\vartheta = E\dot\ell^2_\vartheta$ der Fisher-Information. Also hat $\log dP^n_\tau / dP^n_\vartheta$ genau dann eine nicht ausgeartete Verteilung, wenn $\tau - \vartheta$ von der Ordnung $n^{-1/2}$ ist. Dann sind P^n_τ und P^n_ϑ "benachbart". Für $\tau = \vartheta + n^{-1/2}t$ ist die Grenzverteilung eine Normalverteilung mit Mittelwert $-\frac{1}{2}t^2 I_\vartheta$ und Varianz $t^2 I_\vartheta$. Wir sagen dann, wir haben "lokale asymptotische Normalität" (LAN). Daraus werden wir im folgenden schließen, daß $n^{-1/2}$ die optimale Konvergenzrate für Schätzer glatter Funktionale von ϑ ist. Insbesondere werden wir zeigen, daß es keinen "regulären" Schätzer für ϑ gibt, der eine bessere asymptotische Varianz als I_ϑ^{-1} hat. Das ist die asymptotische Varianz des Maximum-Likelihood-Schätzers. Es ist auch die nichtasymptotische Schranke, die wir für erwartungstreue Schätzer aus der Cramér–Rao-Ungleichung kennen.

Für eine Normalverteilung $N_{tI,I}$ mit Lageparameter tI gilt

$$\log \frac{dN_{tI,I}}{dN_{0,I}}(X) = -\frac{(X-tI)^2}{2I} + \frac{X^2}{2I} = tX - \frac{1}{2}t^2 I.$$

Asymptotisch ist der obige Log-Likelihoodquotient von dieser Form. Also reduziert sich das ursprüngliche Schätzproblem asymptotisch auf diesen Fall.

Gegeben seien Wahrscheinlichkeitsräume $(\Omega_n, \mathcal{F}_n, P_n)$ und reelle Zufallsvariablen V_n darauf. Für eine Menge $A \subset \mathbb{R}$ sei A° das Innere und A^- der Abschluß von A.

Lemma 1 *Sei $f : \mathbb{R} \to \mathbb{R}$ meßbar. Bezeichne D_f die Menge der Unstetigkeitsstellen von f. Gilt $V_n \Rightarrow V$ und $P^V D_f = 0$, so gilt $f(V_n) \Rightarrow f(V)$.*

Beweis. Sei F abgeschlossen. Wegen $V_n \Rightarrow V$ gilt

$$\limsup P_n^{V_n}(f^{-1}F) \leq \limsup P_n^{V_n}(f^{-1}F)^- \leq P^V(f^{-1}F)^-.$$

Es gilt $(f^{-1}F)^- \subset D_f \cup (f^{-1}F)$, also $P^V(f^{-1}F)^- = P^V(f^{-1}F)$ nach Voraussetzung. Die Behauptung folgt jetzt aus dem Portmanteau-Lemma.

Lemma 2 *Sei $f : \mathbb{R} \to \mathbb{R}$ meßbar und beschränkt mit $PD_f = 0$. Gilt $V_n \Rightarrow V$, so gilt $\int f\, dP_n^{V_n} \to \int f\, dP^V$.*

Beweis. Sei $|f| < c$. Lemma 1 und schwache Konvergenz für $g(x) = (-c) \vee x \wedge c$ anwenden.

Lemma 3 *Gilt $V_n \Rightarrow V$, so auch $\liminf P_n|V_n| \geq P|V|$.*

Beweis. Sei $f(x) = |x|\mathbf{1}(|x| \leq c)$. Gilt $P(|V| = c) = 0$, so gilt nach Lemma 2:

$$\int_{|V|\leq c} |V|\, dP = \lim_n \int_{|V_n|\leq c} |V_n|\, dP_n \leq \liminf P_n|V_n|.$$

Jetzt c wachsen lassen. Die Menge der c, für die nicht $P(|V| = c) = 0$ gilt, ist höchstens abzählbar, das Komplement also dicht in $\mathbb{R}$.

Definition. Die Folge V_n heißt *gleichmäßig integrierbar* unter P_n, wenn

$$\sup_n P_n(|V_n|\mathbf{1}(|V_n| > c)) \to 0, \quad c \to \infty.$$

Dann ist die Folge $P_n|V_n|$ beschränkt, denn

$$P_n|V_n| \leq c + P_n(|V_n|\mathbf{1}(|V_n| > c)).$$

Lemma 4 *Die Folge V_n ist gleichmäßig integrierbar genau dann, wenn $P_n|V_n|$ beschränkt ist und*

$$P_n|V_n|A_n \to 0, \quad \textit{wenn} \quad P_n A_n \to 0.$$

Beweis. *Hinreichend.* Es gilt

$$\begin{aligned} P_n|V_n|A_n &= P_n|V_n|A_n\{|V_n| > c\} + P_n|V_n|A_n\{|V_n| \le c\} \\ &\le P_n|V_n|\{|V_n| > c\} + cP_nA_n. \end{aligned}$$

Jetzt $c = c_n$ langsam gegen Unendlich wachsen lassen.

Notwendig. Nach Voraussetzung existiert ein $M > 0$ mit $P_n|V_n| \le M$. Mit der Chebyshev-Ungleichung gilt also

$$P_n(|V_n| > c) \le M/c.$$

Jetzt die Voraussetzung auf $A_n = \{|V_n| > c_n\}$ mit $c_n \to \infty$ anwenden.

Lemma 5 *Gelte $V_n \Rightarrow V$. Dann ist V_n ist gleichmäßig integrierbar genau dann, wenn $P_n|V_n|$ beschränkt ist mit*

$$P_n|V_n| \to P|V| \quad \textit{und} \quad P|V| < \infty.$$

Beweis. *Hinreichend.* Ist V_n gleichmäßig integrierbar, so ist die Folge $P_n|V_n|$ beschränkt, also $P|V| < \infty$ nach Lemma 3. Sei $f(x) = x\mathbf{1}(|x| \le c)$. Wenn $P(|V| = c) = 0$, so folgt wegen $V_n \Rightarrow V$ nach Lemma 2:

$$P_n f(|V_n|) \to Pf(|V|).$$

Schreibe

$$P_n|V_n| - P_n f(|V_n|) = \int_{|V_n|>c} |V_n|\, dP_n, \quad P|V| - Pf(|V|) = \int_{|V|>c} |V|\, dP.$$

Wir erhalten

$$\limsup |P_n|V_n| - P|V|| \le \sup \int_{|V_n|>c} |V_n|\, dP_n + \int_{|V|>c} |V|\, dP.$$

Also folgt $P_n|V_n| \to P|V|$ aus der gleichmäßigen Integrierbarkeit.

Notwendig. Gilt $P_n|V_n| \to P|V| < \infty$ und $P(|V| = c) = 0$, so gilt $P_n f(|V_n|) \to Pf(|V|)$ nach Lemma 2, also folgt aus Obigem

$$\int_{|V_n|>c} |V_n|\, dP_n \to \int_{|V|>c} |V|\, dP.$$

Jetzt c wachsen lassen.

Gegeben seien Wahrscheinlichkeitsmaße P_n, Q_n auf $(\Omega_n, \mathcal{F}_n)$. Wir definieren mit $p_n = dP_n/d(P_n + Q_n)$ und $q_n = dQ_n/d(P_n + Q_n)$ den Dichtequotienten

$$L_n = \frac{dQ_n}{dP_n} = \frac{q_n}{p_n}\mathbf{1}(p_n > 0) + \infty\mathbf{1}(p_n = 0, q_n = 0).$$

Es gilt $P_n L_n = Q_n(L_n < \infty)$ und $P_n(L_n = \infty) = 0$.

Definition. Q_n ist zu P_n *benachbart* ($Q_n \lhd P_n$), wenn für $A_n \in \mathcal{F}_n$ gilt:

$$P_n A_n \to 0 \quad \text{impliziert} \quad Q_n A_n \to 0.$$

Lemma 6 *Die folgenden Bedingungen sind äquivalent:*

(1) $Q_n \lhd P_n$.

(2) L_n *ist straff unter* P_n, *und wenn* $L_n \Rightarrow L$ *für eine Teilfolge, dann gilt* $PL = 1$.

(3) L_n *ist gleichmäßig integrierbar unter* P_n, *und* $Q_n(L_n = \infty) \to 0$.

(4) L_n *ist straff unter* Q_n.

Beweis. *(3) und (1) sind äquivalent:* Zur Vorbereitung schreiben wir

$$\begin{aligned} Q_n A_n &= Q_n\mathbf{1}(L_n < \infty)A_n + Q_n\mathbf{1}(L_n = \infty)A_n \\ &= P_n L_n A_n + Q_n\mathbf{1}(L_n = \infty)A_n. \end{aligned} \tag{2.1}$$

Hier haben wir $\mathbf{1}(L_n < \infty)$ weglassen können, weil $P_n(L_n = \infty) = 0$, also $P_n(L_n < \infty) = 1$.

(3) impliziert (1): Seien $A_n \in \mathcal{F}_n$ mit $P_n A_n \to 0$. Es gilt $P_n L_n A_n \to 0$ nach Lemma 4 "$\Rightarrow$" für $V_n = L_n$. Nach Voraussetzung gilt $Q_n(L_n = \infty) \to 0$, also $Q_n\mathbf{1}(L_n = \infty)A_n \to 0$, also $Q_n A_n \to 0$ nach (2.1).

(1) impliziert (3): Seien $A_n \in \mathcal{F}_n$ mit $P_n A_n \to 0$. Nach Voraussetzung gilt $Q_n A_n \to 0$, also $Q_n\mathbf{1}(L_n = \infty)A_n \to 0$. Nach (2.1) gilt $P_n L_n A_n \to 0$. Außerdem gilt $P_n L_n = Q_n(L_n < \infty) \leq 1$. Nach Lemma 4 "$\Leftarrow$" für $V_n = L_n$ ist L_n gleichmäßig integrierbar unter P_n. Außerdem gilt $P_n(L_n = \infty) = 0$, also $Q_n(L_n = \infty) \to 0$.

(2) impliziert (3): Es genügt zu zeigen: Jede Teilfolge hat eine Teilfolge mit (3). Nach dem Satz von Prohorov besitzt jede Teilfolge eine konvergente Teilfolge mit $L_n \Rightarrow L$. Mit Lemma 3 gilt $\liminf_n P_n L_n \geq PL = 1$. Anderseits ist $P_n L_n = Q_n(L_n < \infty) \leq 1$. Also gilt $P_n L_n \to PL$ und $Q_n(L_n < \infty) \to 1$. Also ist L_n gleichmäßig integrierbar unter P_n nach Lemma 5 "$\Leftarrow$", und $Q_n(L_n = \infty) = 1 - Q_n(L_n < \infty) \to 0$.

(3) impliziert (2): Ist L_n gleichmäßig integrierbar unter P_n, so auch straff, denn für $c > 0$ gilt

$$P_n(L_n > c) \leq \frac{1}{c} P_n L_n \mathbf{1}(L_n > c).$$

Gelte nun $L_n \Rightarrow L$ für eine Teilfolge. Dann gilt $P_n L_n \to PL$ nach Lemma 5 "$\Rightarrow$". Außerdem gilt

$$P_n L_n = Q_n(L_n < \infty) = 1 - Q_n(L_n = \infty) \to 1,$$

also $PL = 1$.

(3) impliziert (4): Wegen

$$Q_n(L_n > c) = P_n L_n \mathbf{1}(L_n > c) + Q_n(L_n = \infty).$$

(4) impliziert (1): Wegen

$$\begin{aligned} Q_n A_n &\leq Q_n(L_n \leq c) A_n + Q_n(L_n > c) \\ &= P_n L_n \mathbf{1}(L_n \leq c) A_n + Q_n(L_n > c) \\ &\leq c P_n A_n + Q_n(L_n > c). \end{aligned}$$

Lemma 7 *(Drittes Lemma von Le Cam) Gilt $Q_n \lhd P_n$ und $(L_n, V_n) \Rightarrow (L, V)$ unter P_n, dann auch $(L_n, V_n) \Rightarrow (L, V)$ unter Q_n, und $dQ = LdP$.*

Beweis. Sei $f : [0, \infty] \times \mathbb{R} \to \mathbb{R}$ stetig und beschränkt. Schreibe

$$Q_n f(L_n, V_n) = P_n L_n f(L_n, V_n) + Q_n \mathbf{1}(L_n = \infty) f(L_n, V_n).$$

Der zweite Term geht gegen 0, weil f beschränkt ist und $Q_n(L_n = \infty) \to 0$ nach Lemma 6 "(1) $\Rightarrow$ (3)". Schreibe den ersten Term als

$$P_n(L_n \wedge c) f(L_n, V_n) + P_n(L_n - c)^+ f(L_n, V_n).$$

Der linke Summand konvergiert gegen $P(L \wedge c) f(L, V)$; für große c geht das gegen $PLf(L, V)$. Der rechte Summand ist klein für großes c nach Lemma 6 "(1) $\Rightarrow$ (3)" mit $P_n(L_n - c)^+ \leq P_n L_n (L_n \geq c)$.

Für den Beweis des nächsten Lemmas erinnern wir daran, daß die charakteristische Funktion einer mehrdimensionalen Normalverteilung $N_{\mu,\Sigma}$ die Form $\varphi(t) = \exp(i\mu^\top t - \frac{1}{2} t^\top \Sigma t)$ hat.

Lemma 8 *Gelte $Q_n \lhd P_n$. Setze $\Lambda_n = \log L_n$. Seien V_n Zufallsvariablen mit $(\Lambda_n, V_n) \Rightarrow (\Lambda, V)$ unter P_n, wo (Λ, V) einer zweidimensionalen Normalverteilung folgt mit Mittelwertvektor und Kovarianzmatrix*

$$\begin{pmatrix} -\frac{1}{2}a \\ b \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} a & c \\ c & d \end{pmatrix}.$$

Dann gilt $(\Lambda_n, V_n) \Rightarrow N$ unter Q_n, wo N eine zweidimensionale Normalverteilung mit derselben Kovarianzmatrix ist wie (Λ, V), und mit Mittelwertvektor

$$\begin{pmatrix} \frac{1}{2}a \\ b+c \end{pmatrix}.$$

Beweis. Unter P_n hat die charakteristische Funktion der Grenzverteilung die Form $\varphi(s,t) = P\exp(is\Lambda + itV)$. Nach Lemma 7 hat die charakteristische Funktion der Grenzverteilung unter Q_n die Form

$$PL\exp(is\Lambda + itV) = P\exp((is+1)\Lambda + itV) = \varphi(s-i,t).$$

Es gilt

$$\varphi(s,t) = \exp\Big(-\frac{1}{2}isa + itb - \frac{1}{2}(s^2a + t^2d + 2stc)\Big).$$

Also ist die charakteristische Funktion unter Q_n von der Form

$$\begin{aligned} \varphi(s-i,t) = \exp\Big(&-\frac{1}{2}isa - \frac{1}{2}a + itb \\ &- \frac{1}{2}(s^2a - 2isa - a + t^2d + 2stc - 2itc)\Big) \\ = \exp\Big(&\frac{1}{2}isa + it(b+c) - \frac{1}{2}(s^2a + t^2d + 2stc)\Big). \end{aligned}$$

3 LAN und Faltungssatz

Sei f stetig und $f_n \to f$ punktweise. Im "regulären" Fall, für "gutartige" f_n, wird die Konvergenz sogar *stetig* sein:

$$f_n(x_n) \to f(x), \quad \text{wenn} \quad x_n \to x.$$

Für die Verteilungen der Schätzerfolgen werden wir nur eine abgeschwächte Version der stetigen Konvergenz annehmen:

$$n^{1/2}(\hat{\vartheta} - (\vartheta + n^{-1/2}t)) \Rightarrow V \quad \text{unter } P^n_{\vartheta + n^{-1/2}t}, \quad t \in \mathbb{R}.$$

Solche Schätzer werden wir *regulär* nennen. Für sie werden wir eine (scharfe) untere Schranke für die asymptotische Varianz angeben. Für Anwendungen auf nichtparametrische Modelle ist es von Vorteil, alle folgenden Aussagen zunächst für *lokale* Modelle zu formulieren, also für P_{nt} statt $P^n_{\vartheta+n^{-1/2}t}$, und für V_n statt $n^{1/2}(\hat\vartheta - \vartheta)$.

Setze $L_{nt} = dP_{nt}/dP_n$ und $\Lambda_{nt} = \log L_{nt}$.

Definition. Die Familie P_{nt}, $t \in \mathbb{R}$, heißt *asymptotisch normal*, wenn

$$\Lambda_{nt} = tS_n - \frac{1}{2}t^2 J + o_{P_n}(1), \quad t \in \mathbb{R},$$
$$S_n \Rightarrow J^{1/2} N \quad \text{unter } P_n,$$

wobei N eine standardnormalverteilte Zufallsvariable und J eine positive Zahl ist.

Lemma 9 *Ist P_{nt}, $t \in \mathbb{R}$, asymptotisch normal, so sind P_{ns} und P_{nt} für alle s und t benachbart.*

Beweis. Weil $\lhd$ transitiv ist, genügt es, $t = 0$ zu betrachten. Es gilt $P_{ns} \lhd P_{n0}$ nach Lemma 6 "(4) $\Rightarrow$ (1)". Es gilt $P_{n0} \lhd P_{ns}$ nach Lemma 6 "(2) $\Rightarrow$ (1)". wenn dort die Rollen von P_n und Q_n vertauscht werden.

Definition. V_n heißt *regulär* mit *Limes* V, wenn

$$V_n - t \Rightarrow V \quad \text{unter } P_{nt}, \quad t \in \mathbb{R}.$$

Satz 1 *(Faltungssatz von Hájek und Le Cam) Ist P_{nt} asymptotisch normal und V_n regulär, so gilt*

$$(J^{-1}S_n, V_n - J^{-1}S_n) \Rightarrow (J^{-1/2}N, U) \quad \textit{unter } P_n$$

für eine Zufallsvariable U, die unabhängig von N ist.

Beweis. Da (S_n, V_n) straff unter P_n ist, gilt $(S_n, V_n) \Rightarrow (J^{1/2}N, V)$ unter P_n für eine Teilfolge. Nach Lemma 9 sind P_{ns} und P_{nt} benachbart. Nach Lemma 7 gilt für jedes stetige und beschränkte f:

$$P_t f(V) = P_s \exp(\Lambda_t - \Lambda_s) f(V).$$

Insbesondere gilt

$$P_t \exp(iu(V-t)) = P_s \exp\Big(iuV - iut + tJ^{1/2}N - \frac{1}{2}t^2 J - sJ^{1/2}N + \frac{1}{2}s^2 J\Big).$$

Da V_n regulär ist, hängt die linke Seite nicht von t ab. Die rechte Seite ist analytisch in t. Ersetze t durch $s - iJ^{-1}u$:

$$P \exp(iuV) = \exp\Big(-\frac{1}{2}u^2 J^{-1}\Big) P_s \exp(iu(V - J^{-1/2}N)).$$

Also hängt die Verteilung von $V - J^{-1/2}N$ unter P_s nicht von s ab. Nach dem Lemma von Basu ist also $V - J^{-1/2}N$ unabhängig von N. Da S_n und V_n in Verteilung konvergieren, hängt die Grenzverteilung von $V - J^{-1/2}N$ nicht von der Teilfolge ab.

Insbesondere ist $V_n = J^{-1}S_n + V_n - J^{-1}S_n$ asymptotisch wie die Faltung $J^{-1/2}N + U$ verteilt. Daher der Name "Faltungssatz". Ist M unabhängig von N, so gilt

$$P(-a < N + M < a) \leq P(-a < N < a), \quad a > 0.$$

Also ist V_n bestenfalls asymptotisch wie $J^{-1/2}N$ in symmetrischen Intervallen um 0 konzentriert. V_n ist asymptotisch optimal genau dann, wenn U nach δ_0 verteilt ist. Genau dann gilt $V_n = J^{-1}S_n + o_{P_n}(1)$.

4 LAN für unabhängige Beobachtungen

Seien $X_1, \dots, X_n$ unabhängige Beobachtungen mit Verteilung P_ϑ aus einer Familie von äquivalenten Wahrscheinlichkeitsmaßen P_ϑ, $\vartheta \in \Theta$, mit $\Theta \subset \mathbb{R}$ offen. Sei ϑ fest. Setze $L_{\tau\vartheta} = dP_\tau / dP_\vartheta$.

Definition. P_τ heißt *Hellinger-differenzierbar* in $\tau = \vartheta$ mit *Ableitung* $\dot{\ell}_\vartheta \in L_{2,0}(P_\vartheta)$, wenn

$$\Big\| L_{\tau\vartheta}^{1/2} - 1 - \frac{1}{2}(\tau - \vartheta)\dot{\ell}_\vartheta \Big\|_2 = o(\tau - \vartheta).$$

Wir benötigen diese Eigenschaft nur für Folgen $\vartheta_{nt} = \vartheta + n^{-1/2}t$ statt für alle τ. Das werden wir in Kapitel 7 brauchen.

Wegen $P_\vartheta L_{\tau\vartheta} = P_\tau \Omega = 1$ und $x - 1 = 2(x^{1/2} - 1) + (x^{1/2} - 1)^2$ gilt

$$P_\vartheta(L_{\tau\vartheta}^{1/2} - 1) = -\frac{1}{2} P_\vartheta (L_{\tau\vartheta}^{1/2} - 1)^2. \tag{4.1}$$

Bezeichne $I_\vartheta = P_\vartheta \dot{\ell}_\vartheta^2$ die *Fisher-Information.* Setze

$$L_{nt} = \frac{dP_{\vartheta_{nt}}^n}{dP_\vartheta^n}, \quad \Lambda_{nt} = \log L_{nt}.$$

Satz 2 *Ist P_τ Hellinger-differenzierbar in $\tau = \vartheta$ mit* Ableitung $\dot\ell_\vartheta$*, so gilt*

$$\Lambda_{nt} = tn^{-1/2}\sum_{i=1}^n \dot\ell_\vartheta(X_i) - \frac{1}{2}t^2 I_\vartheta + o_{P_\vartheta}(1),$$
$$n^{-1/2}\sum_{i=1}^n \dot\ell_\vartheta(X_i) \Rightarrow I_\vartheta^{1/2} N.$$

Wir sagen dann: P_τ ist *lokal asymptotisch normal* (LAN) in ϑ.

Beweis. Schreibe $\log x = 2\log(1 + x^{1/2} - 1)$. Wir verwenden die Taylor-Entwicklung $\log(1+x) = x - \frac{1}{2}x^2 + r(x)$. Für $Z = dP_{\vartheta+n^{-1/2}t}/dP_\vartheta$ ergibt sich mit (4.1):

$$\begin{aligned}\log Z &= 2\log(1+Z^{1/2}-1) = 2(Z^{1/2}-1) - (Z^{1/2}-1)^2 + 2r(Z^{1/2}-1)\\ &= 2\big(Z^{1/2} - 1 - P_\vartheta(Z^{1/2}-1)\big) - 2P_\vartheta(Z^{1/2}-1)^2\\ &\quad + 2r(Z^{1/2}-1) - \big((Z^{1/2}-1)^2 - P_\vartheta(Z^{1/2}-1)^2\big).\end{aligned}$$

Mit $Z_i = Z(X_i)$ also

$$\begin{aligned}\Lambda_{nt} &= \sum_{i=1}^n \log Z_i = 2\sum_{i=1}^n \big(Z_i^{1/2} - 1 - P_\vartheta(Z^{1/2}-1)\big) - 2nP_\vartheta(Z^{1/2}-1)^2\\ &\quad + 2\sum_{i=1}^n r(Z_i^{1/2}-1) - \sum_{i=1}^n \big((Z_i^{1/2}-1)^2 - P_\vartheta(Z^{1/2}-1)^2\big).\end{aligned}$$

Der erste Term ist

$$tn^{-1/2}\sum_{i=1}^n \dot\ell_\vartheta(X_i) + 2\sum_{i=1}^n \Big(Z_i^{1/2} - 1 - P_\vartheta(Z^{1/2}-1) - \frac{1}{2}tn^{-1/2}\dot\ell_\vartheta(X_i)\Big),$$

also gleich $tn^{-1/2}\sum_{i=1}^n \dot\ell_\vartheta(X_i)$ bis auf $o_{P_\vartheta}(1)$. Der zweite Term konvergiert in Wahrscheinlichkeit gegen $-\frac{1}{2}P_\vartheta\dot\ell_\vartheta^2 = -\frac{1}{2}t^2 I_\vartheta$. Für den dritten Term haben wir

$$\begin{aligned}nP_\vartheta(|Z^{1/2}-1| > \varepsilon) &\le \frac{n}{\varepsilon^2}P_\vartheta(Z^{1/2}-1)^2\mathbf{1}(|Z^{1/2}-1| > \varepsilon)\\ &= \frac{t^2}{4\varepsilon^2}P_\vartheta\dot\ell_\vartheta^2\mathbf{1}(|\dot\ell_\vartheta| > 2t^{-1}n^{1/2}\varepsilon) + o(1) \to 0.\end{aligned}$$

Das gilt auch für eine Folge $\varepsilon_n \to 0$. Mit $|r(x)| \leq 2|x|^3$ für $|x| \leq \frac{1}{2}$ und dem Gesetz der großen Zahl gilt

$$\sum_{i=1}^{n} |r(Z_i^{1/2} - 1)| \mathbf{1}(|Z_i^{1/2} - 1| \leq \varepsilon_n) \leq 2\varepsilon_n \sum_{i=1}^{n} (Z_i^{1/2} - 1)^2$$
$$= \frac{t^2}{2n} \varepsilon_n \sum_{i=1}^{n} \dot{\ell}_\vartheta^2(X_i) + o_{P_\vartheta}(1) = o_{P_\vartheta}(1).$$

Der vierte Term ist $\frac{t^2}{4n} \sum_{i=1}^n (\dot{\ell}_\vartheta^2(X_i) - P_\vartheta \dot{\ell}_\vartheta^2)$ bis auf $o_{P_\vartheta}(1)$, also von der Ordnung $o_{P_\vartheta}(1)$ nach dem Gesetz der großen Zahl.

5 Effiziente Schätzer für parametrische Familien

Gegeben sei eine Familie von äquivalenten Wahrscheinlichkeitsmaßen P_ϑ, $\vartheta \in \Theta$, mit $\Theta \subset \mathbb{R}$ offen. Sei ϑ fest. Seien $X_1, \ldots, X_n$ nach P_ϑ verteilt.

Definition. Ein Schätzer $\hat{\vartheta}$ heißt *regulär* in ϑ mit *Limes* V, wenn

$$n^{1/2}(\hat{\vartheta} - (\vartheta + n^{-1/2}t)) \Rightarrow V \quad \text{unter } P_{\vartheta + n^{-1/2}t}, \quad t \in \mathbb{R}.$$

Definition. Ein Schätzer $\hat{\vartheta}$ heißt *asymptotisch linear* in ϑ mit *Einflußfunktion* h, wenn $h \in L_{2,0}(P_\vartheta)$ und

$$n^{1/2}(\hat{\vartheta} - \vartheta) = n^{-1/2} \sum_{i=1}^{n} h(X_i) + o_{P_\vartheta}(1).$$

Ein solcher Schätzer ist asymptotisch normal mit Varianz $P_\vartheta h^2$.

Satz 3 *Sei P_τ Hellinger-differenzierbar in ϑ und $\hat{\vartheta}$ asymptotisch linear in ϑ mit Einflußfunktion h. Dann ist $\hat{\vartheta}$ regulär genau dann, wenn $h - I_\vartheta^{-1}\dot{\ell}_\vartheta \perp \dot{\ell}_\vartheta$ unter P_ϑ.*

Beweis. Es ist $(tn^{-1/2} \sum_{i=1}^n \dot{\ell}_\vartheta(X_i), n^{-1/2} \sum_{i=1}^n h(X_i))$ asymptotisch normal mit Kovarianz $tP_\vartheta \dot{\ell}_\vartheta h$. Nach Satz 2 und Lemma 8 gilt

$$n^{1/2}(\hat{\vartheta} - (\vartheta + n^{-1/2}t)) \Rightarrow (P_\vartheta h^2)^{1/2} N + tP_\vartheta \dot{\ell}_\vartheta h - t \quad \text{unter } P_{\vartheta + n^{-1/2}t}.$$

Es ist also $\hat{\vartheta}$ regulär genau dann, wenn $P_\vartheta \dot{\ell}_\vartheta h = 1$, das heißt

$$P_\vartheta(h - I_\vartheta^{-1}\dot{\ell}_\vartheta)\dot{\ell}_\vartheta = P_\vartheta \dot{\ell}_\vartheta h - 1 = 0.$$

Satz 4 *Sei P_τ Hellinger-differenzierbar in ϑ und $\hat{\vartheta}$ regulär mit Limes V. Dann gilt $V = I_\vartheta^{-1/2} N + U$ in Verteilung für eine Zufallsvariable U, die unabhängig von N ist.*

Ist U verteilt nach δ_0, so ist $\hat{\vartheta}$ asymptotisch linear in ϑ mit Einflußfunktion $I_\vartheta^{-1}\dot{\ell}_\vartheta$.

Beweis. Der erste Teil folgt aus Satz 1 mit $V_n = J^{-1}S_n + V_n - J^{-1}S_n$. Ist U verteilt nach δ_0, so folgt aus Satz 1, daß $V_n - J^{-1}S_n \Rightarrow \delta_0$. Das ist der zweite Teil der Behauptung.

Bemerkung 1 Wenn man nur asymptotisch lineare Schätzer betrachtet, braucht man den Faltungssatz nicht. Ist $\hat{\vartheta}$ asymptotisch linear in ϑ mit Einflußfunktion $h \in L_{2,0}(P_\vartheta)$ und regulär, dann gilt $h - I_\vartheta^{-1}\dot{\ell}_\vartheta \perp \dot{\ell}_\vartheta$ unter P_ϑ nach Satz 3. Weil $\hat{\vartheta}$ asymptotisch linear ist, ist $\hat{\vartheta}$ asymptotisch normal mit Varianz $\|h\|_2^2$. Es gilt mit Pythagoras:

$$\|h\|_2^2 = \|I_\vartheta^{-1}\dot{\ell}_\vartheta + h - I_\vartheta^{-1}\dot{\ell}_\vartheta\|_2^2 = I_\vartheta^{-1} + \|h - I_\vartheta^{-1}\dot{\ell}_\vartheta\|_2^2 \geq I_\vartheta^{-1}.$$

Insbesondere ist $\hat{\vartheta}$ nicht nur in allen symmetrischen, sondern sogar in *allen* 0 enthaltenden Intervallen asymptotisch höchstens wie $I_\vartheta^{-1/2}N$ konzentriert.

6 LAN und Faltungssatz für Funktionale

Die Ergebnisse von Kapitel 3 lassen sich von parametrischen auf nichtparametrische und semiparametrische Modelle und auf das Schätzen differenzierbarer Funktionale übertragen. Um den Begriff der asymptotischen Linearität aus Kapitel 5 ebenfalls zu verallgemeinern, benötigen wir eine Einbettung des Modells in einen größeren Raum.

Sei A ein *Parameterraum.* Für $n \in \mathbb{N}$ sei P_{na}, $a \in A$, eine Familie von Wahrscheinlichkeitsmaßen auf $(\Omega_n, \mathcal{F}_n)$, das *Modell.* Sei $\kappa : A \to \mathbb{R}$ ein Funktional. Wir wollen $\kappa(a)$ schätzen. Sei $a \in A$ fest. Setze $P_n = P_{na}$.

Sei H ein Hilbertraum mit innerem Produkt (h, h') und Norm $\|h\| = (h,h)^{1/2}$. Für $h \in H$ und $n \in \mathbb{N}$ sei $S_n(h)$ eine Zufallsvariable auf $(\Omega_n, \mathcal{F}_n)$, die linear in h ist mit

$$S_n(h) \Rightarrow \|h\|N \quad \text{unter } P_n.$$

Sei K ein linearer Raum, der *lokale Parameterraum.* Für $k \in K$ sei a_{nk} eine Folge in A. Setze $L_{nk} = dP_{na_{nk}}/dP_n$ und $\Lambda_{nk} = \log L_{nk}$.

Definition. Das Modell heißt *lokal asymptotisch normal* in a, wenn eine lineare Abbildung $D : K \to H$ existiert mit

$$\Lambda_{nk} = S_n(Dk) - \frac{1}{2}\|Dk\|^2 + o_{P_n}(1), \quad k \in K.$$

Definition. Das Funktional κ heißt *differenzierbar* in a mit *Gradient* g, wenn $g \in H$ und

$$n^{1/2}(\kappa(a_{nk}) - \kappa(a)) \to (Dk, g), \quad k \in K.$$

Der *kanonische* Gradient g_0 ist die Projektion eines beliebigen Gradienten auf $(DK)^-$.

Beispiel. In Kapitel 4 war $A = \Theta \subset \mathbb{R}$, $P_{na} = P_\vartheta^n$ und $\kappa(\vartheta) = \vartheta$. Setze $K = \mathbb{R}$ und $\vartheta_{nt} = \vartheta + n^{-1/2}t$. Setze $H = L_{2,0}(P_\vartheta)$ mit dem natürlichen inneren Produkt $(h, h') = P_\vartheta hh'$. Setze $S_n(h) = n^{-1/2}\sum_{i=1}^n h(X_i)$ und $Dt = t\dot{\ell}_\vartheta$. Nach Satz 2 gilt

$$\Lambda_{nt} = tn^{-1/2}\sum_{i=1}^n \dot{\ell}_\vartheta(X_i) - \frac{1}{2}t^2 I_\vartheta + o_{P_\vartheta}(1) = S_n(t\dot{\ell}_\vartheta) - \frac{1}{2}(t\dot{\ell}_\vartheta, t\dot{\ell}_\vartheta) + o_{P_\vartheta}(1).$$

Für das Funktional $\kappa(\vartheta) = \vartheta$ gilt

$$n^{1/2}(\kappa(\vartheta_{nt}) - \kappa(\vartheta)) = t = (t\dot{\ell}_\vartheta, I_\vartheta^{-1}\dot{\ell}_\vartheta).$$

Also ist $I_\vartheta^{-1}\dot{\ell}_\vartheta$ der kanonische Gradient von $\kappa(\vartheta) = \vartheta$.

Definition. Ein Schätzer $\hat{\kappa}$ heißt *regulär* für κ in a mit *Limes* V, wenn

$$n^{1/2}(\hat{\kappa} - \kappa(a_{nk})) \Rightarrow V \quad \text{unter } P_{a_{nk}}, \quad k \in K. \tag{6.1}$$

Definition. Ein Schätzer $\hat{\kappa}$ heißt *asymptotisch linear* für κ in a mit *Einflußfunktion* f, wenn $f \in H$ und

$$n^{1/2}(\hat{\kappa} - \kappa(a)) = S_n(f) + o_{P_n}(1).$$

Satz 5 *(Faltungssatz) Ist das Modell lokal asymptotisch normal, das Funktional κ differenzierbar mit kanonischem Gradienten g_0 und der Schätzer $\hat{\kappa}$ regulär in a, so gilt für eine von N unabhängige Zufallsvariable U:*

$$\big(S_n(g_0), n^{1/2}(\hat{\kappa} - \kappa(a)) - S_n(g_0)\big) \Rightarrow (\|g_0\|N, U) \quad \text{unter } P_n.$$

Beweis. Sei $k \in K$ und Dk nicht orthogonal zu g_0. Setze $a_{nt} = a_{n,t(Dk,g_0)^{-1}k}$ und $P_{nt} = P_{na_{nt}}$. Wir haben lokale asymptotische Normalität, also

$$\Lambda_{nt} = tS_n - \frac{1}{2}t^2 J + o_{P_n}(1)$$

mit $S_n = (Dk, g_0)^{-1} S_n(Dk)$ und $J = (Dk, g_0)^{-2} \|Dk\|^2$.

Aus der Differenzierbarkeit von κ folgt

$$n^{1/2}(\kappa(a_{nt}) - \kappa(a)) \to t.$$

Für $V_n = n^{1/2}(\hat{\kappa} - \kappa(a))$ folgt aus der Regularität von $\hat{\kappa}$:

$$V_n - t = n^{1/2}(\hat{\kappa} - \kappa(a_{nt})) + o(1) \Rightarrow V \quad \text{unter } P_{nt}.$$

Es gilt

$$J^{-1} S_n = (Dk, g_0)\|Dk\|^{-2} S_n(Dk).$$

Aus Satz 1 folgt also

$$\begin{aligned}&\big((Dk, g_0)\|Dk\|^{-2} S_n(Dk), n^{1/2}(\hat{\kappa} - \kappa(a)) - (Dk, g_0)\|Dk\|^{-2} S_n(Dk)\big)\\ &\quad \Rightarrow \big((Dk, g_0)\|Dk\|^{-1} N, U_k\big) \quad \text{unter } P_n\end{aligned}$$

mit U_k unabhängig von N. Jetzt Dk gegen g_0 konvergieren lassen. Dann konvergiert (Dk, g_0) gegen $\|g_0\|^2$ und $\|Dk\|$ gegen $\|g_0\|$.

Der Beweis schreibt sich etwas einfacher, wenn $g_0 \in DK$ ist und $Dk = g_0$ gewählt werden kann. — Aus Satz 5 folgt durch Faltung, daß

$$n^{1/2}(\hat{\kappa} - \kappa(a)) \Rightarrow \|g_0\| N + U \quad \text{unter } P_n.$$

Also ist $n^{1/2}(\hat{\kappa} - \kappa(a))$ genau dann asymptotisch maximal in symmetrischen Intervallen um 0 konzentriert, wenn $U = 0$. Das rechtfertigt folgende Definition.

Definition. Ein Schätzer $\hat{\kappa}$ heißt *effizient* für κ in a, wenn

$$n^{1/2}(\hat{\kappa} - \kappa(a)) \Rightarrow \|g_0\| N.$$

Bemerkung 2 Aus Satz 5 folgt, daß ein regulärer und effizienter Schätzer asymptotisch linear ist und den kanonischen Gradienten g_0 als Einflußfunktion hat.

Bemerkung 3 Nach dem zentralen Grenzwertsatz ist ein asymptotisch linearer Schätzer effizient, wenn seine Einflußfunktion der kanonische Gradient ist.

Bemerkung 4 Ein asymptotisch linearer Schätzer ist genau dann regulär, wenn seine Einflußfunktion ein Gradient ist. Mit Bemerkung 2 erhalten wir also folgende Charakterisierung:

Ein Schätzer ist genau dann regulär und effizient, wenn er asymptotisch linear ist und den kanonischen Gradienten als Einflußfunktion hat.

Beweis. Sei f die Einflußfunktion von $\hat{\kappa}$ in a. Sei $k \in K$. Da S_n und D linear sind, folgt aus dem Satz von Cramér und Wold, daß $(S_n(Dk), S_n(f))^\top$ asymptotisch normal unter P_n ist; der Mittelwertvektor ist 0 und die Kovarianzmatrix ist

$$\begin{pmatrix} \|Dk\|^2 & (Dk, f) \\ (Dk, f) & \|f\|^2 \end{pmatrix}.$$

Aus der lokalen asymptotischen Normalität des Modells folgt also, daß der Vektor $(\Lambda_{nk}, S_n(f))^\top$ auch asymptotisch normal unter P_n ist; der Mittelwertvektor ist jetzt $(-\|Dk\|^2/2, 0)^\top$, die Kovarianzmatrix ist unverändert. Nach Lemma 8 ist also $(\Lambda_{nk}, S_n(f))^\top$ auch unter $P_{na_{nk}}$ asymptotisch normal mit derselben Kovarianzmatrix und Mittelwertvektor $(\|Dk\|^2/2, (Dk, f))^\top$. Wir interessieren uns nur für die zweite Komponente: Unter $P_{na_{nk}}$ ist $S_n(f)$ asymptotisch normal mit Varianz $\|f\|^2$ und Mittelwert (Dk, f). Insbesondere folgt aus der asymptotischen Linearität von $\hat{\kappa}$ und der Differenzierbarkeit von κ in a, daß

$$n^{1/2}(\hat{\kappa} - \kappa(a_{nk})) \Rightarrow \|f\|N + (Dk, f - g) \quad \text{unter } P_{na_{nk}}, \quad k \in K.$$

Also ist $\hat{\kappa}$ regulär genau dann, wenn $(Dk, f - g) = 0$ für $k \in K$, also genau dann, wenn f ein Gradient ist.

Bemerkung 5 Sei $A_1 \subset A$ und $a \in A_1$. Sei $K_1 \subset K$ ein lokaler Parameterraum für das Teilmodell P_{na}, $a \in A_1$. Für $k \in K_1$ seien a_{nk} so gewählt, daß sie in A_1 liegen. Sei κ differenzierbar in a mit Gradient g. Dann ergibt sich der kanonische Gradient g_1 für das Teilmodell als Projektion des Gradienten g oder des kanonischen Gradienten g_0 auf $(DK_1)^-$. Für die untere Varianzschranke im Faltungssatz gilt also $\|g_0\| = \|g_1\| + \|g_0 - g_1\| \geq \|g_1\|$.

Bemerkung 6 Sei $g_0 \in DK$. Dann können wir im Beweis von Satz 5 $Dk = g_0$ und $a_{nt} = a_{n,t\|g_0\|^{-2}g_0}$ wählen. Wir brauchen also Regularität nur für das eindimensionale lokale Modell P_{n,tg_0}, $t \in \mathbb{R}$, zu verlangen.

Sei $k \in K$. Nach Bemerkung 5 ist der kanonische Gradient für das eindimensionale lokale Modell $P_{n,tk}$, $t \in \mathbb{R}$, die Projektion g_k von g_0 auf den von k aufgespannten eindimensionalen Raum. Also gilt $\|g_k\| \leq \|g_0\|$. Die Varianzschranke im Faltungssatz wird also durch das *ungünstigste* eindimensionale (lokale) Modell bestimmt.

Es ist anschaulich klar, daß das ungünstigste Modell durch den kanonischen Gradienten beschrieben wird. Er ist die Richtung des steilsten Anstiegs von κ. Bei vorgegebener Differenz zwischen zwei Werten von κ liegen in dieser Richtung die zugehörigen Verteilungen also am engsten beieinander und lassen sich deshalb am schlechtesten unterscheiden.

Bemerkung 7 Der kanonische Gradient ist $g_0 \in (DK)^-$ mit

$$n^{1/2}(\kappa(a_{nk}) - \kappa(a)) \to (Dk, g_0), \quad k \in K.$$

Er läßt sich nicht immer leicht bestimmen. Hat man schon einen Gradienten g, so ergibt sich g_0 durch Projektion von g auf $(DK)^-$. Hat man schon einen regulären Schätzer für κ mit Einflußfunktion f, so ist f ein solcher Gradient, wie in Bemerkung 4 gesehen.

In den Anwendungen gibt es gewöhnlich auch auf K ein inneres Produkt $(k, k')_K$. Es ist oft leichter, den kanonischen Gradienten k_* von κ bezüglich dieses inneren Produkts zu bestimmen,

$$n^{1/2}(\kappa(a_{nk}) - \kappa(a)) \to (k, k_*)_K, \quad k \in K.$$

Ist $g_0 \in DK$, so hat es die Form $g_0 = Dk_0$ für ein $k_0 \in K$. Also bestimmt sich g_0 aus

$$(Dk, Dk_0) = (k, k_*)_K, \quad k \in K.$$

Hat D eine Adjungierte $D^* : DK \to K$, so gilt $(Dk, Dk_0) = (k, D^*Dk_0)_K$, also $k_* = D^*Dk_0$. Ist D^*D invertierbar, so ergibt sich $g_0 = Dk_0$ mit $k_0 = (D^*D)^{-1}k_*$.

Es reicht, einen invertierbaren Operator $C : K \to K$ zu finden mit

$$(Dk, Dk_0) = (k, Ck_0)_K, \quad k \in K.$$

Dann ist $g_0 = Dk_0$ mit $k_0 = C^{-1}k_*$. Falls C^{-1} schwer zu bestimmen ist, kann man manchmal C als Störung $C = I - B$ der Identität I schreiben und C^{-1} in die von-Neumann-Reihe $C^{-1} = \sum_{j=0}^{\infty} B^j$ entwickeln.

Der Faltungssatz überträgt sich auf m-dimensionale Funktionale. Differenzierbarkeit eines Funktionals und asymptotische Linearität eines Schätzer verstehen sich dann komponentenweise; die Regularität (6.1) eines Schätzers muß jetzt mehrdimensional gelesen werden. Bezeichne N_m einen m-dimensionalen $N_{0,I}$-verteilten Zufallsvektor.

Satz 6 *(Mehrdimensionaler Faltungssatz) Ist das Modell lokal asymptotisch normal, das Funktional* κ *differenzierbar mit kanonischem Gradienten* $g_0 = (g_{01}, \dots, g_{0m})^\top$, $(g_0, g_0^\top)$ *positiv definit und der Schätzer* $\hat{\kappa}$ *regulär in* a*, so gilt für einen von* N_m *unabhängigen Zufallsvektor* U*:*

$$(S_n(g_0), n^{1/2}(\hat{\kappa} - \kappa(a)) - S_n(g_0)) \Rightarrow ((g_0, g_0^\top)^{1/2} N_m, U) \quad \textit{unter } P_n.$$

Beweis. Für $b \in \mathbb{R}^m$ ist $b^\top \kappa$ differenzierbar in a mit kanonischem Gradienten $b^\top g_0$. Der Schätzer $b^\top \hat{\kappa}$ ist regulär für $b^\top \kappa$ in a mit Limes $b^\top V$. Nach Satz 5 existiert eine von N_m unabhängige Zufallsvariable U_b mit

$$\begin{aligned} &\big(b^\top S_n(g_0), b^\top n^{1/2}(\hat{\kappa} - \kappa(a)) - b^\top S_n(g_0)\big) \\ &\Rightarrow \big(\big(b^\top (g_0, g_0^\top) b\big)^{1/2} N, U_b\big) \quad \text{unter } P_n. \end{aligned} \tag{6.2}$$

Anderseits sind $S_n(g_0)$ und $n^{1/2}(\hat{\kappa} - \kappa(a))$ straff, also existiert ein Zufallsvektor U, so daß für eine Teilfolge gilt:

$$n^{1/2}(\hat{\kappa} - \kappa(a)) - S_n(g_0) \Rightarrow U.$$

Weil auch (6.2) gilt, muß $b^\top U$ wie U_b verteilt sein. Insbesondere hängt die Verteilung von U nicht von der Teilfolge ab. Also können wir in (6.2) den Limes als $b^\top ((g_0, g_0^\top)^{1/2} N_m, U)$ schreiben. Die Behauptung folgt jetzt aus dem Satz von Cramér und Wold.

Aus Satz 6 folgt wieder durch Faltung, daß

$$n^{1/2}(\hat{\kappa} - \kappa(a)) \Rightarrow (g_0, g_0^\top)^{1/2} N_m + U \quad \text{unter } P_n.$$

Sei Σ eine positiv definite symmetrische $(m \times m)$-Matrix. Für jede um 0 symmetrische konvexe Menge C gilt nach dem Lemma von Anderson (1955)

$$P(\Sigma^{1/2} N_m + U \in C) \le P(\Sigma^{1/2} N_m \in C).$$

Wir nennen deshalb einen Schätzer $\hat{\kappa}$ für ein m-dimensionales Funktional κ *effizient* in a, wenn

$$n^{1/2}(\hat{\kappa} - \kappa(a)) \Rightarrow (g_0, g_0^\top)^{1/2} N_m.$$

Die Charakterisierung regulärer und effizienter Schätzer in Bemerkung 4 bleibt gültig.

7 Effizienz empirischer Schätzer

Seien $X_1, \dots, X_n$ unabhängige Beobachtungen mit unbekannter Verteilung $P \in \mathcal{P}$ auf $(\Omega, \mathcal{F})$. Sei P fest. Wir dürfen annehmen, daß die Familie $\mathcal{P}$ aller möglicherweise vorkommenden Verteilungen zu P äquivalent ist, müssen es aber nicht.

Sei K die Menge der beschränkten Funktionen in $L_{2,0}(P)$. Setze

$$P_{nk}(dx) = P(dx)(1 + n^{-1/2}k(x)), \quad k \in K.$$

Dann ist P_{nk} Hellinger-differenzierbar in P mit Ableitung k,

$$\left\|(dP_{nk}/dP)^{1/2} - 1 - \frac{1}{2}n^{-1/2}k\right\|_2 = o(n^{-1/2}).$$

Setze $L_{nk} = dP_{nk}^n/dP^n$ und $\Lambda_{nk} = \log L_{nk}$. Aus Satz 1 folgt, daß das Modell lokal asymptotisch normal ist,

$$\Lambda_{nk} = n^{-1/2}\sum_{i=1}^n k(X_i) - \frac{1}{2}\|k\|_2^2 + o_P(1), \quad k \in K.$$

Wir sind also in der Situation von Kapitel 6, mit $A = \mathcal{P}$, $H = L_{2,0}(P)$, natürlichem inneren Produkt auf K und H, $D = I$ und

$$S_n(k) = n^{-1/2}\sum_{i=1}^n k(X_i).$$

Sei $f \in L_2(P)$ und $\kappa(P) = Pf$ ein lineares Funktional. Es gilt

$$n^{1/2}(\kappa(P_{nk}) - \kappa(P)) = (k, f) = (k, f_0), \quad k \in K,$$

mit $f_0 = f - Pf \in H$. Also ist κ differenzierbar in P mit kanonischem Gradienten f_0.

Der empirische Schätzer für $\kappa(P) = Pf$ ist

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^n f(X_i).$$

Es gilt

$$n^{1/2}(\hat{\kappa} - \kappa(P)) = n^{-1/2}\sum_{i=1}^n f_0(X_i).$$

Also ist $\hat{\kappa}$ (asymptotisch) linear mit Einflußfunktion f_0, also regulär und effizient nach Bemerkung 4 und Satz 5.

Bemerkung 8 Im nichtparametrischen Modell gibt es nur einen Gradienten. Alle regulären asymptotisch linearen Schätzer sind also asymptotisch äquivalent. Das liegt daran, daß wir H so gewählt haben, daß K dicht in H ist. Es ist ohne weiteres möglich, den Begriff der Regularität zu erweitern. Hier sind zwei einfache Beispiele. In beiden Beispielen schreiben wir die Beobachtungen in Paaren $Y_i = (X_{2i-1}, X_{2i})$, $i = 1, \dots, n$. Zur Vereinfachung der Schreibweise haben wir hier den ursprünglichen Stichprobenumfang verdoppelt. Es sind $Y_1, \dots, Y_n$ unabhängig mit Verteilung $P^2 = P \otimes P$. Sei $H_2 = L_{2,0}(P^2)$.

1. Sei $f_1(x, y) = f(y)$ mit $f \in L_2(P)$. Ein neuer empirischer Schätzer für $\kappa(P) = P^2 f_1 = Pf$ ist

$$\hat{\kappa}_1 = \frac{1}{n} \sum_{i=1}^{n} f_1(Y_i) = \frac{1}{n} \sum_{i=1}^{n} f(X_{2i}).$$

Er ist asymptotisch linear im neuen Sinn, aber nicht im alten. Seine asymptotische Varianz ist doppelt so groß wie die des alten empirischen Schätzers für Pf (jetzt mit $2n$ Beobachtungen).

2. Sei $f_2 \in L_2(P^2)$. Ein empirischer Schätzer für $\kappa(P) = P^2 f_2$ ist

$$\hat{\kappa}_2 = \frac{1}{n} \sum_{i=1}^{n} f_2(Y_i) = \frac{1}{n} \sum_{i=1}^{n} f_2(X_{2i-1}, X_{2i}).$$

Auch er ist asymptotisch linear im neuen Sinn, aber nicht im alten.

Bemerkung 9 *(Nebenbedingungen)* Wenn die Verteilungen strukturellen Nebenbedingungen genügen, können wir häufig diese Nebenbedingungen recht einfach auf das lokale Modell und den empirischen Schätzer übertragen und so effiziente Schätzer gewinnen.

Sei $h : \Omega \to \mathbb{R}$ $\mathcal{F}$-meßbar und $\mathcal{P}$ die Familie der Verteilungen auf $\mathcal{F}$, für die $h \in L_2(P)$ mit $Ph = 0$ und $Ph^2 > 0$ ist. Seien $X_1, \dots, X_n$ unabhängig mit Verteilung P. Sei $f \in L_2(P)$ und $\kappa(P) = Pf$. Ein Schätzer für $\kappa(P)$ ist der empirische Schätzer. Wir wissen schon, wie wir ihn mit der linearen Einschränkung $Ph = 0$ verbessern können. Wir wollen zeigen, daß uns die Theorie der Effizienz auch diese Verbesserung liefert und zusätzlich die Effizienz des verbesserten Schätzers zeigt.

Definiere P_{nk} wie oben. Wegen der Nebenbedingung

$$0 = P_{nk} h = P(1 + n^{-1/2} k) h$$

muß $(k, h) = Pkh = 0$ gelten. Das lokale Modell besteht also aus den Folgen P_{nk} mit $k \in K_1$, wobei

$$K_1 = \{k \in K : (k, h) = 0\}.$$

Natürlich bleibt κ auf diesem Teilmodell differenzierbar, und $f_0 = f - Pf$ ist weiterhin ein Gradient von κ in P. Er ist aber i.a. nicht in K_1^-, also nicht kanonisch. Der *kanonische* Gradient ist die Projektion von f_0 auf K_1^-, also die Funktion $f_1 \in K_1^-$, für die gilt:

$$(k, f_1) = (k, f_0), \quad k \in K_1.$$

Da K_1 aus den $k \in K$ besteht, die zu h orthogonal sind, ist die Projektion von f_0 auf K_1^- von der Form $f_1 = f_0 - ch$. Die Konstante c ergibt sich aus

$$0 = (f_1, h) = (f_0, h) - c(h, h) = (f, h) - c(h, h).$$

Wir erhalten $c = (f, h)/(h, h) = Pfh/Phh$. Ein Schätzer mit der Einflußfunktion f_1 ergibt sich als

$$\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - \hat{c}h(X_i)) \quad \text{mit} \quad \hat{c} = \frac{\sum_{i=1}^{n} f(X_i)h(X_i)}{\sum_{i=1}^{n} h^2(X_i)}.$$

Er ist deshalb regulär und effizient im nichtparametrischen Modell mit der Nebenbedingung $Ph = 0$.

8 LAN für Markov-Ketten

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$ eine Filtration. Für $n \in \mathbb{N}$ seien Y_n $\mathcal{F}_n$-meßbare und integrierbare Zufallsvariablen. Es heißt $(Y_n, \mathcal{F}_n)$, $n \in \mathbb{N}$, *Folge von Martingal-Differenzen*, wenn $E(Y_{n+1}|\mathcal{F}_n) = 0$ für $n \in \mathbb{N}$. Das gilt genau dann, wenn $(Z_n, \mathcal{F}_n)$ mit $Z_n = \sum_{i=1}^{n} Y_i$ ein Martingal ist.

Satz 7 *(Zentraler Grenzwertsatz für Martingale) Sei $(Y_n, \mathcal{F}_n)$ eine Folge von Martingal-Differenzen. Es gelte*

$$\frac{1}{n}\sum_{i=1}^{n} E(Y_{i+1}^2|\mathcal{F}_i) \to \sigma^2 > 0 \quad (P),$$

$$\frac{1}{n}\sum_{i=1}^{n} EY_i^2 \mathbf{1}(|Y_i| > n^{1/2}\varepsilon) \to 0, \quad \varepsilon > 0.$$

Dann gilt $n^{-1/2}\sum_{i=1}^{n} Y_i \Rightarrow \sigma N$.

Dann konvergiert auch der Partialsummenprozeß $t \to n^{-1/2} \sum_{i=1}^{[nt]} Y_i$, $t \in [0,1]$, in Verteilung gegen $t \to \sigma B_t$, $t \in [0,1]$, wobei B die Brownsche Bewegung bezeichnet. Das werden wir aber nicht brauchen. Ein Beweis des Satzes steht in Durrett (1991), Theorem (7.4).

Sei $(E, \mathcal{E})$ ein meßbarer Raum. Eine Folge $X_0, X_1, \ldots$ von Zufallselementen mit Werten in E heißt *stochastischer Prozeß* im *Zustandsraum E*.

Definition. Sei Q ein Kern auf $E \times \mathcal{E}$. Eine (homogene) *Markov-Kette* (bezüglich der natürlichen Filtration) mit *Übergangsverteilung* Q ist charakterisiert durch

$$P(X_{n+1} \in A | X_0, \ldots, X_n) = P(X_{n+1} \in A | X_n) = Q(X_n, A), \quad A \in \mathcal{E}, n \in \mathbb{N}.$$

Die Verteilung μ von X_0 heißt *Startverteilung*. Wir schreiben P_μ für die Verteilung von $X_0, X_1, \ldots$ und P_x, wenn $\mu = \delta_x$. Für Kerne Q, R und eine Verteilung μ schreiben wir

$$Q \otimes R(x, dy, dz) = Q(x, dy) R(y, dz), \quad QR(x, A) = \int Q(x, dy) R(y, A),$$
$$\mu \otimes Q(dx, dy) = \mu(dx) Q(x, dy), \quad \mu Q A = \int \mu(dx) Q(x, A).$$

Zur Abkürzung setzen wir $Q^n = QQ^{n-1}$ und $Q_x f = \int Q(x, dy) f(y)$. Ist $X_0, X_1, \ldots$ eine Markov-Kette mit Startverteilung μ und Übergangsverteilung Q, so hat $(X_n, \ldots, X_{n+m})$ die Verteilung $\mu Q^n \otimes Q \otimes \cdots \otimes Q$. Insbesondere hat X_n die Verteilung μQ^n.

Definition. Eine Verteilung π auf $\mathcal{E}$ heißt *invariant unter* Q, wenn $\pi Q = \pi$.

Ein Prozeß $X_0, X_1, \ldots$ heißt *stationär*, wenn für alle $t, m \in \mathbb{N}_0$ der Vektor $(X_t, \ldots, X_{t+m})$ wie $X_0, \ldots, X_m$ verteilt ist. Eine Markov-Kette mit invarianter Verteilung π ist stationär genau dann, wenn π die Startverteilung ist.

Definition. Eine Markov-Kette heißt *positiv Harris-rekurrent*, wenn sie eine invariante Verteilung π hat und für jedes $x \in E$ und jedes $A \in \mathcal{E}$ mit $\pi A > 0$ gilt:

$$P_x\Big(\sum_{n=0}^{\infty} \mathbf{1}(X_n \in A) = \infty\Big) = 1.$$

Sei $\|\mu\| = \sup_{A \in \mathcal{E}} |\mu A| = |\mu| E$.

Definition. Eine Markov-Kette heißt *aperiodisch*, wenn für alle Startverteilungen μ, ν gilt:

$$\|\mu Q^t - \nu Q^t\| \to 0.$$

Für eine aperiodische Markov-Kette mit invarianter Verteilung π gilt für alle Startverteilungen μ:

$$\|\mu Q^t - \pi\| \to 0.$$

Satz 8 *(Starkes Gesetz der großen Zahl für Markov-Ketten) Sei $X_0, X_1, \dots$ eine positiv Harris-rekurrente und aperiodische Markov-Kette mit invarianter Verteilung π. Dann gilt für jedes π-integrierbare f:*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \to \pi f \quad f.s.$$

Ein Beweis steht in Meyn und Tweedie (1993), Theorem 17.2.7.

Es habe Q eine invariante Verteilung π. Bezeichne $\| \ \|_2$ die Norm auf $L_2(\pi \otimes Q)$. Wir setzen

$$H = \{h \in L_2(\pi \otimes Q) : Qh = 0\}.$$

Definition. Eine Folge Q_n von Übergangsverteilungen heißt *Hellinger-differenzierbar* in Q mit *Ableitung* h, wenn $h \in H$ und

$$\int Q(x,dy)\Big(\Big(\frac{dQ_n}{dQ}\Big)^{1/2}(x,y) - 1 - \frac{1}{2}n^{-1/2}h(x,y)\Big)^2 \leq n^{-1} r_n(x)$$

mit $r_n \downarrow 0$ punktweise und π-integrierbar für große n.

Seien $X_0, \dots, X_n$ Beobachtungen einer Markov-Kette mit Startverteilung μ und Übergangsverteilung Q. Dann haben die Bobachtungen die Verteilung

$$P_{n+1}(dx_0, \dots, dx_n) = \mu(dx_0)\prod_{i=1}^{n} Q(x_{i-1}, dx_i).$$

Bezeichne P'_{n+1} diese Verteilung unter μ_n und Q_n. Dann ist der *Likelihood-Quotient* von der Form

$$L_n(x_0, \dots, x_n) = \frac{dP'_{n+1}}{dP_{n+1}}(x_0, \dots, x_n) = \frac{d\mu_n}{d\mu}(x_0)\prod_{i=1}^{n}\frac{dQ_n}{dQ}(x_{i-1}, x_i).$$

Setze $\Lambda_n = \log L_n$.

Satz 9 *(LAN für Markov-Ketten) Sei die Markov-Kette positiv Harris-rekurrent und aperiodisch mit invarianter Verteilung π, und gelte*

$$\int \left(\left(\frac{d\mu_n}{d\mu} \right)^{1/2} - 1 \right)^2 d\mu \to 0. \tag{8.1}$$

Ist Q_n Hellinger-differenzierbar mit Ableitung h, so gilt

$$\Lambda_n = n^{-1/2} \sum_{i=1}^{n} h(X_{i-1}, X_i) - \frac{1}{2} \|h\|_2^2 + o_p(1),$$

$$n^{-1/2} \sum_{i=1}^{n} h(X_{i-1}, X_i) \Rightarrow \|h\|_2 N.$$

Der Beweis geht wie der von Satz 2. Statt des (schwachen) Gesetzes der großen Zahl für i.i.d. Beobachtungen wenden wir Satz 8 an. Die asymptotische Normalität folgt aus Satz 7, weil $Y_n = h(X_{n-1}, X_n)$, $n \in \mathbb{N}$, eine Folge von Martingal-Differenzen ist.

9 Gleichmäßige Ergodizität und Martingal-Approximation

Sei $V : E \to [1, \infty)$ meßbar. Für eine meßbare Funktion f auf E setze

$$|f|_V = \sup_{x \in E} \frac{|f(x)|}{V(x)}.$$

Es gilt $|f|_V \leq 1$ genau dann, wenn $|f| \leq |V|$. Für ein signiertes Maß μ auf $\mathcal{E}$ setze

$$\|\mu\|_V = \sup_{|f|_V \leq 1} |\mu f|.$$

Sei $\mathcal{L}_V$ die Menge der f mit $|f| < \infty$ und $\mathcal{M}_V$ die Menge der μ mit $\|\mu\|_V < \infty$. Die Mengen $\mathcal{L}_V$ und $\mathcal{M}_V$ sind Banach-Räume für die Normen $|\ |_V$ und $\|\ \|_V$. Es gilt

$$|f|_V = \sup_{\|\mu\|_V \leq 1} |\mu f|.$$

Für einen signierten Kern K auf $E \times \mathcal{E}$ setze

$$\|K\|_V = \sup_{|f|_V \leq 1} |Kf|_V.$$

Aus der Definition der Operatornorm folgt $|Kf|_V \leq \|K\|_V |f|_V$. Für zwei signierte Kerne J, K gilt also $\|JK\|_V \leq \|J\|_V |Kf|_V$, also

$$\|JK\|_V \leq \|J\|_V \|K\|_V.$$

Es gilt

$$\|K\|_V = \sup_{\|\mu\|_V \leq 1} \|\mu K\|_V.$$

Gilt $\|K\|_V < \infty$, so sind also $f \to Kf$ und $\mu \to \mu K$ beschränkte lineare Operatoren auf $\mathcal{L}_V$ und $\mathcal{M}_V$.

Definition. Eine Markov-Kette mit Übergangsverteilung Q heißt *V-gleichmäßig ergodisch*, wenn sie eine invariante Verteilung π hat und

$$\|Q^t - \Pi\|_V \to 0$$

für die *stationäre Projektion* $\Pi(x, dy) = \pi(dy)$ gilt.

Dann ist die Kette auch positiv Harris-rekurrent und aperiodisch; siehe Revuz (1984), Chapter 6, Theorem 3.4. Es gilt

$$Q\Pi(x, dz) = \int Q(x, dy)\Pi(y, dz) = \Pi(x, dz),$$
$$\Pi Q(x, dz) = \int \Pi(x, dy)Q(y, dz) = \Pi(x, dz).$$

Es ergibt sich $(Q^{t-1} - \Pi)(Q - \Pi) = Q^t - \Pi$, also durch Iteration

$$(Q - \Pi)^t = Q^t - \Pi.$$

Wähle t_0, so daß $\varrho = \|(Q - \Pi)^{t_0}\|_V < 1$. Dann gilt

$$\|Q^t - \Pi\|_V = \|(Q - \Pi)^t\|_V = \|(Q - \Pi)^{t_0}\|_V^{t-t_0} = \varrho^{t-t_0} = \varrho^{-t_0}\varrho^t.$$

Also existiert ein $C > 0$ und ein $\varrho < 1$, so daß

$$\|Q^t - \Pi\|_V \leq C\varrho^t. \tag{9.1}$$

Die Kette ist also sogar *geometrisch V-gleichmäßig ergodisch*. Bezeichne $I(x, dy) = \delta_x(dy)$ den *Einheitskern*. Setze $Q^0 = I$. Der Operator

$$U = \sum_{t=0}^{\infty} (Q - \Pi)^t = \sum_{t=0}^{\infty} (Q^t - \Pi)$$

ist beschränkt auf $\mathcal{L}_V$ und die Inverse von $I - Q + \Pi$.

Definition. Eine Markov-Kette mit Übergangsverteilung Q heißt *stark stabil*, wenn es ein $\varepsilon > 0$ gibt, so daß jede Übergangsverteilung Q' mit $\|Q' - Q\|_V < \varepsilon$ eine invariante Verteilung π' besitzt und $\|\pi' - \pi\|_V \to 0$ gilt, wenn $\|Q' - Q\|_V \to 0$.

Wenn $I - Q + \Pi$ eine beschränkte Inverse hat, ist die Kette stark stabil; siehe Kartashov (1985), Theorem 4. Für Verteilungen μ, ν gilt $(\mu - \nu)\Pi = 0$, also

$$\pi' - \pi = \pi'(Q' - Q) + (\pi' - \pi)(Q - \Pi).$$

Durch Iteration ergibt sich

$$\pi' - \pi = \pi'(Q' - Q) + \sum_{t=1}^{m} \pi'(Q' - Q)(Q^t - \Pi) + \pi'(Q' - Q)(Q^{m+1} - \Pi).$$

Wegen (9.1) konvergiert die rechte Seite in der $\| \, \|_V$-Norm gegen $\pi'(Q' - Q)U$, also

$$\pi' - \pi = \pi'(Q' - Q)U.$$

Durch erneute Iteration ergibt sich die Störungsentwicklung

$$\pi' - \pi = \pi \sum_{t=1}^{m} \big((Q' - Q)U\big)^t + \pi'\big((Q' - Q)U\big)^{m+1}. \tag{9.2}$$

Wenn $\|(Q' - Q)U\|_V < 1$ gilt oder zumindest $\|\big((Q' - Q)U\big)^t\|_V < 1$ für ein t, so folgt

$$\pi' - \pi = \pi \sum_{t=1}^{\infty} \big((Q' - Q)U\big)^t.$$

Satz 10 *(Martingal-Approximation) Sei die Markov-Kette V-gleichmäßig ergodisch, sei V π-integrierbar und $|f|_V \leq 1$. Dann gilt*

$$\frac{1}{n} \sum_{i=1}^{n} (f(X_i) - \pi f) = \frac{1}{n} \sum_{i=1}^{n} Af(X_{i-1}, X_i) + O_{P_{n+1}}(n^{-1})$$

mit

$$Af(x, y) = U_y(f - \pi f) - Q_x U(f - \pi f) = \sum_{t=0}^{\infty} (Q_y^t f - Q_x^{t+1} f)$$

und $Q_X Af(X, Y) = 0$.

Beweis. Schreibe $f-\pi f = (I-\Pi)f-(Q-\Pi)f+(Q-\Pi)f$. Durch Iteration ergibt sich

$$\begin{aligned}\sum_{i=1}^{n}(f(X_i)-\pi f) &= \sum_{i=1}^{n}(Uf(X_i)-Q_{X_i}Uf)\\ &= \sum_{i=1}^{n}(Uf(X_i)-Q_{X_{i-1}}Uf)+\sum_{i=1}^{n}(Q_{X_{i-1}}Uf-Q_{X_i}Uf)\\ &= \sum_{i=1}^{n}Af(X_{i-1},X_i)+Q_{X_0}Uf-Q_{X_n}Uf.\end{aligned}$$

Jetzt $P(|Q_{X_n}Uf|>c)\le c^{-1}\mu Q^n|Uf|$ anwenden.

Satz 11 *(Zentraler Grenzwertsatz für Markov-Ketten) Sei die Markov-Kette V-gleichmäßig ergodisch, sei V π-integrierbar und $f^2\le V$. Dann gilt*

$$n^{-1/2}\sum_{i=1}^{n}(f(X_i)-\pi f)\Rightarrow \|Af\|_2 N.$$

Beweis. Nach Satz 10 gilt

$$n^{-1/2}\sum_{i=1}^{n}(f(X_i)-\pi f)=n^{-1/2}\sum_{i=1}^{n}Af(X_{i-1},X_i)+O_{P_{n+1}}(n^{-1/2}).$$

Nach Satz 7 für $Y_i=Af(X_{i-1},X_i)$ gilt

$$n^{-1/2}\sum_{i=1}^{n}Af(X_{i-1},X_i)\Rightarrow \|Af\|_2 N.$$

Die Varianz $\|Af\|_2^2$ von Af ist eine Doppelsumme, aber auf allen Diagonalen stehen alternierende Folgen, denn

$$\begin{aligned}&\iint \pi(dx)Q(x,dy)(Q_y^s f-Q_x^{s+1}f)(Q_y^t f-Q_x^{t+1}f)\\ &\qquad=\iint \pi(dx)Q(x,dy)(Q_y^s f-Q_x^{s+1}f)Q_y^t f,\\ &\iint \pi(dx)Q(x,dy)Q_y^s fQ_y^t f=\pi Q^s fQ^t f,\\ &\iint \pi(dx)Q(x,dy)Q_x^{s+1} fQ_y^t f=\pi Q^{s+1} fQ^{t+1} f.\end{aligned}$$

Also gilt

$$\|Af\|_2^2 = \pi(f-\pi f)^2 + 2\sum_{t=1}^{\infty} \pi(f-\pi f)Q^t f.$$

Die Varianz von $n^{-1/2}\sum_{i=1}^n (f(X_i) - \pi f)$ ist

$$\pi(f-\pi f)^2 + \frac{2}{n}\sum_{1\le i<j\le n} E(f(X_i)-\pi f)(f(X_j)-\pi f).$$

Für $i < j$ gilt

$$E(f(X_i)-\pi f)(f(X_j)-\pi f) = \pi(f-\pi f)Q^{j-i}f.$$

Also konvergiert die Varianz von $n^{-1/2}\sum_{i=1}^n (f(X_i)-\pi f)$ gegen $\|Af\|_2^2$.

Die beiden Sätze lassen sich auf Funktionen f mit mehr als einem Argument verallgemeinern. Sei f eine meßbare Funktion auf E^2. Sei $f^2(x,y) \le V^\alpha(x)V^\beta(y)$ mit $\alpha+\beta = 1$, $\alpha,\beta \le 0$. Wir schreiben

$$(9.3)\quad f(X_{i-1},X_i) - \pi\otimes Qf = f(X_{i-1},X_i) - Q_{X_{i-1}}f + Q_{X_{i-1}}f - \pi\otimes Qf$$

und wenden Satz 10 auf $Q_x f$ statt $f(x)$ an. Es ergibt sich die Martingal-Approximation

$$\frac{1}{n}\sum_{i=1}^n (f(X_{i-1},X_i) - \pi\otimes Qf) = \frac{1}{n}\sum_{i=1}^n Af(X_{i-1},X_i) + O_{P_{n+1}}(n^{-1})$$

mit

$$Af(x,y) = f(x,y) - Q_x f + \sum_{t=1}^{\infty}(Q_y^t f - Q_x^{t+1} f).$$

10 Effizienz empirischer Schätzer für Markov-Ketten

Seien $X_0,\dots,X_n$ Beobachtungen einer Markov-Kette mit Übergangsverteilung Q. Sei Q fest, mit invarianter Verteilung π. Wie in Kapitel 8 sei

$$H = \{h \in L_2(\pi\otimes Q) : Qh = 0\}.$$

Sei K die Menge der beschränkten Funktionen in H. Für $k \in K$ setze

$$Q_{nk}(x,dy) = Q(x,dy)(1 + n^{-1/2}k(x,y)).$$

Dann ist Q_{nk} Hellinger-differenzierbar mit Ableitung k.

Die Kette sei V-gleichmäßig ergodisch unter Q, also nach Kapitel 9 auch positiv Harris-rekurrent und aperiodisch. Es gelte (8.1) für die Startverteilung μ_{nk} zu Q_{nk}. Dann ist nach Satz 9 das Modell lokal asymptotisch normal,

$$\Lambda_{nk} = n^{-1/2} \sum_{i=1}^{n} k(X_{i-1}, X_i) - \frac{1}{2}\|k\|_2^2 + o_p(1),$$

$$n^{-1/2} \sum_{i=1}^{n} k(X_{i-1}, X_i) \Rightarrow \|k\|_2 N.$$

Für $|k| \leq c$ gilt

$$\|Q_{nk} - Q\|_V \leq n^{-1/2} c \|Q\|_V.$$

Sei $\|Q\|_V < \infty$. Nach Kapitel 9 ist die Kette stark stabil unter Q. Für $n \geq n_c$ hat also jedes Q_{nk} mit $|k| \leq c$ eine invariante Verteilung π_{nk}, und

$$\sup_{|k|\leq c} \|\pi_{nk} - \pi\|_V \to 0.$$

Da U beschränkt ist, gilt andererseits auch

$$\sup_{|k|\leq c} \|(Q_{nk} - Q)U\|_V = O(n^{-1/2}).$$

Die Störungsentwicklung (9.2) für $m = 1$ liefert dann gleichmäßig für $|k| \leq c$:

$$\|\pi_{nk} - \pi - \pi(Q_{nk} - Q)U\|_V = \|\pi_{nk}\big((Q_{nk} - Q)U\big)^2\|_V = O(n^{-1}).$$

Ist V π-integrierbar und $f^2 \leq V$, so ergibt sich

$$n^{1/2}(\pi_{nk} f - \pi f) \to \iint \pi(dx) Q(x, dy) k(x, y) U_y f = \pi \otimes QkAf.$$

Es gilt $Af \in K^- = H$. Also ist πf differenzierbar in Q mit kanonischem Gradienten Af.

Das überträgt sich auf Funktionen mit mehr als einem Argument. Sei f eine meßbare Funktion auf E^2. Sei $f^2(x, y) \leq V^\alpha(x) V^\beta(y)$ mit $\alpha + \beta \leq 1$ und $\alpha, \beta \geq 0$. Dann gilt

$$n^{1/2}(\pi_{nk} \otimes Q_{nk} f - \pi \otimes Qf) \to \pi \otimes QkAf. \tag{10.1}$$

Also ist $\pi \otimes Qf$ differenzierbar in Q mit kanonischem Gradienten Af.

Satz 12 *Seien* $X_0, \ldots, X_n$ *Beobachtungen einer Markov-Kette mit Übergangsverteilung* Q *und invarianter Verteilung* π*. Die Kette sei* V*-gleichmäßig ergodisch mit* $\|Q\|_V < \infty$*, und* V *sei* π*-integrierbar. Sei* f *eine meßbare Funktion auf* E^2 *mit* $f^2(x,y) \leq V^\alpha(x)V^\beta(y)$ *und* $\alpha + \beta \leq 1$*,* $\alpha, \beta \geq 0$*. Der lokale Parameter-Raum sei* K*, und die Startverteilungen* μ_{nk} *zu* Q_{nk} *erfüllen (8.1). Dann ist der empirische Schätzer* $(1/n)\sum_{i=1}^n f(X_{i-1}, X_i)$ *regulär und effizient für* $\pi \otimes Qf$*.*

Beweis. Das Modell ist lokal asymptotisch normal. Nach (10.1) ist Af der kanonische Gradient von $\pi \otimes Qf$. Nach (9.3) ist Af die Einflußfunktion des empirischen Schätzers. Die Behauptung folgt aus der Charakterisierung in Bemerkung 4.

Bemerkung 10 *(Störungsentwicklung mit Kontiguitätsargument)* Den kanonischen Gradienten aus (10.1) erhalten wir auch mit dem dritten Lemma von Le Cam. Wir brauchen also eigentlich die Störungsentwicklung aus Kapitel 9 nicht, wohl aber die Martingal-Approximation. Bezeichne $\mathbb{E}f = \frac{1}{n}\sum_{i=1}^n f(X_{i-1}, X_i)$ den empirischen Schätzer für $\pi \otimes Qf$. Mit Satz 10 ist $(\Lambda_{nk}, n^{1/2}(\mathbb{E}f - \pi \otimes Qf))^\top$ unter P_{n+1} asymptotisch normal mit Mittelwertvektor $(-\|k\|_2^2/2, 0)^\top$ und Kovarianz $(k, Af)_2$. Nach Lemma 8 ist also $n^{1/2}(\mathbb{E}f - \pi \otimes Qf)$ unter $P_{n+1,nk}$ asymptotisch normal mit Mittelwert $(k, Af)_2$. Andererseits hat $\mathbb{E}f$ unter $P_{n+1,nk}$ den Mittelwert $\pi_{nk} \otimes Q_{nk}f$. Also gilt

$$n^{1/2}(\pi_{nk} \otimes Q_{nk}f - \pi \otimes Qf) \to (k, Af)_2.$$

Definition. Sei $m \in \mathbb{N}$ und Q ein Kern auf $E^m \times \mathcal{E}$. Wir nennen eine Folge $X_{-m+1}, X_{-m+2}, \ldots$ von Zufallselementen mit Werten in E *Markov-Kette* der *Ordnung* m, wenn

$$\begin{aligned} P(X_{n+1} \in A | X_{-m+1}, \ldots, X_n) &= P(X_{n+1} \in A | X_{n-m+1}, \ldots, X_n) \\ &= Q(X_{n-m+1}, \ldots, X_n, A), \quad A \in \mathcal{E}^m,\ n \in \mathbb{N}. \end{aligned}$$

Wir setzen

$$\begin{aligned} &Q \otimes Q(x_1, \ldots, x_m, dx_{m+1}, dx_{m+2}) \\ &= Q(x_1, \ldots, x_m, dx_{m+1}) Q(x_2, \ldots, x_{m+1}, dx_{m+2}) \end{aligned}$$

etc. Eine Verteilung π auf $\mathcal{E}^m$ heißt *invariant* unter Q, wenn $\pi \otimes Q^m = \pi$. Die Martingal-Approximation und die Effizienz empirischer Schätzer übertragen sich auf Markov-Ketten höherer Ordnung. Letztere lassen sich auch als spezielle Markov-Ketten erster Ordnung interpretieren. Setze $\mathbf{X}_n = (X_{n-m+1}, \dots, X_n)^\top$. Dann ist $\mathbf{X}_0, \mathbf{X}_1, \dots$ eine Markov-Kette der Ordnung 1 mit Übergangsverteilung $Q^{\otimes m} = Q \otimes \cdots \otimes Q$ auf $E^m \times \mathcal{E}^m$.

Bemerkung 11 *(Mehrdimensionale stationäre Verteilung)* Sei $X_0, X_1, \dots$ eine Markov-Kette (der Ordnung 1). Bezeichne

$$B_3(dx, dy, dz) = \pi(dx) Q(x, dy) Q(y, dz)$$

die stationäre Verteilung von (X_0, X_1, X_2). Sei $f \in L_2(B_3)$. Wir wollen $B_3 f$ schätzen. Der empirische Schätzer ist

$$\mathbb{B}_3 f = \frac{1}{n-1} \sum_{i=2}^n f(X_{i-2}, X_{i-1}, X_i).$$

Wir erhalten eine Martingal-Approximation wie folgt. Setze

$$\begin{aligned} Q_{xy} f &= \int Q(y, dz) f(x, y, z), \\ Q_x^2 f &= \int Q(x, dy) Q_{xy} f, \qquad Q^t f = Q Q^{t-1} f, \quad t = 3, 4, \dots. \end{aligned}$$

Schreibe

$$f(x, y, z) = f(x, y, z) - Q_{xy} f + Q_{xy} - Q_x^2 f + Q_x^2 f.$$

Mit Satz 10 für $Q^2 f$ statt f ergibt sich

$$n^{1/2}(\mathbb{B}_3 f - B_3 f) = n^{-1/2} \sum_{i=2}^n Af(X_{i-2}, X_{i-1}, X_i) + o_p(1)$$

mit

$$Af(x, y, z) = f(x, y, z) - Q_{xy} f + Q_{yz} f - Q_y^2 f + AQ^2 f(y, z).$$

Der Schätzer $\mathbb{B}_3 f$ ist *nicht* asymptotisch linear, denn Af ist nicht in H. Insbesondere ist er nicht effizient. Er ist aber effizient im größeren Modell der Markov-Ketten der Ordnung 2. Dazu schreiben wir $B_3(dx, dy, dz) = B_2(dx, dy) Q(x, y, dz)$ in Analogie zur Darstellung der Ketten der Ordnung 1. Die Einflußfunktion ist dann

$$Af(x, y, z) = f(x, y, z) - Q_{xy} f + AQf(x, y, z).$$

Es gibt verschiedene Möglichkeiten, effiziente Schätzer für B_3f (für Ketten der Ordnung 1) zu konstruieren. Sie verwenden die Faktorisierung

$$B_3(dx, dy, dz) = \pi(dx)Q(x, dy)Q(y, dz).$$

Sei die Kette reellwertig, und habe $Q(x, dy)$ die Lebesgue-Dichte $q(x, y)$. Dann hat π ebenfalls eine Lebesgue-Dichte, sagen wir b. Bezeichne $b_2(x, y) = b(x)q(x, y)$ die Dichte von $B_2 = \pi \otimes Q$. Dann hat B_3 die Dichte

$$b(x)q(x, y)q(y, z) = \frac{b_2(x, y)b_2(y, z)}{b(y)}.$$

Kernschätzer für b und b_2 sind

$$\hat{b}(y) = \frac{1}{n}\sum_{i=1}^{n} k_b(y - X_i), \quad \hat{b}_2(x, y) = \frac{1}{n}\sum_{i=1}^{n} k_b(x - X_{i-1})k_b(y - X_i).$$

Wir schätzen B_3f durch den Plug-in-Schätzer

$$\hat{B}_3f = \iiint \frac{\hat{b}_2(x, y)\hat{b}_2(y, z)}{\hat{b}(y)} f(x, y, z)\, dx\, dy\, dz$$

oder durch

$$\bar{B}_3f = \frac{1}{n}\sum_{i=1}^{n} \int \frac{\hat{b}_2(X_i, z)}{\hat{b}(X_i)} f(X_{i-1}, X_i, z)\, dz.$$

Der erste Schätzer ist eine geglättete Version des zweiten. Wir wollen hier die Einflußfunktion nicht ausrechnen. Wir erwarten aber, daß sie mit dem kanonischen Gradienten übereinstimmt. Um letzteren zu bestimmen, schreiben wir

$$\begin{aligned}
&n^{1/2}(B_{2nk} \otimes Q_{nk}f - B_2 \otimes Qf)\\
&= n^{1/2}(B_{2nk} - B_2) \otimes Qf + B_2 \otimes n^{1/2}(Q_{nk} - Q)f\\
&\qquad + n^{1/2}(B_{2nk} - B_2) \otimes (Q_{nk} - Q)f\\
&= B_2kAQf + E_\pi k(X_1, X_2)f(X_0, X_1, X_2) + O(n^{-1/2}).
\end{aligned}$$

Die *umgekehrte Übergangsverteilung* Q^- ist definiert durch

$$\pi(dx)Q(x, dy) = \pi(dy)Q^-(y, dx).$$

Es gilt

$$B_3(dx, dy, dz) = B_2(dy, dz)Q^-(y, dx).$$

Setze

$$Q_{yz}^{-}f = \int Q^{-}(y, dx)f(x, y, z).$$

Es ergibt sich

$$E_{\pi}k(X_1, X_2)f(X_0, X_1, X_2) = \int B_2(dy, dz)k(y, z)Q_{yz}^{-}f.$$

Wir projizieren $Q^{-}f$ auf H und erhalten den kanonischen Gradienten

$$g_0(x, y) = AQf(x, y) + Q_{xy}^{-}f - Q_x Q^{-}f.$$

Das muß die Projektion der Einflußfunktion des empirischen Schätzers im größeren Modell sein. Im kleineren Modell gilt $Q(x, y, dz) = Q(y, dz)$, und $Af(x, y, z)$ vereinfacht sich zu

$$Af(x, y, z) = f(x, y, z) - Q_{xy}f + AQf(y, z).$$

Wir erhalten unter Weglassen der bedingten Zentrierungen

$$\begin{aligned}
&\iiint B_3(dx, dy, dz)(Af(x, y, z) - g_0(y, z))k(y, z) \\
&= \iiint B_3(dx, dy, dz)(f(x, y, z) + AQf(y, z) - AQf(y, z) - Q_{yz}^{-}f)k(y, z) \\
&= 0.
\end{aligned}$$

Bemerkung 12 *(Falsch spezifiziertes Modell)* Im folgenden untersuchen wir heuristisch ein *nicht*lineares Funktional auf einem nichtparametrischen Modell. Sei Q_ϑ, $\vartheta \in \Theta \subset \mathbb{R}$, eine parametrische Familie von Übergangsverteilungen mit Dichten q_ϑ. Sei die wahre Übergangsverteilung Q unbekannt. Was schätzt dann der Maximum-Likelihood-Schätzer?

Sei π die invariante Verteilung von Q. Setze $B = \pi \otimes Q$. Die *Kullback–Leibler-Information* ist $B \log q_\vartheta$. Sei $\vartheta(B)$ der Parameter ϑ, der die Kullback–Leibler-Information maximiert. Ein plausibler Schätzer für $\vartheta(B)$ ist die empirische Version dieses Funktionals. Bezeichne

$$\mathbb{B} = \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_{i-1}, X_i)}.$$

Sei $\hat{\vartheta}$ der Parameter ϑ, der die empirische Kullback–Leibler-Information

$$\vartheta(\mathbb{B}) = \frac{1}{n}\sum_{i=1}^{n} \log q_\vartheta(X_{i-1}, X_i)$$

maximiert. Das ist der Maximum-Likelihood-Schätzer.

Welche Einflußfunktion hat der Maximum-Likelihood-Schätzer? Wir zeigen, daß er auch im falsch spezifizierten Modell effizient bleibt. Sei $\dot\ell_\vartheta = \dot q_\vartheta / q_\vartheta$. Nach Definition von $\vartheta(B)$ gilt

$$0 = \partial_{\vartheta=\vartheta(B)} B \log q_\vartheta = B\dot\ell_{\vartheta(B)}.$$

Die Einflußfunktion von $\hat\vartheta$ unter Q ergibt sich durch Taylor-Entwicklung

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n \dot\ell_{\hat\vartheta}(X_{i-1}, X_i) \\ &= \frac{1}{n}\sum_{i=1}^n \dot\ell_{\vartheta(B)}(X_{i-1}, X_i) + (\hat\vartheta - \vartheta(B))\frac{1}{n}\sum_{i=1}^n \ddot\ell_{\vartheta(B)}(X_{i-1}, X_i) \\ &\quad + o_{P_{n+1}}(n^{-1/2}), \end{aligned}$$

also durch Auflösen nach $\hat\vartheta$ und mit dem Gesetz der großen Zahl:

$$n^{1/2}(\hat\vartheta - \vartheta(B)) = -(B\ddot\ell_{\vartheta(B)})^{-1} n^{-1/2} \sum_{i=1}^n \dot\ell_{\vartheta(B)}(X_{i-1}, X_i) + o_{P_{n+1}}(n^{-1/2}).$$

Es gilt zwar, wie oben gezeigt, $B\dot\ell_{\vartheta(B)} = 0$, aber $\dot\ell_{\vartheta(B)}(X_{i-1}, X_i)$ ist i.a. keine Martingal-Differenz mehr. Mit Satz 10 folgt aber, daß $\hat\vartheta$ unter Q asymptotisch linear für $\vartheta(B)$ mit Einflußfunktion $-(B\ddot\ell_{\vartheta(B)})^{-1} A\dot\ell_{\vartheta(B)}$ ist.

Ist der Maximum-Likelihood-Schätzer effizient? Sei $B_{nk} = \pi_{nk} \otimes Q_{nk}$. Durch Taylor-Entwicklung erhalten wir

$$0 = B_{nk}\dot\ell_{\vartheta(B_{nk})} = B_{nk}\dot\ell_{\vartheta(B)} + (\vartheta(B_{nk}) - \vartheta(B))B_{nk}\ddot\ell_{\vartheta(B)} + o(n^{-1/2}).$$

Es gilt $B_{nk}\ddot\ell_{\vartheta(B)} \to B\ddot\ell_{\vartheta(B)}$ und

$$n^{1/2}(B_{nk}\dot\ell_{\vartheta(B)} - B\dot\ell_{\vartheta(B)}) \to BkA\dot\ell_{\vartheta(B)}.$$

Also ergibt sich durch Auflösen nach $\vartheta(B_{nk})$:

$$n^{1/2}(\vartheta(B_{nk}) - \vartheta(B)) \to -(B\ddot\ell_{\vartheta(B)})^{-1} BkA\dot\ell_{\vartheta(B)}.$$

Es gilt $A\dot\ell_{\vartheta(B)} \in K^- = H$. Also ist $g = -(B\ddot\ell_{\vartheta(B)})^{-1} A\dot\ell_{\vartheta(B)}$ der kanonische Gradient von $\vartheta(B)$. Das ist die Einflußfunktion des Maximum-Likelihood-Schätzers.

11 ARMA-Prozesse

Sei X_t, $t \in \mathbb{Z}$, ein reellwertiger Prozeß (eine *Zeitreihe*) auf $(\Omega, \mathcal{F}, P)$.

Definition. Es gelte $EX_t^2 < \infty$ für $t \in \mathbb{Z}$. Dann ist die *Autokovarianz-Funktion* definiert durch

$$\gamma(r,s) = \operatorname{Cov}(X_r, X_s), \quad r, s \in \mathbb{Z}.$$

Es gilt $\gamma(r,s) = \gamma(s,r)$.

Definition. Eine Zeitreihe (X_t) heißt *schwach stationär*, wenn für ein $m \in \mathbb{R}$ und für $r, s, t \in \mathbb{Z}$ gilt:

$$EX_t^2 < \infty, \quad EX_t = m, \quad \gamma(r,s) = \gamma(r+t, s+t).$$

Dann ist die Autokovarianz-Funktion durch die Funktion

$$\gamma(s) = \gamma(s,0) = \gamma(t+s,t)$$

beschrieben. Es gilt $\gamma(r,s) = \gamma(r-s)$. Ist die Zeitreihe (strikt) stationär mit endlichen zweiten Momenten, dann ist sie schwach stationär.

Definition. Eine Zeitreihe (X_t) heißt *Gaußsch*, wenn $(X_{t_1}, \dots, X_{t_m})^\top$ normalverteilt ist für alle $\{t_1, \dots, t_m\} \subset \mathbb{Z}$ und $m \in \mathbb{N}$. Sie heißt *zentriert*, wenn $EX_t = 0$ für $t \in \mathbb{Z}$.

Ein schwach stationärer Gaußscher Prozess ist stationär, denn eine mehrdimensionale Normalverteilung ist durch ihre Mittelwerte, Varianzen und Kovarianzen bestimmt.

Sei μ ein m-dimensionaler Vektor und Σ eine symmetrische und positiv definite $m \times m$ Matrix. Die *m-dimensionale Normalverteilung* $N_{\mu,\Sigma}$ hat die Dichte

$$p_{\mu,\Sigma}(x) = (2\pi)^{-m/2} (\det \Sigma)^{-1/2} \exp\Big(-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\Big).$$

Sei I die m-dimensionale Einheitsmatrix. Die *m-dimensionale Standard-Normalverteilung* $N_{0,I}$ hat die Dichte

$$p_{0,I}(x) = (2\pi)^{-m/2} \exp\Big(-\frac{1}{2}x^\top x\Big) = (2\pi)^{-m/2} \prod \exp(-x_j^2/2).$$

Das ist ein unabhängiges Produkt, die Verteilung eines Vektors mit unabhängigen eindimensionalen standard-normalverteilten Komponenten. Die

Kovarianzmatrix von $N_{0,I}$ ist also I. Ist X standard-normalverteilt, so folgt $\Sigma^{1/2}X + \mu$ der Verteilung $N_{\mu,\Sigma}$. Also hat $N_{\mu,\Sigma}$ den Mittelwertvektor μ und die Kovarianzmatrix Σ.

Ist Σ symmetrisch und positiv *semi*definit, so nennen wir weiterhin die Verteilung von $\Sigma^{1/2}X + \mu$ eine m-dimensionale Normalverteilung. Sie lebt auf einem affinen Teilraum von $\mathbb{R}^m$.

Für die Autokovarianz-Funktion eines schwach stationären Prozesses gilt

$$\begin{aligned} \gamma(0) &= E(X_0 - \mu)^2 \geq 0, \\ |\gamma(s)| &= |E(X_s - \mu)(X_0 - \mu)| \leq \gamma(0), \\ \gamma(s) &= \gamma(s, 0) = \gamma(0, -s) = \gamma(-s, 0) = \gamma(-s). \end{aligned}$$

Definition. Eine Funktion $\gamma : \mathbb{Z} \to \mathbb{R}$ heißt *positiv semidefinit*, wenn

$$\sum_{i,j=1}^{m} a_i \gamma(t_i - t_j) a_j \geq 0$$

für $a_i \in \mathbb{R}$, $t_i \in \mathbb{Z}$, $m \in \mathbb{N}$.

Satz 13 *(Charakterisierung von Autokovarianz-Funktionen) Eine Funktion $\gamma : \mathbb{Z} \to \mathbb{R}$ ist Autokovarianz-Funktion einer schwach stationären Zeitreihe genau dann, wenn sie positiv semidefinit und gerade ist.*

Beweis. *Hinreichend.* Sei γ die Autokovarianz-Funktion einer schwach stationären Zeitreihe. Setze $a = (a_1, \ldots, a_m)^\top$, $t = (t_1, \ldots, t_m)^\top$,

$$Z = (X_{t_1} - EX_{t_1}, \ldots, X_{t_m} - EX_{t_m})^\top.$$

Es gilt

$$0 \leq \operatorname{Var}(a^\top Z) = a^\top EZZ^\top a = \sum_{i,j=1}^{m} a_i \gamma(t_i - t_j) a_j.$$

Notwendig. Sei γ positiv semidefinit und gerade. Zu $t = (t_1, \ldots, t_m)^\top$ sei Γ_t die Matrix mit Einträgen $\gamma(t_i - t_j)$. Dann ist Γ_t symmetrisch und positiv semidefinit. Sei $(X_{t_1}, \ldots, X_{t_m})^\top$ verteilt wie N_{0,Γ_t}. Die Familie dieser Verteilungen ist offensichtlich projektiv. Nach dem Fortsetzungssatz von Kolmogorov is (X_t) also ein zentrierter stationärer Gaußscher Prozeß mit Autokovarianz-Funktion γ.

Definition. Ein Prozeß (Z_t) heißt *weißes Rauschen* mit (Mittelwert 0 und) Varianz σ^2 , wenn er ein schwach stationärer Prozeß mit $EZ_t = 0$ und $\gamma(0) = \sigma^2$ und $\gamma(t) = 0$ für $t \neq 0$ ist.

Der *Backshift* B ist definiert durch $BX_t = X_{t-1}$.

Definition. Sei $p, q \in \mathbb{N}$. Setze $\varrho(z) = 1 - \varrho_1 z - \cdots - \varrho_p z^p$ und $\varphi(z) = 1 + \varphi_1 z + \cdots + \varphi_q z^q$. Ein schwach stationärer Prozeß (X_t) heißt *ARMA(p,q)*, wenn $\varrho(B)X_t = \varphi(B)Z_t$ für ein weißes Rauschen (Z_t). Ausführlich:

$$X_t - \varrho_1 X_{t-1} - \cdots - \varrho_p X_{t-p} = Z_t + \varphi_1 Z_{t-1} + \cdots + \varphi_q Z_{t-q}.$$

Wir setzen $\varrho_0 = 1$ und $\varphi_0 = 1$.

Beispiel. Für $p = 0$, also $\varrho(z) = 1 = z^0$, ergibt sich der MA(q)-Prozeß $X_t = \varphi(B)Z_t$. Es gilt $EX_t = \sum_{j=0}^q EZ_{t-j} = 0$ und

$$\mathrm{Cov}\,(X_{t+s}, X_t) = \sigma^2 \sum_{j=0}^{q-s} \varphi_j \varphi_{j+s} = \gamma(s).$$

Also ist (X_t) schwach stationär mit Autokovarianz-Funktion γ, und $\gamma(t) = 0$ für $|t| > q$.

Beispiel. Für $q = 0$, also $\varphi(z) = 1 = z^0$, ergibt sich der AR(p)-Prozeß $\varrho(B)X_t = Z_t$. Er ist nicht immer schwach stationär. Für $p = 1$ ist er von der Form $X_t = Z_t + \varrho X_{t-1}$. Durch Iteration ergibt sich

$$X_t = Z_t + \varrho Z_{t-1} + \varrho^2 X_{t-2} = \sum_{j=0}^{m-1} \varrho^j Z_{t-j} + \varrho^m X_{t-m}.$$

Falls EX_t konstant ist und $|\varrho| < 1$ gilt, folgt

$$\Big\| X_t - \sum_{j=0}^{m-1} \varrho^j Z_{t-j} \Big\|_2^2 = \varrho^{2m} \|X_{t-m}\|_2^2 \to 0, \quad m \to \infty.$$

Also gilt in $L_2(P)$:

$$X_t = \sum_{j=0}^{\infty} \varrho^j Z_{t-j}.$$

Wir haben also eine MA(∞)-Darstellung für AR(1). Die schwache Stationarität ergibt sich wie im vorigen Beispiel. Umgekehrt ergibt sich wieder die AR(1)-Darstellung:

$$X_t = Z_t + \sum_{j=1}^{\infty} \varrho^j Z_{t-j} = Z_t + \varrho \sum_{j=0}^{\infty} \varrho^j Z_{t-j-1} = Z_t + \varrho X_{t-1}.$$

Das gilt allgemeiner, und auch f.s.

Definition. Eine Zeitreihe (X_t) heißt kausal, wenn

$$X_t = \sum_{j=0}^{\infty} \vartheta_j Z_{t-j} \quad \text{mit} \quad \sum_{j=0}^{\infty} |\vartheta_j| < \infty.$$

Satz 14 *Gilt* $\sup E|X_t| < \infty$ *und* $\sum_{j=-\infty}^{\infty} |\vartheta_j| < \infty$*, dann ist die Reihe*

$$\vartheta(B)X_t = \sum_{j=-\infty}^{\infty} \vartheta_j X_{t-j}$$

f.s. absolut konvergent. Gilt auch $\sup EX_t^2 < \infty$*, so konvergiert sie auch in* $L_2(P)$*.*

Beweis. Mit dem Satz von der monotonen Konvergenz gilt

$$E \sum_{j=-\infty}^{\infty} |\vartheta_j||X_{t-j}| = \lim_{n\to\infty} E \sum_{j=-n}^{n} |\vartheta_j||X_{t-j}| \leq \sup E|X_t| \sum_{j=-\infty}^{\infty} |\vartheta_j|.$$

Also konvergieren $\sum_{j=-\infty}^{\infty} |\vartheta_j||X_{t-j}|$ und $\vartheta(B)X_t$ f.s. Gelte $\sup EX_t^2 < \infty$. Für $n > m > 0$ gilt

$$\begin{aligned} E\Big(\sum_{m<|j|\leq n} \vartheta_j X_{t-j} \Big)^2 &= \sum_{m<|i|,|j|\leq n} \vartheta_i \vartheta_j EX_{t-i}EX_{t-j} \\ &\leq \sup EX_t^2 \Big(\sum_{m<|j|\leq n} |\vartheta_j| \Big)^2 \to 0, \quad m, n \to \infty. \end{aligned}$$

Also konvergiert nach dem Cauchy-Kriterium $\vartheta(B)X_t$ in $L_2(P)$, also in Wahrscheinlichkeit. Bezeichnet $\vartheta(B)X_t$ den f.s. Limes und S den Limes in $L_2(P)$, so gilt mit dem Lemma von Fatou

$$\begin{aligned} E(S - \vartheta(B)X_t)^2 &= E \lim_{n\to\infty} \Big(S - \sum_{j=-n}^{n} \vartheta_j X_{t-j} \Big)^2 \\ &\leq \liminf_{n\to\infty} E\Big(S - \sum_{j=-n}^{n} \vartheta_j X_{t-j} \Big)^2 = 0. \end{aligned}$$

Satz 15 *Sei* (Y_t) *schwach stationär mit Autokovarianz-Funktion* γ_Y*, und gelte* $\sum_{j=-\infty}^{\infty} |\vartheta_j| < \infty$*. Dann ist* $X_t = \vartheta(B)Y_t$ *schwach stationär mit* $EX_t = EY_t \sum_{j=-\infty}^{\infty} \vartheta_j$ *und Autokovarianz-Funktion*

$$\gamma_X(s) = \sum_{i,j=-\infty}^{\infty} \vartheta_i \vartheta_j \gamma_Y(s-i+j).$$

Beweis. Nach Satz 14 ist $\vartheta(B)Y_t$ f.s. absolut konvergent und in $L_2(P)$ konvergent. Also gilt

$$EX_t = \lim_{n\to\infty} \sum_{j=-n}^{n} \vartheta_j EY_{t-j} = EY_t \sum_{j=-\infty}^{\infty} \vartheta_j$$

und

$$\begin{aligned} EX_{t+s}X_t &= \lim_{n\to\infty} \sum_{i,j=-n}^{n} \vartheta_i \vartheta_j EY_{t+s-i}Y_{t-j} \\ &= \sum_{i,j=-\infty}^{\infty} \vartheta_i \vartheta_j (\gamma_Y(s-i+j) + (EY_t)^2). \end{aligned}$$

Für γ_X ergibt sich

$$\gamma_X(s) = EX_{t+s}X_t - (EX_t)^2 = \sum_{i,j=-\infty}^{\infty} \vartheta_i \vartheta_j \gamma_Y(s-i+j).$$

Für Potenzreihen im Backshift-Operator gelten die üblichen Rechenregeln. Ist $\sum |\alpha_j| < \infty$ und $\sum |\beta_j| < \infty$ und (X_t) schwach stationär, dann ist $\alpha(B)\beta(B)X_t$ wohldefiniert, und es gilt

$$\alpha(B)\beta(B)X_t = \beta(B)\alpha(B)X_t = \vartheta(B)X_t$$

mit

$$\vartheta_j = \sum_{i=-\infty}^{\infty} \alpha_i \beta_{j-i} = \sum_{i=-\infty}^{\infty} \beta_i \alpha_{j-i}.$$

Denn mit $Z_t = \sum_i \alpha_{t-i}$ und $Y_t = \sum_j \beta_j X_{t-j}$ gilt

$$Z_t = \sum_i \alpha_i \sum_j \beta_j X_{t-i-j} = \sum_i \alpha_i \sum_j \beta_{j-i} X_{t-j} = \sum_j \Big(\sum_i \alpha_i \beta_{j-i}\Big) X_{t-j}.$$

Satz 16 *Sei (X_t) ARMA(p,q), und es gelte, daß die zugehörigen Polynome ϱ und φ keine gemeinsamen Nullstellen haben. Dann ist (X_t) genau dann kausal, wenn ϱ keine Nullstellen auf der abgeschlossenen komplexen Einheits-Kreisscheibe hat. Die kausalen Koeffizienten ergeben sich aus $\psi = \varphi/\varrho$.*

Beweis. *Notwendig.* Sei $\varrho(z) \neq 0$ für $|z| \leq 1$. Dann gilt $\varrho(z) \neq 0$ für $|z| < 1 + \varepsilon$; also gibt es eine Reihenentwicklung

$$\frac{1}{\varrho(z)} = \sum_{j=0}^{\infty} \vartheta_j z^j = \vartheta(z), \quad |z| < 1 + \varepsilon.$$

Insbesondere gilt $\vartheta_j (1 + \varepsilon/2)^j \to 0$. Also existiert ein $C > 0$ mit $|\vartheta_j| \leq C(1 + \varepsilon/2)^{-j}$. Insbesondere gilt $\sum |\vartheta_j| < \infty$. Wegen $\vartheta(z)\varrho(z) = 1$ folgt mit Satz 15:

$$X_t = \vartheta(B)\varrho(B)X_t = \vartheta(B)\varphi(B)Z_t.$$

Hinreichend. Sei (X_t) kausal, $X_t = \psi(B)Z_t$. Dann gilt

$$\varphi(B)Z_t = \varrho(B)X_t = \varrho(B)\psi(B)Z_t,$$

also $\varphi(z) = \varrho(z)\psi(z)$ für $|z| \leq 1$, also $\varrho(z) \neq 0$, falls $\varphi(z) \neq 0$. Nach Voraussetzung gilt also $\varrho(z) \neq 0$ für alle $|z| \leq 1$.

Definition. Sei (X_t) ARMA(p,q), $\varrho(B)X_t = \varphi(B)Z_t$. Dann heißt (X_t) *invertierbar*, wenn es π_j mit $\sum_{j=0}^{\infty} |\pi_j| < \infty$ gibt, so daß $Z_t = \pi(B)X_t$.

Satz 17 *Sei (X_t) ARMA(p,q), und es gelte, daß ϱ und φ keine gemeinsamen Nullstellen haben. Dann ist (X_t) genau dann invertierbar, wenn φ keine Nullstellen auf der abgeschlossenen komplexen Einheits-Kreisscheibe hat. Die Koeffizienten der inversen Darstellung ergeben sich aus $\pi = \varrho/\varphi$.*

Beweis. *Notwendig.* Sei $\varphi(z) \neq 0$ für $|z| \leq 1$. Wie im Beweis zu Satz 16 gibt es eine Reihenentwicklung

$$\frac{1}{\varphi(z)} = \sum_{j=0}^{\infty} \vartheta_j z^j = \vartheta(z), \quad |z| < 1 + \varepsilon,$$

und $\sum |\vartheta_j| < \infty$. Mit Satz 15 folgt

$$\vartheta(B)\varrho(B)X_t = \vartheta(B)\varphi(B)Z_t = Z_t.$$

Also gilt die inverse Darstellung mit $\pi = \vartheta\varrho = \varrho/\varphi$.

Hinreichend. Sei (X_t) invertierbar, $Z_t = \pi(B)X_t$. Dann gilt

$$\varrho(B)Z_t = \varrho(B)\pi(B)X_t = \pi(B)\varphi(B)Z_t,$$

also $\varrho(z) = \pi(z)\varphi(z)$ für $|z| \leq 1$, also $\varphi(z) \neq 0$ für $|z| \leq 1$.

Satz 18 *Sei (X_t) ARMA(p,q), und ϱ habe keine Nullstellen auf dem komplexen Einheitskreis. Dann gibt es eine eindeutige schwach stationäre Lösung $\sum_{j=-\infty}^{\infty} \vartheta_j Z_{t-j}$, und die Koeffizienten ergeben sich aus der Laurent-Reihe*

$$\frac{\varphi(z)}{\varrho(z)} = \sum_{j=-\infty}^{\infty} \vartheta_j z^j = \vartheta(z), \quad \frac{1}{r} < |z| < r.$$

Beweis. *Existenz.* Nach Satz 15 ist $X_t = \vartheta(B)Z_t$ schwach stationär, und $\varrho(B)X_t = \varrho(B)\vartheta(B)Z_t = \varphi(B)Z_t$.

Eindeutigkeit. Sei (X_t) eine schwach stationäre Lösung von $\varrho(B)X_t = \varphi(B)Z_t$. Nach Voraussetzung gibt es ein $r > 1$, so daß

$$\frac{1}{\varrho(z)} = \sum_{j=-\infty}^{\infty} \xi_j z^j = \xi(z), \quad \frac{1}{r} < |z| < r,$$

absolut konvergent ist. Es folgt

$$\xi(B)\varrho(B)X_t = \xi(B)\varphi(B)Z_t = \xi(B)\varrho(B)\vartheta(B)Z_t,$$

also $X_t = \vartheta(B)Z_t$.

12 Hilberträume und Orthonormalsysteme

In $\mathbb{C}$ definiert $x\bar{y}$ ein inneres Produkt. Für $x = a + ib$ gilt $\|x\|^2 = x\bar{x} = (a+ib)(a-ib) = a^2 + b^2$. Sei $\mathcal{H}$ ein komplexer Vektorraum. Eine Abbildung $(\cdot,\cdot) : \mathcal{H}^2 \to \mathbb{C}$ heißt *inneres Produkt*, wenn für $x, y, z \in \mathcal{H}$ und $\alpha \in \mathbb{C}$ gilt:

$$\begin{aligned}
(x, y) &= \overline{(y, x)}, \\
(x + y, z) &= (x, z) + (y, z), \\
(\alpha x, y) &= \alpha(x, y), \\
(x, x) &\geq 0, \\
(x, x) &= 0 \quad \text{genau dann, wenn} \quad x = 0.
\end{aligned}$$

Die zugehörige *Norm* ist definiert durch $\|x\| = (x,x)^{1/2}$. Es gilt die Cauchy-Ungleichung $|(x,y)| \le \|x\|\|y\|$, mit Gleichheit genau dann, wenn $x = 0$ oder $y = 0$, oder wenn x und y proportional sind. Die Cauchy-Ungleichung impliziert, daß das innere Produkt stetig ist. Für $x, y \in \mathcal{H}$ gilt

$$\begin{aligned}
\|x+y\| &\le \|x\| + \|y\|, \\
\|\alpha x\| &= |\alpha|\|x\|, \\
\|x\| &\ge 0, \\
\|x\| &= 0 \quad \text{genau dann, wenn} \quad x = 0.
\end{aligned}$$

Ein komplexer Vektorraum $\mathcal{H}$ heißt *Hilbertraum*, wenn $\mathcal{H}$ abgeschlossen ist. Sei $\mathcal{H}$ ein (komplexer oder reeller) Hilbertraum. Sei $\mathcal{M}$ eine Teilmenge von $\mathcal{H}$. Das *orthogonale Komplement* $\mathcal{M}^\perp$ von $\mathcal{M}$ besteht aus allen $x \in \mathcal{H}$ mit $(x,y) = 0$ für $y \in \mathcal{M}$.

Satz 19 *$\mathcal{M}^\perp$ ist ein abgeschlossener linearer Teilraum von $\mathcal{H}$.*

Beweis. Offensichtlich ist $\mathcal{M}^\perp$ linear. Sei $x_n \in \mathcal{M}^\perp$ mit $\|x_n - x\| \to 0$. Dann gilt für $y \in \mathcal{M}$ mit der Cauchy-Ungleichung

$$|(x_n - x, y)| \le \|x_n - x\|\|y\| \to 0,$$

also $(x,y) = 0$, also $x \in \mathcal{M}^\perp$.

Satz 20 *Ist $\mathcal{M}$ ein abgeschlossener linearer Teilraum von $\mathcal{H}$ und $x \in \mathcal{H}$, dann gibt es ein eindeutiges $x_0 \in \mathcal{M}$ mit*

$$\|x - x_0\| = \inf_{y \in \mathcal{M}} \|x - y\|.$$

Das Element x_0 ist charakterisiert durch $x_0 \in \mathcal{M}$ und $x - x_0 \in \mathcal{M}^\perp$.

Das Element x_0 heißt die *orthogonale Projektion* von x auf $\mathcal{M}$. Im folgenden schreiben wir dafür Px oder $P_{\mathcal{M}}x$.

Beweis. *Existenz.* Setze $d = \inf_{y \in \mathcal{M}} \|x - y\|$. Es gibt $y_n \in \mathcal{M}$ mit $\|y_n - x\| \to d$. Mit der Parallelogramm-Gleichung für $y_m - x$ und $y_n - x$ gilt

$$\begin{aligned}
0 \le \|y_m - y_n\|^2 &= -4\left\|\frac{y_m + y_n}{2} - x\right\|^2 + 2(\|y_n - x\|^2 + \|y_m - x\|^2) \\
&\le -4d^2 + 2(\|y_n - x\|^2 + \|y_m - x\|^2) \to 0, \quad m, n \to \infty.
\end{aligned}$$

Nach dem Cauchy-Kriterium existiert also ein x_0 mit $\|y_n - x_0\| \to 0$.

Eindeutigkeit. Sei x_1 ebenfalls Projektion von x. Wie eben gilt

$$\begin{aligned}0 &\leq \|x_0 - x_1\|^2 = -4\left\|\frac{x_0 + x_1}{2} - x\right\|^2 + 2(\|x_0 - x\|^2 + \|x_1 - x\|^2) \\ &\leq -4d^2 + 4d^2 = 0.\end{aligned}$$

Charakterisierung. Sei $x_0 \in \mathcal{M}$ und $x - x_0 \in \mathcal{M}^\perp$. Dann gilt für $y \in \mathcal{M}$:

$$\|x - y\|^2 = \|x - x_0\|^2 + \|x_0 - y\|^2 \geq \|x - x_0\|^2.$$

Gleichheit gilt nur dann, wenn $y = x_0$. Sei umgekehrt $x_0 \in \mathcal{M}$ und $x - x_0 \notin \mathcal{M}^\perp$. Wähle $y \in \mathcal{M}$ mit $a = (x - x_0, y) \neq 0$ und $\|y\| = 1$. Definiere x_1 durch $x - x_1 = x - x_0 - ay$. (Dann ist ay die Projektion von $x - x_0$ auf den von y aufgespannten Raum.) Es gilt

$$\|x - x_1\|^2 = \|x - x_0\|^2 + |a|^2 - 2|a|^2 = \|x - x_0\|^2 - |a|^2.$$

Definition. Sei $\{x_t, t \in T\} \subset \mathcal{H}$. Der (abgeschlossene) *Spann* $\mathrm{sp}\,(x_t, t \in T)$ ist die kleinste abgeschlossene Teilmenge, die $\{x_t, t \in T\}$ enthält.

Der Spann von $z_1, \ldots, z_m$ ist die Menge aller Linearkombinationen von $z_1, \ldots, z_m$. Für $x \in \mathcal{H}$ ist also die Projektion Px auf den Spann von der Form $\alpha^\top z$ und charakterisiert durch $(x - \alpha^\top z, z_j) = 0$ für $j = 1, \ldots, m$, also $(x, z) = (z, z^\top)\alpha$, also $\alpha = (z, z^\top)^{-1}(x, z)$, falls $(z, z^\top)$ positiv definit ist.

Definition. Eine Menge $\{e_t, t \in T\} \subset \mathcal{H}$ heißt *Orthonormalsystem (ONS)*, wenn $(e_s, e_t) = \mathbf{1}(s = t)$ für $s, t \in T$ gilt.

Satz 21 *Sei $\{e_1, \ldots, e_m\}$ ein ONS und $\mathcal{M}$ sein Spann. Dann gilt für $x \in \mathcal{H}$:*

$$\begin{aligned}P_\mathcal{M} x &= \sum_{j=1}^m (x, e_j)e_j, \\ \|P_\mathcal{M} x\|^2 &= \sum_{j=1}^m |(x, e_j)|^2, \\ \|x - P_\mathcal{M} x\| &\leq \left\|x - \sum_{j=1}^m c_j e_j\right\|,\end{aligned}$$

mit Gleichheit genau dann, wenn $c_j = (x, e_j)$ für $j = 1, \ldots, m$.

Beweis. Die erste Behauptung folgt aus obiger Bemerkung mit $z_j = e_j$, die zweite durch Nachrechnen, und die dritte ist die definierende Eigenschaft der Projektion. Nach Satz 20 gilt Gleichheit genau dann, wenn

$$\sum_{j=1}^{m} c_j e_j = P_{\mathcal{M}} x = \sum_{j=1}^{m} (x, e_j) e_j.$$

Bildet man die inneren Produkte mit e_j, ergibt sich die Charakterisierung.

Aus Satz 21 folgt die *Bessel-Ungleichung*

$$\sum_{j=1}^{m} |(x, e_j)|^2 \leq \|x\|^2.$$

Definition. Ein ONS heißt *vollständig (Orthonormalbasis, ONB)*, wenn sein Spann gleich $\mathcal{H}$ ist.

Definition. Ein Hilbertraum $\mathcal{H}$ heißt *separabel*, wenn es eine abzählbare ONB gibt.

Satz 22 *Sei $\{e_j, j \in \mathbb{N}\}$ eine ONB. Dann gilt:*

1. *Die endlichen Linearkombinationen der ONB sind dicht in $\mathcal{H}$.*
2. *Für $x \in \mathcal{H}$ gilt $x = \sum (x, e_j) e_j$ und $\|x\|^2 = \sum |(x, e_j)|^2$.*
3. *Für $x, y \in \mathcal{H}$ gilt die* Parseval-Gleichung $(x, y) = \sum (x, e_j)(e_j, y)$.
4. *Es gilt $x = 0$ genau dann, wenn $(x, e_j) = 0$ für $j \in \mathbb{N}$.*

Beweis. 1. Die Menge der endlichen Linearkombinationen bildet einen linearen Raum. Sein Abschluß enthält die ONB, also auch ihren Spann $\mathcal{H}$.

2. Sei $\varepsilon > 0$. Nach 1. existiert ein m und $c_1, \ldots, c_m$ mit

$$\left\| x - \sum_{j=1}^{m} (x, e_j) e_j \right\| < \varepsilon.$$

Aus der Eigenschaft der Projektion folgt

$$\left\| x - \sum_{j=1}^{n} (x, e_j) e_j \right\| < \varepsilon, \quad n \geq m,$$

also

$$x = \sum_{j=1}^{\infty} (x, e_j) e_j.$$

Für $n \geq m$ folgt aus der letzten Ungleichung

$$\|x\|^2 \leq \varepsilon^2 + \sum_{j=1}^{\infty} |(x, e_j)|^2.$$

Aus der Bessel-Ungleichung folgt

$$\sum_{j=1}^{\infty} |(x, e_j)|^2 \leq \|x\|^2,$$

also Gleichheit.

3. Für $x, y \in \mathcal{H}$ gilt wegen der Stetigkeit des inneren Produkts

$$(x, y) = \lim_{m \to \infty} \Big(\sum_{j=1}^{m} (x, e_j) e_j, \sum_{j=1}^{m} (y, e_j) e_j \Big) = \lim_{m \to \infty} \sum_{j=1}^{m} (x, e_j)(e_j, y).$$

4. folgt aus 2.

13 Kovarianz-Funktion und Ordnung linearer Zeitreihen

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum. Sei Z ein Zufallsvektor und $\mathcal{M}(Z)$ die Menge aller Zufallsvariablen der Form $f(Z) \in L_2(P)$ für Borel-meßbares f. Dann ist $\mathcal{M}(Z)$ ein abgeschlossener und linearer Teilraum von $L_2(P)$. Sei $X \in L_2(P)$. Dann ist der bedingte Erwartungswert $E(X|Z)$ die Projektion von X auf $\mathcal{M}(Z)$. Denn $E(X|Z)$ ist in $L_2(P)$ und $\mathcal{F}(Z)$-meßbar, also in $\mathcal{M}(Z)$, und nach Definition des bedingten Erwartungswerts gilt

$$E(X - E(X|Z))\mathbf{1}_A = 0, \quad A \in \mathcal{F}(Z),$$

also

$$E(X - E(X|Z))f(Z) = 0, \quad f(Z) \in \mathcal{M}(Z).$$

Wir schreiben $E(X|Z) = E(X|\mathcal{M}(Z))$.

Definition. Sei $\mathcal{M}$ ein abgeschlossener linearer Teilraum von $L_2(P)$, der die Konstanten enthält. Wir definieren den *bedingten Erwartungswert* $E(X|\mathcal{M})$ als Projektion von X auf $\mathcal{M}$.

Definition. Ein Prozeß (X_t) heißt $MA(\infty)$, wenn er die Darstellung $X_t = \psi(B)Z_t$ mit $\sum_{j=0}^{\infty} |\psi_j| < \infty$ und weißem Rauschen (Z_t) hat.

Ein MA(q)-Prozeß $X_t = \varphi(B)Z_t$ ist MA(∞) mit $\psi_0 = 1$, $\psi_1 = \varphi_1$, ..., $\psi_q = \varphi_q$ und $\psi_j = 0$ sonst. Unter den Voraussetzungen von Satz 16 (aber i.a. nicht unter denen von Satz 17) ist ein ARMA(p,q)-Prozeß auch MA(∞).

Satz 23 *Sei (X_t) schwach stationär und zentriert mit $\gamma(s) = 0$ für $|s| > q$. Dann ist (X_t) MA(q).*

Beweis. Sei $\mathcal{M}_t = sp(X_t, X_{t-1}, \dots)$ und $\mathcal{M}_t^k = sp(X_t, X_{t-1}, \dots, X_{t-k+1})$. Setze

$$Z_t = X_t - P_{\mathcal{M}_{t-1}} X_t.$$

Dann ist $Z_t \in \mathcal{M}_t \cap \mathcal{M}_{t-1}^{\perp}$. Für $s < t$ gilt $Z_s \in \mathcal{M}_s$, also $EZ_sZ_t = 0$. Nach Satz 21,1 und 22,1 gilt

$$\|P_{\mathcal{M}_{t-1}^k} X_t - P_{\mathcal{M}_{t-1}} X_t\|_2 \to 0, \quad k \to \infty.$$

Also gilt wegen der schwachen Stationarität von (X_t):

$$\begin{aligned} \|Z_{t+1}\|_2 = \|X_{t+1} - P_{\mathcal{M}_t} X_{t+1}\|_2 &= \lim_{k\to\infty} \|X_{t+1} - P_{\mathcal{M}_t^k} X_{t+1}\|_2 \\ &= \lim_{k\to\infty} \|X_t - P_{\mathcal{M}_{t-1}^k} X_t\|_2 = \|X_t - P_{\mathcal{M}_{t-1}} X_t\|_2 = \|Z_t\|_2. \end{aligned}$$

Also ist (Z_t) ein weißes Rauschen mit $\sigma^2 = \|Z_t\|_2^2$. Es gilt

$$\mathcal{M}_{t-1} = \mathrm{sp}\,(Z_{t-1}, X_{t-2}, \dots) = \mathrm{sp}\,(Z_{t-1}, \dots Z_{t-q}, X_{t-q-1}, \dots),$$

also ist $\mathcal{M}_{t-1}$ die orthogonale Summe von $\mathcal{M}_{t-q-1}$ und $\mathrm{sp}\,(Z_{t-1}, \dots Z_{t-q})$. Wegen $\gamma(s) = 0$ für $|s| > q$ gilt $X_t \perp \mathcal{M}_{t-q-1}$, also

$$P_{\mathcal{M}_{t-1}} X_t = P_{\mathrm{sp}\,(Z_{t-1},\dots Z_{t-q})} X_t = \sigma^{-2} \sum_{j=1}^{q} E(X_t Z_{t-j}) Z_{t-j}.$$

Die Koeffizienten $\varphi_j = \sigma^{-2} E(X_t Z_{t-j})$ hängen nicht von t ab. Es ergibt sich

$$X_t = Z_t + P_{\mathcal{M}_{t-1}} X_t = \varphi(B) Z_t$$

mit $\varphi(z) = 1 + \varphi_1 z + \dots + \varphi_q z^q$.

14 Spektralverteilung stationärer Prozesse

In diesem Kapitel sei (X_t) ein *komplex*-wertiger Prozeß auf $(\Omega, \mathcal{F}, P)$ mit $E|X_t|^2 < \infty$, und er sei schwach stationär, das heißt jetzt, daß EX_t und $EX_{t+s}\bar{X}_t$ nicht von t abhängen. Die *Autokovarianz-Funktion* ist definiert durch

$$\gamma(s) = EX_{t+s}\bar{X}_t - EX_{t+s}E\bar{X}_t.$$

Es gilt $\gamma(0) \geq 0$, $|\gamma(s)| \leq \gamma(0)$ und $\gamma(s) = \bar{\gamma}(-s)$, d.h., γ ist *hermitesch.* Wie Satz 13 beweist man: Eine Funktion $\gamma : \mathbb{Z} \to \mathbb{C}$ ist Autokovarianz-Funktion einer schwach stationären (komplexen) Zeitreihe genau dann, wenn sie hermitesch und *positiv semidefinit* ist, also

$$\sum_{j,k=1}^{m} a_j \gamma(t_j - t_k)\bar{a}_k \geq 0$$

für $a_j \in \mathbb{C}$, $t_j \in \mathbb{Z}$, $m \in \mathbb{N}$.

Wir beginnen mit drei Beispielen für Spektraldarstellungen.

1. Sei A eine komplexe Zufallsvariable mit $EA = 0$ und $EA\bar{A} = \sigma^2$. Sei $\lambda \in (-\pi, \pi]$ und $X_t = Ae^{it\lambda}$. Dann gilt $EX_t = 0$ und $EX_{t+s}\bar{X}_t = \sigma^2 e^{is\lambda}$. Also ist (X_t) schwach stationär mit Autokovarianz-Funktion

$$\gamma(s) = \int e^{isu} dF(u),$$

wobei F die Verteilungsfunktion des Einpunktmaßes $\sigma^2\delta_\lambda$ ist.

2. Seien A, B unkorrelierte komplexe Zufallsvariable mit $A = \bar{B}$, $EA = EB = 0$ und $EA\bar{A} = \sigma^2$. Sei $\lambda \in (-\pi, \pi]$ und

$$X_t = Ae^{it\lambda} + Be^{-it\lambda}.$$

Dann gilt

$$\begin{aligned} X_t &= A\cos t\lambda + iA\sin t\lambda + B\cos(-t\lambda) + iB\sin(-t\lambda) \\ &= (A + B)\cos t\lambda + i(A - B)\sin t\lambda. \end{aligned}$$

Wegen $A = \bar{B}$ ist $A + B$ reell und $A - B$ imaginär, also X_t reell. Es gilt

$$EX_{t+s}X_t = \sigma^2(e^{is\lambda} + e^{-is\lambda}) = 2\sigma^2 \cos s\lambda.$$

Also ist (X_t) schwach stationär mit Autokovarianz-Funktion

$$\gamma(s) = \int e^{isu} dF(u),$$

wobei F die Verteilungsfunktion von $\sigma^2(\delta_\lambda + \delta_{-\lambda})$ ist.

3. Für $-\pi < \lambda_1 < \cdots < \lambda_m = \pi$ und unkorrelierte komplexe Zufallsvariable $A(\lambda_1)$, ..., $A(\lambda_m)$ mit $EA(\lambda_j) = 0$ und $EA(\lambda_j)\bar{A}(\lambda_j) = \sigma^2$ sei

$$X_t = \sum_{j=1}^{m} A(\lambda_j) e^{it\lambda_j}.$$

(Dieser Prozeß ist reell, wenn $A(\lambda_m)$ reell ist und $\lambda_j = -\lambda_{m-j}$ und $A(\lambda_j) = \bar{A}(\lambda_{m-j})$ für $j = 1, \ldots, m$ gilt.) Es gilt $EX_t = 0$ und

$$EX_{t+s}\bar{X}_t = \sigma^2 \sum_{j=1}^{m} e^{is\lambda_j}.$$

Also ist (X_t) schwach stationär. Seine Autokovarianz-Funktion läßt sich schreiben als $\gamma(s) = \int e^{isu} dF(u)$ mit Verteilungsfunktion

$$F(u) = \sum_{\lambda_j \leq u} \|A(\lambda_j)\|_2^2 = \sigma^2 \#\{j : \lambda_j \leq u\}.$$

Satz 24 *(Herglotz) Eine Funktion* $\gamma : \mathbb{Z} \to \mathbb{C}$ *ist positiv semidefinit genau dann, wenn*

$$\gamma(s) = \int_{(-\pi,\pi]} e^{isu} dF(u), \quad s \in \mathbb{Z},$$

mit F *rechtsstetig, nichtfallend, beschränkt,* $F(-\pi) = 0$.

Die Funktion F heißt *Spektralverteilungsfunktion* von γ; falls eine Dichte existiert, heißt sie *Spektraldichte.*

Beweis. *Notwendig.* Hat γ die obige Darstellung, so gilt

$$\begin{aligned}
\sum_{j,k=1}^{m} a_j \gamma(t_j - t_k) \bar{a}_k &= \int \sum_{j,k=1}^{m} a_j \bar{a}_k e^{iu(t_j - t_k)} dF(u) \\
&= \int \Big| \sum_{j=1}^{m} a_j e^{iut_j} \Big|^2 dF(u) \geq 0.
\end{aligned}$$

Also ist γ positiv semidefinit.

Hinreichend. Sei γ positiv semidefinit. Für $m \in \mathbb{N}$ und $u \in (-\pi, \pi]$ setze

$$f_m(u) = \frac{1}{2\pi m} \sum_{j,k=1}^{m} e^{-iju} \gamma(j-k) e^{iku} = \frac{1}{2\pi m} \sum_{|j|<m} (m - |j|) e^{-iju} \gamma(j).$$

Da γ positiv semidefinit ist, ist f_m nichtnegativ. Sei F_m die zugehörige Verteilungsfunktion. Es gilt für $|s| < m$:

$$\int e^{isu} dF_m(u) = \frac{1}{2\pi} \sum_{|j|<m} (1 - |j|/m)\gamma(j) \int_{-\pi}^{\pi} e^{i(s-j)u} du = (1 - |s|/m)\gamma(s).$$

Insbesondere gilt $F_m(\pi) = \gamma(0)$. Nach dem Satz von Helly gilt $F_m \Rightarrow F$ für eine Teilfolge, also $\int e^{isu} dF(u) = \gamma(s)$.

Eine Funktion γ mit der Darstellung aus Satz 24 ist offensichtlich hermitesch. Also ist γ Autokovarianz-Funktion eines schwach stationären Prozesses genau dann, wenn sie diese Darstellung hat. Man kann zeigen, daß F durch γ eindeutig bestimmt ist.

Satz 25 *Sei $\gamma : \mathbb{Z} \to \mathbb{C}$ mit $\sum |\gamma(s)| < \infty$. Dann gilt*

$$\gamma(s) = \int e^{isu} f(u)\, du$$

mit

$$f(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-isu} \gamma(s).$$

Beweis. Mit Fubini gilt

$$\int e^{isu} f(u)\, du = \frac{1}{2\pi} \sum_{t=-\infty}^{\infty} \gamma(t) \int_{-\pi}^{\pi} e^{i(s-t)u} du = \gamma(s).$$

Satz 26 *Sei $\gamma : \mathbb{Z} \to \mathbb{C}$ mit $\sum |\gamma(s)| < \infty$. Dann ist γ Autokovarianz-Funktion eines schwach stationären Prozesses genau dann, wenn*

$$f(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} e^{-isu} \gamma(s) \geq 0, \quad -\pi \leq u \leq \pi,$$

und f ist dann die Spektraldichte.

Beweis. *Hinreichend.* Sei γ Autokovarianz-Funktion. Dann ist γ positiv semidefinit. Es gilt

$$\begin{aligned} 0 \leq f_m(u) &= \frac{1}{2\pi m} \sum_{j,k=1}^{m} e^{-iju} \gamma(j-k) e^{iku} \\ &= \frac{1}{2\pi m} \sum_{|j|<m} (m - |j|) e^{-iju} \gamma(j) \to f(u), \quad m \to \infty. \end{aligned}$$

Aus Satz 25 folgt, daß f Spektraldichte von γ ist.

Notwendig. Sei $f \geq 0$. Nach Satz 25 gilt

$$\gamma(s) = \int e^{isu} f(u)\, du = \int_{(-\pi,\pi]} e^{isu} dF(u),$$

wenn F die Verteilungsfunktion von f bezeichnet. Nach obiger Bemerkung ist also γ eine Autokovarianz-Funktion.

Satz 27 *Sei (Y_t) schwach stationär und zentriert mit Spektralverteilungsfunktion F_Y, und sei $\sum |\vartheta_j| < \infty$. Dann ist $X_t = \vartheta(B)Y_t$ schwach stationär und zentriert mit Spektralverteilungsfunktion*

$$F_X(u) = \int_{(-\pi,u]} \Big| \sum_{j=-\infty}^{\infty} \vartheta_j e^{-ijv} \Big|^2 dF_Y(v). \quad -\pi \leq u \leq \pi,$$

Beweis. Wie in Satz 15 zeigt man, daß (X_t) schwach stationär und zentriert ist mit Autokovarianz-Funktion

$$\gamma_X(s) = \sum_{j,k=-\infty}^{\infty} \vartheta_j \bar{\vartheta}_k \gamma_Y(s-j+k).$$

Mit der Spektraldarstellung für γ_Y ergibt sich

$$\begin{aligned}
\gamma_X(s) &= \sum_{j,k=-\infty}^{\infty} \vartheta_j \bar{\vartheta}_k \int_{(-\pi,\pi]} e^{i(s-j+k)u} dF_Y(u) \\
&= \int_{(-\pi,\pi]} \sum_{j=-\infty}^{\infty} \vartheta_j e^{-iju} \sum_{k=-\infty}^{\infty} \bar{\vartheta}_k e^{iku} e^{isu} dF_Y(u) \\
&= \int_{(-\pi,\pi]} e^{isu} \Big| \sum_{j=-\infty}^{\infty} \vartheta_j e^{-iju} \Big|^2 dF_Y(u).
\end{aligned}$$

Also ist F_X die Spektralverteilungsfunktion von γ_X.

Satz 28 *(Spektraldichte einer ARMA(p,q)-Zeitreihe) Sei (Z_t) ein weißes Rauschen mit Varianz σ^2. Sei (X_t) ARMA(p,q), $\varrho(B)X_t = \varphi(B)Z_t$. Es gelte, daß die Polynome ϱ und φ keine Nullstellen auf dem komplexen Einheitskreis haben. Dann hat (X_t) die Spektraldichte*

$$f_X(u) = \frac{\sigma^2}{2\pi} \frac{|\varphi(\exp(-iu))|^2}{|\varrho(\exp(-iu))|^2}.$$

Beweis. Nach Satz 18 gibt es genau eine schwach stationäre Lösung $X_t = \vartheta(B)Z_t$ mit $\vartheta = \varphi/\varrho$, und $\sum |\vartheta_j| < \infty$. Nach Satz 27 hat (X_t) eine Spektraldichte. Satz 27 für $V_t = \varrho(B)X_t = \varphi(B)Z_t$ liefert

$$f_V(u) = |\varrho(\exp(-iu))|^2 f_X(u) = |\varphi(\exp(-iu))|^2 f_Z(u).$$

Mit $\varrho(\exp(-iu)) \neq 0$ folgt die Behauptung.

15 Spektraldarstellung stationärer Prozesse

Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $L_2(P)$ der Raum der komplexen Zufallsvariablen darauf. Bezeichne $(X, Y)_2 = EX\bar{Y}$ das innere Produkt und $\|X\|_2 = (E|X|^2)^{1/2}$ die zugehörige Norm.

Definition. Ein komplexer Prozeß $Z(u)$, $-\pi \leq u \leq \pi$, heißt Prozeß mit *orthogonalen Inkrementen (OIP)*, wenn $Z(u) \in L_{2,0}(P)$ für $-\pi \leq u \leq \pi$ und

$$(Z(v) - Z(u), Z(x) - Z(w))_2 = 0, \quad -\pi \leq u < v \leq w < x \leq \pi,$$
$$\|Z(u+t) - Z(u)\|_2 \to 0, \quad t \downarrow 0, \quad -\pi \leq u < \pi.$$

Satz 29 *Für einen OIP Z definiert*

$$F(v) - F(u) = \|Z(v) - Z(u)\|_2^2, \quad u \leq v,$$

eine Verteilungsfunktion; sie ist durch $F(-\pi) = 0$ eindeutig festgelegt.

Beweis. Insbesondere gilt $F(u) = \|Z(u) - Z(-\pi)\|_2^2$. Für $v > u$ gilt wegen der Orthogonalität der Inkremente

$$F(v) = \|Z(v) - Z(u) + Z(u) - Z(-\pi)\|_2^2 \geq F(u).$$

Die Rechtsstetigkeit von F folgt aus der von Z.

Sei Z ein OIP und F die zugehörige Verteilungsfunktion. Bezeichne $\mathcal{F}_0 \subset L_2(F)$ die Treppenfunktionen der Form

$$f = \sum_{j=0}^{m} f_j \mathbf{1}_{(u_j, u_{j+1}]}, \quad -\pi = u_0 < \cdots < u_{m+1} = \pi.$$

Setze

$$\int f\, dZ = \sum_{j=0}^{m} f_j (Z(u_{j+1}) - Z(u_j)).$$

Wie üblich zeigt man, daß der Operator $f \mapsto \int f\,dZ$ von $\mathcal{F}_0$ nach $L_2(P)$ wohldefiniert und linear ist. Wir zeigen, daß er das innere Produkt erhält. Seien

$$f = \sum_{j=0}^{m} f_j \mathbf{1}_{(u_j,u_{j+1}]}, \quad g = \sum_{j=0}^{m} g_j \mathbf{1}_{(u_j,u_{j+1}]},$$

mit einer gemeinsamen Partition dargestellt. Dann gilt

$$\begin{aligned}\Big(\int f\,dZ, \int g\,dZ\Big)_2 &= \sum f_j \bar{g}_j \|Z(u_{j+1}) - Z(u_j)\|_2^2 \\ &= \sum f_j \bar{g}_j (F(u_{j+1}) - F(u_j)) = \int f\bar{g}\,dF = (f,g)_{L_2(F)}.\end{aligned}$$

$\mathcal{F}_0$ ist dicht in den stetigen Funktionen. Die stetigen Funktionen sind dicht in $L_2(F)$. Wie üblich läßt sich also das Integral stetig auf $L_2(F)$ fortsetzen. Die Fortsetzung ist linear und erhält das innere Produkt. Aus $EZ(u) = 0$ folgt $E\int f\,dZ = 0$. Wir haben insbesondere gezeigt: Ist Z ein OIP und F die zugehörige Verteilungsfunktion, so ist der Prozeß

$$X_t = \int e^{itu} dZ(u), \quad t \in \mathbb{Z},$$

schwach stationär und zentriert mit Autokovarianz-Funktion

$$\gamma(s) = EX_{t+s}\bar{X}_t = \int e^{isu} dF(u).$$

Wir zeigen nun umgekehrt, daß jeder zentrierte schwach stationäre Prozeß eine solche Spektraldarstellung hat. Sei (X_t) ein schwach stationärer zentrierter Prozeß mit Spektralverteilungsfunktion F. Bezeichne $\mathcal{H}_0$ die Linearkombinationen der X_t, $t \in \mathbb{Z}$, und $\mathcal{K}_0$ die Linearkombinationen der $e^{it\cdot}$, $t \in \mathbb{Z}$. Das sind lineare Teilräume von $L_2(P)$ und $L_2(F)$. Setze

$$T\Big(\sum_{j=1}^{m} a_j X_{t_j}\Big) = \sum_{j=1}^{m} a_j e^{it_j\cdot}.$$

Der Operator T von $\mathcal{H}_0$ nach $\mathcal{K}_0$ ist wohldefiniert: Gelte

$$\Big\|\sum_{j=1}^{m} a_j X_{s_j} - \sum_{k=1}^{n} b_k X_{t_k}\Big\|_2 = 0.$$

Dann gilt mit Satz 24

$$\begin{aligned}
&\Big\|T\Big(\sum_{j=1}^{m} a_j X_{s_j}\Big) - T\Big(\sum_{k=1}^{n} b_k X_{t_k}\Big)\Big\|^2_{L_2(F)} \\
&= \int_{(-\pi,\pi]} \Big|\sum_{j=1}^{m} a_j e^{is_j u} - \sum_{k=1}^{n} b_k e^{it_k u}\Big|^2 dF(u) \\
&= E\Big|\sum_{j=1}^{m} a_j X_{s_j} - \sum_{k=1}^{n} b_k X_{t_k}\Big|^2 = 0.
\end{aligned}$$

Insbesondere ist T linear. Außerdem erhält T das innere Produkt:

$$\begin{aligned}
&\Big(T\Big(\sum_{j=1}^{m} a_j X_{s_j}\Big), T\Big(\sum_{k=1}^{n} b_k X_{t_k}\Big)\Big)_{L_2(F)} \\
&= \sum_{j=1}^{m}\sum_{k=1}^{n} a_j \bar{b}_k \big(e^{is_j\cdot}, e^{it_k\cdot}\big)_{L_2(F)} \\
&= \sum_{j=1}^{m}\sum_{k=1}^{n} a_j \bar{b}_k \int_{(-\pi,\pi]} e^{i(s_j-t_k)u} dF(u) \\
&= \sum_{j=1}^{m}\sum_{k=1}^{n} a_j \bar{b}_k (X_{s_j}, X_{t_k})_2 = \Big(\sum_{j=1}^{m} a_j X_{s_j}, \sum_{k=1}^{n} b_k X_{t_k}\Big)_2.
\end{aligned}$$

Also ist T ein Isomorphismus von $\mathcal{H}_0$ auf $\mathcal{K}_0$.

$\mathcal{K}_0$ ist dicht bezüglich der sup-Norm in den stetigen Funktionen $f : [\pi,\pi] \to \mathbb{R}$ mit $f(-\pi) = f(\pi)$, also dicht in $L_2(F)$. Bezeichne $\mathcal{H}$ den Abschluß von $\mathcal{H}_0$ in $L_2(P)$. Der Operator T läßt sich wie folgt stetig zu einem Isomorphismus von $\mathcal{H}$ auf $L_2(F)$ fortsetzen. Sei $Y \in \mathcal{H}$ und $Y_n \in \mathcal{H}_0$ mit $\|Y_n - Y\|_2 \to 0$. Dann ist Y_n Cauchy. Weil T ein Isomorphismus ist, ist TY_n Cauchy in $L_2(F)$. Definiere TY als Limes von TY_n. Die Fortsetzung ist wohldefiniert, linear und erhält das innere Produkt. Wir haben also folgendes Resultat.

Satz 30 *Sei (X_t) ein schwach stationärer und zentrierter Prozeß mit Spektralverteilungsfunktion F. Dann existiert genau ein Isomorphismus T von* $\mathrm{sp}\,(X_t, t \in \mathbb{Z})$ *auf $L_2(F)$ mit*

$$TX_t = e^{it\cdot}, \quad t \in \mathbb{Z}.$$

Satz 31 *Sei (X_t) ein schwach stationärer und zentrierter Prozeß mit Spektralverteilungsfunktion F. Der Prozeß*

$$Z(u) = T^{-1}\mathbf{1}_{(-\pi,u]}, \quad -\pi \le u \le \pi,$$

ist ein OIP; die zugehörige Verteilungsfunktion ist F.

Beweis. Es gilt $Z(u) \in \mathrm{sp}\,(X_t, t \in \mathbb{Z}) \subset L_{2,0}(P)$. Für $-\pi \le u < v \le w < x \le \pi$ gilt mit Satz 30:

$$\begin{aligned}(Z(v) - Z(u), Z(x) - Z(w))_2 &= (\mathbf{1}_{(u,v]}, \mathbf{1}_{(w,x]})_{L_2(F)} \\ &= \int_{(-\pi,\pi]} \mathbf{1}_{(u,v]}\mathbf{1}_{(w,x]}\, dF = 0.\end{aligned}$$

Für $-\pi \le u \le v \le \pi$ gilt mit Satz 30:

$$\|Z(v) - Z(u)\|_2^2 = \|\mathbf{1}_{(u,v]}\|_{L_2(F)}^2 = F(v) - F(u).$$

Also hat Z Verteilungsfunktion F und ist insbesondere rechtsstetig.

Satz 32 *Sei (X_t) ein schwach stationärer und zentrierter Prozeß mit Spektralverteilungsfunktion F. Dann existiert ein OIP Z mit*

$$\begin{aligned}\|Z(u) - Z(-\pi)\|_2^2 &= F(u), \quad -\pi \le u \le \pi, \\ X_t &= \int_{(-\pi,\pi]} e^{itu} dZ(u), \quad t \in \mathbb{Z}.\end{aligned}$$

Beweis. Sei Z der Prozeß aus Satz 31. Für $f \in \mathcal{F}_0$ gilt

$$\begin{aligned}\int f\, dZ &= \sum f_j(Z(u_{j+1}) - Z(u_j)) \\ &= \sum f_j\big(T^{-1}\mathbf{1}_{(-\pi,u_{j+1}]} - T^{-1}\mathbf{1}_{(-\pi,u_j]}\big) = T^{-1}f.\end{aligned}$$

Das gilt auch für $f \in L_2(F)$, da $\mathcal{F}_0$ dicht in $L_2(F)$ ist. Mit Satz 30 folgt

$$X_t = T^{-1}e^{it\cdot} = \int_{(-\pi,\pi]} e^{it\cdot} dZ.$$

16 Vorhersage stationärer Prozesse

Sei (X_t) reellwertig, schwach stationär und zentriert mit Autokovarianz-Funktion γ. Wir beobachten $X_1, \ldots, X_n$ und wollen X_{n+1} vorhersagen. Setze

$$\mathcal{M}_n = \mathrm{sp}\,(X_n, \ldots, X_1).$$

Der *Prädiktor* von X_{n+1} ist die Projektion in $L_2(P)$ von X_{n+1} auf $\mathcal{M}_n$, also

$$P_{\mathcal{M}_n} X_{n+1} = E(X_{n+1} | X_n, \ldots, X_1).$$

Mit $\mathbf{X}_n = (X_n, \ldots, X_1)$ und untenstehendem Lemma 10 ergibt sich nach Kapitel 12:

$$P_{\mathcal{M}_n} X_{n+1} = \alpha_n^\top \mathbf{X}_n$$

mit

$$\alpha_n = (\mathbf{X}_n, \mathbf{X}_n^\top)_2^{-1} (X_{n+1}, \mathbf{X}_n)_2 = \Gamma_n^{-1} \gamma_n,$$

wobei $\Gamma_{nij} = \gamma(i-j)$ und $\gamma_{nj} = \gamma(j)$ für $i, j = 1, \ldots, n$ sei. Der mittlere quadratische Fehler ist

$$E(X_{n+1} - P_{\mathcal{M}_n} X_{n+1})^2 = E(X_{n+1} - \gamma_n^\top \Gamma_n^{-1} \mathbf{X}_n)^2 = \gamma(0) - \gamma_n^\top \Gamma_n^{-1} \gamma_n.$$

Lemma 10 *Gilt $\gamma(0) > 0$ und $\gamma(s) \to 0$ für $s \to \infty$, so ist Γ_n positiv definit für $n \in \mathbb{N}$.*

Beweis. *Indirekt.* Sei Γ_n singulär. Dann existiert ein r, so daß Γ_r nicht singulär ist und $X_{r+1} = \sum_{j=1}^r a_j X_j$, wegen der Stationarität also $X_{r+s} = \sum_{j=1}^r a_j X_{j+s-1}$. Für $n \geq r$ gilt also $X_n = a_n^\top \mathbf{X}_r$ mit $a_n = (a_{n1}, \ldots, a_{nr})^\top$ und $\mathbf{X}_r = (X_1, \ldots, X_r)^\top$, also $\gamma(0) = a_n^\top \Gamma_r a_n$. Schreibe $\Gamma_r = P \Lambda P^\top$ mit $PP^\top = I_r$ und $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_r)$ mit $0 < \lambda_1 \leq \cdots \leq \lambda_r$. Es folgt

$$\gamma(0) \geq \lambda_1 a_n^\top P P^\top a_n = \lambda_1 \sum_{j=1}^r a_{nj}^2.$$

Andererseits gilt

$$\gamma(0) = \mathrm{Cov}\,(X_n, a_n^\top \mathbf{X}_r) \leq \sum_{j=1}^r a_{nj} \gamma(n-j).$$

Da die a_{nj} beschränkt sind und $\gamma(s) \to 0$ für $s \to \infty$, ergibt sich ein Widerspruch gegen $\gamma(0) > 0$.

Der *(m-Schritt-)Prädiktor* von X_{n+m} ist analog

$$P_{\mathcal{M}_n} X_{n+m} = E(X_{n+m}|X_n, \dots, X_1) = \alpha_n^\top \mathbf{X}_n$$

mit

$$\alpha_n = (\mathbf{X}_n, \mathbf{X}_n^\top)_2^{-1}(X_{n+m}, \mathbf{X}_n)_2 = \Gamma_n^{-1}\gamma_n^{(m)},$$

jetzt mit $\gamma_{nj}^{(m)} = \gamma(j+m-1)$.

17 Schätzen des Mittelwerts und der Autokovarianz-Funktion

Sei (X_t) reellwertig und schwach stationär mit Mittelwert μ und Autokovarianz-Funktion γ. Ein erwartungstreuer Schätzer für μ ist das *Stichprobenmittel*

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Satz 33 *Es gilt*

$$\begin{aligned} &\operatorname{Var}\hat{\mu} \to 0, \quad \textit{wenn} \quad \gamma(s) \to 0, \\ &n\operatorname{Var}\hat{\mu} \to \sum_{s\in\mathbb{Z}} \gamma(s), \quad \textit{wenn} \quad \sum_{s\in\mathbb{Z}} |\gamma(s)| < \infty. \end{aligned}$$

Beweis. Es gilt

$$\operatorname{Var}\hat{\mu} = \frac{1}{n}\sum_{i,j=1}^n \operatorname{Cov}(X_i, X_j) = \sum_{|s|<n} (1-|s|/n)\gamma(s) \le \sum_{|s|<n} |\gamma(s)|.$$

Daraus folgt die erste Behauptung. Gilt $\sum_{s\in\mathbb{Z}} |\gamma(s)| < \infty$, so liefert der Satz über die dominierte Konvergenz die zweite Behauptung.

Satz 34 *Sei $X_t = \mu + \vartheta(B)\varepsilon_t$, und seien ε_t i.i.d. und zentriert mit Varianz σ^2. Es gelte $\sum_{j\in\mathbb{Z}} |\vartheta_j| < \infty$ und $\sum_{j\in\mathbb{Z}} \vartheta_j \neq 0$. Dann gilt*

$$n^{1/2}(\hat{\mu} - \mu) \Rightarrow N_{0,\tau^2}$$

mit

$$\tau^2 = \sum_{s\in\mathbb{Z}} \gamma(s) = \sigma^2\Big(\sum_{s\in\mathbb{Z}} \vartheta_s\Big)^2.$$

Die Form der Varianz ergibt sich aus Satz 15. Ein Schätzer für $\gamma(s)$ für $0 \le s < n$ ist

$$\hat{\gamma}(s) = \frac{1}{n} \sum_{i=1}^{n-s} (X_{i+s} - \hat{\mu})(X_i - \hat{\mu}).$$

Setze $\hat{\gamma}(-s) = \hat{\gamma}(s)$ und $\hat{\Gamma}_{nij} = \hat{\gamma}(i-j)$ für $i, j = 1, \dots, n$. Es gilt $\hat{\Gamma}_n = TT^\top/n$ mit $n \times 2n$ Matrix

$$T = \begin{pmatrix} 0 & \cdots & X_1 - \hat{\mu} & \cdots & X_n - \hat{\mu} \\ \vdots & & & & \vdots \\ 0 & X_1 - \hat{\mu} & \cdots & X_n - \hat{\mu} & 0 \end{pmatrix}.$$

Also ist $\hat{\Gamma}_n$ positiv semidefinit. Das ist der Grund, warum mit $\hat{\gamma}(s)$ mit $1/n$ statt $1/(n-s)$ normiert wird.

Ein Schätzer für die Autokorrelation $\varrho(s) = \gamma(s)/\gamma(0)$ ist

$$\hat{\varrho}(s) = \hat{\gamma}(s)/\hat{\gamma}(0).$$

Satz 35 *Sei $X_t = \mu + \vartheta(B)\varepsilon_t$, und seien ε_t i.i.d. und zentriert mit Varianz σ^2 und endlichem vierten Moment. Es gelte $\sum_{j\in\mathbb{Z}} |\vartheta_j| < \infty$. Setze $\varrho_s = \varrho(1), \dots, \varrho(s))^\top$ und $\hat{\varrho}_s = (\hat{\varrho}(1), \dots, \hat{\varrho}(s))^\top$. Dann gilt*

$$n^{1/2}(\hat{\varrho}_s - \varrho_s) \Rightarrow N_{0,W}$$

mit

$$\begin{aligned} W_{ij} = \sum_{s\in\mathbb{Z}} \big(&\varrho(s+i)\varrho(s+j) + \varrho(s-i)\varrho(s+j) \\ &+ 2\varrho(i)\varrho(j)\varrho^2(s) - 2\varrho(i)\varrho(s)\varrho(s+j) - 2\varrho(j)\varrho(s)\varrho(s+i)\big). \end{aligned}$$

Beweis. Brockwell und Davis (1991), Section 7.3.

18 Schätzer für die Parameter von AR(p)-Prozessen

Sei $\varrho(z) = 1 - \varrho_1 z - \cdots - \varrho_p z^p$ und (X_t) der AR(p)-Prozeß $\varrho(B)X_t = Z_t$ mit weißem Rauschen (Z_t) mit Varianz σ^2. Bezeichne γ die Autokovarianz-Funktion von (X_t). Es gelte $\varrho(z) \ne 0$ für $|z| \le 1$. Nach Satz 16 ist (X_t) kausal und hat die MA(∞)-Darstellung $X_t = \psi(B)Z_t$ mit $\psi = 1/\varrho$. Insbesondere gilt $EX_{t-j}Z_t = 0$ für $j \ge 1$, also die *Yule–Walker-Gleichungen*

$$\Gamma_p \boldsymbol{\varrho} = \gamma_p, \quad \sigma^2 = \gamma(0) - \boldsymbol{\varrho}^\top \gamma_p.$$

mit $\boldsymbol{\varrho} = (\varrho_1, \ldots, \varrho_p)^\top$ und $\gamma_p = (\gamma(1), \ldots, \gamma(p))^\top$. Daraus ergeben sich die *Yule–Walker-Schätzgleichungen* für $\boldsymbol{\varrho}$ und σ^2,

$$\hat{\Gamma}_p \hat{\boldsymbol{\varrho}} = \hat{\gamma}_p, \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\varrho}}^\top \hat{\gamma}_p,$$

also mit Lemma 10 die *Yule–Walker-Schätzer*

$$\hat{\boldsymbol{\varrho}} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p, \quad \hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\gamma}_p^\top \hat{\Gamma}_p^{-1} \hat{\gamma}_p.$$

Satz 36 *Sei (X_t) der AR(p)-Prozeß $\varrho(B)X_t = \varepsilon_t$, und seien ε_t i.i.d. und zentriert mit Varianz σ^2. Gelte $\varrho(z) \neq 0$ für $|z| \leq 1$. Dann gilt*

$$n^{1/2}(\hat{\boldsymbol{\varrho}} - \boldsymbol{\varrho}) \Rightarrow N_{0,\sigma^2\Gamma_p^{-1}} \quad \textit{und} \quad \hat{\sigma}^2 \to \sigma^2 \quad \textit{in Wahrscheinlichkeit.}$$

Beweis. Brockwell und Davis (1991), Section 8.10.

19 Schätzen der Spektraldichte stationärer Prozesse

Sei (X_t) ein komplexer schwach stationärer Prozeß mit Mittelwert μ und absolut summierbarer Autokovarianz-Funktion γ. Nach Satz 26 ist die Spektraldichte

$$f(u) = \frac{1}{2\pi} \sum_{s \in \mathbb{Z}} e^{-isu} \gamma(s).$$

Ein naheliegender Schätzer ergibt sich, wenn γ durch die empirische Autokovarianz-Funktion $\hat{\gamma}$ aus Kapitel 17 ersetzt wird (für $|s| < n$, und für eine endliche Anzahl von Frequenzen u). Das resultierende sogenannte Periodogramm ist jedoch nicht konsistent; erst eine geeignete Glättung führt zu einem konsistenten Schätzer. Die Beweise der folgenden Sätze finden sich in Brockwell und Davis (1991), Chapter 10.

Die *Fourier-Frequenzen* sind $\omega_r = 2\pi r/n$ für r in

$$F_n = \{r \in \mathbb{Z} : \omega_r \in (-\pi, \pi]\} = \{-[(n-1)/2], \ldots, [n/2]\}.$$

Setze

$$e_r = n^{-1/2} \left(e^{i\omega_r}, \ldots, e^{in\omega_r}\right)^\top, \quad r \in F_n.$$

Für $u, v \in \mathbb{C}^n$ setze

$$(u, v) = \sum_{j=1}^n u_j \bar{v}_j, \quad \|u\| = (u, u)^{1/2}.$$

Es gilt

$$(e_r, e_r) = \frac{1}{n}\sum_{j=1}^{n} e^{ij(\omega_r - \omega_r)} = 1$$

und, für $r \neq s$,

$$(e_r, e_s) = \frac{1}{n} e^{i(\omega_r - \omega_s)} \frac{1 - e^{in(\omega_r - \omega_s)}}{1 - e^{i(\omega_r - \omega_s)}} = 0.$$

Also ist $\{e_r, r \in F_n\}$ eine ONB für $\mathbb{C}^n$. Jedes $x \in \mathbb{C}^n$ hat die Darstellung

$$x = \sum_{r=1}^{n} a_r e_r$$

mit

$$a_r = (x, e_r) = n^{-1/2} \sum_{j=1}^{n} x_j e^{-ij\omega_r}.$$

Die Folge a_r, $r \in F_n$, heißt *diskrete Fouriertransformation* von x. Das *Periodogramm* von x ist definiert durch

$$I(\omega_r) = |a_r|^2 = |(x, e_r)|^2 = \left| n^{-1/2} \sum_{j=1}^{n} x_j e^{-ij\omega_r} \right|^2.$$

Es gilt

$$\|x\|^2 = \sum_{r=1}^{n} |(x, e_r)|^2 = \sum_{r=1}^{n} I(\omega_r).$$

Setze

$$\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j, \quad \hat{\gamma}(s) = \frac{1}{n}\sum_{j=1}^{n-s}(X_{j+s} - \hat{\mu})(\bar{X}_j - \bar{\hat{\mu}}), \quad s \geq 0.$$

Weil γ hermitesch ist, setzen wir $\hat{\gamma}(-s) = \overline{\hat{\gamma}(s)}$. Es gilt $I(0) = n|\hat{\mu}|^2$.

Lemma 11 *Für $\omega_r \neq 0$ gilt*

$$I(\omega_r) = \sum_{|s|<n} e^{-is\omega_r} \hat{\gamma}(s).$$

Beweis. Nach Definition gilt

$$I(\omega_r) = \frac{1}{n} \sum_{j,k=1}^{n} X_j e^{-ij\omega_r} \bar{X}_k e^{ik\omega_r}.$$

Für $\omega_r \neq 0$ gilt $\sum_{j=1}^{n} e^{-ij\omega_r} = \sum_{k=1}^{n} e^{ik\omega_r} = 0$, also

$$I(\omega_r) = \frac{1}{n} \sum_{j,k=1}^{n} (X_j - \hat{\mu})(\bar{X}_k - \bar{\hat{\mu}}) e^{-i(j-k)\omega_r} = \sum_{|s|<n} e^{-is\omega_r} \hat{\gamma}(s).$$

Im folgenden sei (X_t) *reellwertig*. Der Beweis von Lemma 11 zeigt, daß wir die Zentrierung mit $\hat{\mu}$ auch durch die mit μ ersetzen dürfen. Es ergibt sich die Darstellung

$$I(\omega_r) = \sum_{|s|<n} e^{-is\omega_r} \frac{1}{n} \sum_{j=1}^{n-|s|} (X_{j+s} - \mu)(X_j - \mu).$$

Wir setzen das Periodogramm wie folgt auf $[-\pi, \pi]$ fort. Für $u \geq 0$ sei u^* der Fourier-Koeffizient ω_r mit $\omega_r - \pi/n < u \leq \omega_r + \pi/n$. Setze $I(u) = I(u^*)$ und $I(-u) = I(u)$. Das Periodogramm ist asymptotisch erwartungstreu.

Lemma 12 *Sei (X_t) schwach stationär mit Mittelwert μ und absolut summierbarer Autokovarianz-Funktion γ. Dann gilt*

$$\begin{aligned} EI(0) - n\mu^2 &\to 2\pi f(0), \\ EI(u) &\to 2\pi f(u), \quad u \neq 0. \end{aligned}$$

Beweis. Nach Satz 33 und 26 gilt

$$EI(0) - n\mu^2 = n \operatorname{Var} \hat{\mu} \to \sum_{s \in \mathbb{Z}} \gamma(s) = 2\pi f(0).$$

Für $u > 0$ und n hinreichend groß gilt $u^* > 0$. Also gilt mit Satz 25:

$$\begin{aligned} EI(u) = EI(u^*) &= \sum_{|s|<n} e^{-isu^*} \frac{1}{n} \sum_{j=1}^{n-|s|} E(X_{j+s} - \mu)(X_j - \mu) \\ &= \sum_{|s|<n} e^{-isu^*} (1 - |s|/n) \gamma(s) \to \sum_{s \in \mathbb{Z}} e^{-isu} \gamma(s) = 2\pi f(u). \end{aligned}$$

Nach Satz 27 hat ein linearer Prozeß $X_t = \vartheta(B) Z_t$ mit absolut summierbaren Koeffizienten die Spektraldichte $f_X(u) = |\vartheta(e^{-iu})|^2 f_Z(u)$. Ein ähnlicher Zusammenhang gilt näherungsweise für Periodogramme.

Satz 37 *Sei* $X_t = \vartheta(B)\varepsilon_t$, *und seien* ε_t *i.i.d. und zentriert mit endlicher Varianz. Es gelte* $\sum_{j\in\mathbb{Z}} |\vartheta_j| < \infty$. *Für die Periodogramme* I_X *und* I_ε *setze*

$$T_r = I_X(\omega_r) - |\vartheta(e^{-i\omega_r})|^2 I_\varepsilon(\omega_r).$$

Dann gilt $\max_{r\in F_n} E|T_r| \to 0$. *Gilt zusätzlich* $\sum_{j\in\mathbb{Z}} |j|^{1/2}|\vartheta_j| < \infty$, *so gilt* $\max_{r\in F_n} E|T_r|^2 = O(n^{-1})$.

Das Periodogramm ist asymptotisch erwartungstreu, aber die Streuung geht asymptotisch nicht gegen 0.

Satz 38 *Unter den Voraussetzungen von Satz 37 gilt: Ist die Spektraldichte positiv und* $0 < u_1 < \cdots < u_m < \pi$, *dann konvergiert* $(I(u_1), \ldots, I(u_m))^\top$ *in Verteilung gegen einen Vektor unabhängiger exponentialverteilter Zufallsvariablen mit Mittelwertvektor* $2\pi(f(u_1), \ldots, f(u_m))^\top$.

Unter der Zusatzvoraussetzung von Satz 37 gilt

$$\begin{aligned}
\operatorname{Var} I(u) &= 2(2\pi)^2 f^2(u) + O(n^{-1/2}), \quad u = 0, \pi, \\
\operatorname{Var} I(u) &= (2\pi)^2 f^2(u) + O(n^{-1/2}), \quad u \in (0, \pi), \\
\operatorname{Cov}(I(u), I(v)) &= O(n^{-1}), \quad |\omega_r| \neq |\omega_s|,
\end{aligned}$$

gleichmäßig in r, s.

Nach Satz 38 sind $I(\omega_r)$, $r = 1, \ldots, [(n-1)/2]$, asymptotisch unabhängig mit Mittelwerten $2\pi f(\omega_r)$. Aus der Darstellung der Spektraldichte f in Satz 25 für schwach stationäre Prozesse mit absolut summierbarer Autokovarianz-Funktion folgt, daß f stetig ist. Wir erwarten also, daß ein “Kernschätzer” aus den “Beobachtungen” $I(\omega_r)$ konsistent ist. Setze

$$\hat{f}(\omega_r) = \frac{1}{2\pi} \sum_{|s|\le S} w(s) I(\omega_{r+s})$$

mit $S \to \infty$, $S/n \to 0$ und

$$w(s) = w(-s), \quad w(s) \ge 0, \quad \sum_{|s|\le S} w(s) = 1, \quad \sum_{|s|\le S} w^2(s) \to 0.$$

Der *diskrete Spektraldichteschätzer* ist $\hat{f}(u) = \hat{f}(u^*)$.

Satz 39 *Unter den Voraussetzungen und der Zusatzvoraussetzung von Satz 37 gilt:*

$$E\hat{f}(u) \to f(u), \quad u \in [0, \pi],$$

$$\frac{1}{\sum_{|s|\le S} w^2(s)} \operatorname{Cov}(\hat{f}(u), \hat{f}(v)) \to \begin{cases} 2f^2(u), & u = v \in \{0, \pi\}, \\ f^2(u), & u = v \in (0, \pi), \\ 0, & u \neq v. \end{cases}$$

20 Effiziente Schätzer im AR(1)-Modell

Sei $X_t = \vartheta X_{t-1} + \varepsilon_t$ ein AR(1)-Prozeß, und seien ε_t i.i.d. und zentriert mit endlicher Varianz σ^2 und positiver Dichte f. Diese Zeitreihe ist eine Markov-Kette (der Ordnung 1) mit Übergangsverteilung

$$Q(x, dy) = f(y - \vartheta x)\, dy.$$

Die Kette ist irreduzibel und aperiodisch.

Im folgenden gelte $|\vartheta| < 1$. Dann ist die Kette positiv Harris-rekurrent. Es gilt $1 - \vartheta z \neq 0$ für $|z| \leq 1$. Nach Satz 16 hat (X_t) die kausale MA(∞)-Darstellung

$$X_t = \sum_{j=0}^{\infty} \vartheta^j \varepsilon_{t-j}.$$

Das liefert eine Darstellung für die stationäre Verteilung π von X_t. Außerdem ergibt sich $EX_t = 0$ und die Autokovarianz-Funktion von (X_t) für $s \geq 0$ als

$$\gamma(s) = \operatorname{Cov}(X_{t+s}, X_t) = \sum_{j,k=0}^{\infty} \vartheta^j \vartheta^k E\varepsilon_{t+s-j}\varepsilon_{t-k} = \sigma^2 \sum_{j=0}^{\infty} \vartheta^{2j+s} = \vartheta^s \gamma(0).$$

Das folgt natürlich auch aus dem allgemeinen Satz 15 oder durch Einsetzen der AR(1)-Darstellung. Die Autokovarianz-Funktion ist absolut summierbar.

Seien $X_0, \ldots, X_n$ Beobachtungen des AR(1)-Prozesses. Im folgenden sehen wir uns die Schätzer aus den Kapiteln 16–18 für diesen Spezialfall an.

Vorhersage. Sei $\mathcal{M}_n = \operatorname{sp}(X_n, \ldots, X_1)$. Der Prädiktor von X_{n+1} ist die Projektion

$$P_{\mathcal{M}_n} X_{n+1} = E(X_{n+1}|X_n) = E(\vartheta X_n + \varepsilon_{n+1}|X_n) = \vartheta X_n.$$

Für die Ordnung 2 ergibt sich

$$P_{\mathcal{M}_n} X_{n+2} = P_{\mathcal{M}_n} P_{\mathcal{M}_{n+1}} X_{n+2} = \vartheta P_{\mathcal{M}_n} X_{n+1} = \vartheta^2 X_n,$$

und für die Ordnung m ergibt sich $P_{\mathcal{M}_n} X_{n+m} = \vartheta^m X_n$. Das muß mit Kapitel 16 übereinstimmen. Dort erhalten wir

$$\Gamma_n = \gamma(0) \begin{pmatrix} 1 & \vartheta & \cdots & \vartheta^{n-1} \\ \vdots & \vdots & & \vdots \\ \vartheta^{n-1} & \vartheta^{n-2} & \cdots & 1 \end{pmatrix}, \qquad \gamma_n^{(m)} = \begin{pmatrix} \vartheta^m \\ \vdots \\ \vartheta^{m+n-1} \end{pmatrix}.$$

Für $\alpha_n = (\vartheta^m, 0, \dots, 0)^\top$ ergibt sich $\Gamma_n \alpha_n = \gamma_n^{(m)}$, also tatsächlich

$$P_{\mathcal{M}_n} X_{n+m} = \alpha_n^\top \mathbf{X}_n = \vartheta^m X_n.$$

Schätzen der Autokovarianz-Funktion. Wegen $EX_t = 0$ ist ein Schätzer für $\gamma(s)$ für $0 \le s < n$ gegeben durch

$$\hat{\gamma}(s) = \frac{1}{n} \sum_{i=0}^{n-s} X_{i+s} X_i.$$

Es gilt

$$\begin{aligned}\frac{1}{n} \sum_{i=0}^{n-s} (X_{i+s} - \hat{\mu})(X_i - \hat{\mu}) &= \hat{\gamma}(s) - \hat{\mu} \frac{1}{n} \sum_{i=0} (X_{i+s} + X_i) + \frac{n-s+1}{n} \hat{\mu}^2 \\ &= \hat{\gamma}(s) + O_p(n^{-1}).\end{aligned}$$

Der Schätzer ist also äquivalent zu dem aus Kapitel 16.

Schätzer für ϑ. Ein Schätzer für ϑ ist der *Kleinste-Quadrate-Schätzer (LSE)*, das Minimum von $\sum_{i=1}^n (X_i - \vartheta X_{i-1})^2$, also die Lösung der Martingal-Schätzgleichung $\sum_{i=1}^n X_{i-1}(X_i - \vartheta X_{i-1}) = 0$, nämlich

$$\hat{\vartheta}_{LS} = \frac{\sum_{i=1}^n X_{i-1} X_i}{\sum_{i=1}^n X_{i-1}^2}.$$

Das ist der Yule–Walker-Schätzer $\hat{\gamma}(1)/\hat{\gamma}(0)$ aus Kapitel 18. Es gilt die Martingal-Approximation

$$n^{1/2}(\hat{\vartheta}_{LS} - \vartheta) = \frac{n^{-1/2} \sum_{i=1}^n X_{i-1}\varepsilon_i}{\frac{1}{n} \sum_{i=1}^n X_{i-1}^2} = \gamma(0)^{-1} n^{-1/2} \sum_{i=1}^n X_{i-1}\varepsilon_i + o_p(1).$$

Aus Satz 7 oder Satz 11 folgt

$$n^{1/2}(\hat{\vartheta}_{LS} - \vartheta) \Rightarrow N_{0, \sigma^2/\gamma(0)}.$$

Das ist das Resultat von Satz 36 für $p = 1$.

Schätzer für σ^2. Die kausale Darstellung und die AR(1)-Darstellung ergeben

$$\sigma^2 = E\varepsilon_t^2 = EX_t\varepsilon_t = EX_t^2 - \vartheta EX_t X_{t-1} = \gamma(0) - \vartheta\gamma(1).$$

Ein Schätzer für σ^2 ist

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\vartheta}_{LS}\hat{\gamma}(1) = \hat{\gamma}(0) - \hat{\gamma}(1)^2/\hat{\gamma}(0).$$

Das ist der Yule–Walker-Schätzer aus Kapitel 18.

Lokale asymptotische Normalität. Der AR(1)-Prozeß ist mit (ϑ, f) parametrisiert. Setze $\vartheta_{nt} = \vartheta + n^{-1/2}t$ für $t \in \mathbb{R}$ und $f_{nu}(x) = f(x)(1 + n^{-1/2}u(x))$ für u aus der Menge U der beschränkten Funktionen in $L_{2,0}(f)$ mit $E\varepsilon u(\varepsilon) = 0$. Der lokale Parameterraum ist dann $\mathbb{R} \times U$. Unter (ϑ_{nt}, f_{nu}) hat die Kette die Übergangsverteilung

$$Q_{ntu}(x, dy) = f_{nu}(y - \vartheta_{nt}x)\, dy.$$

Wir nehmen an, daß f *endliche Fisher-Information* hat. Das heißt, f ist absolut stetig, und die f.s. Ableitung f' erfüllt $I = E\dot{\ell}^2(\varepsilon) < \infty$. (Die Fisher-Information ist die für die Lageparameter-Familie von Dichten $f_\vartheta(x) = f(x-\vartheta)$.) Dann ist Q_{ntu} Hellinger-differenzierbar mit Ableitung

$$D(t, u)(x, y) = tx\dot{\ell}(y - \vartheta x) + u(y - \vartheta x)$$

in

$$H = \{h \in L_{2,0}(\pi \otimes Q) : Qh = 0\},$$

denn

$$\int f(y - \vartheta x)x\dot{\ell}(y - \vartheta x)\, dy = x \int f(y)\dot{\ell}(y)\, dy = xE\dot{\ell}(\varepsilon) = 0,$$
$$\int f(y - \vartheta x)u(y - \vartheta x)\, dy = \int f(y)u(y)\, dy = Eu(\varepsilon) = 0.$$

Die Kette ist geometrisch V-gleichmäßig ergodisch für $V(x) = (1 + |x|)^2$, also stark stabil, also gilt für die stationären Verteilungen

$$\|\pi_{ntu} - \pi\|_V \to 0.$$

Sei P_n die Verteilung von $(X_0, \dots, X_n)$ unter (ϑ, f) und P_{ntu} die unter (ϑ_{nt}, f_{nu}). Setze $\Lambda_{ntu} = \log dP_{ntu}/dP_n$. Nach Satz 9 gilt LAN,

$$\Lambda_{ntu} = n^{-1/2} \sum_{i=1}^{n} D(t, u)(X_{i-1}, X_i) - \frac{1}{2}\|D(t, u)\|_2^2 + o_p(1),$$

mit

$$D(t, u)(X_{i-1}, X_i) = tX_{i-1}\dot{\ell}(\varepsilon_i) + u(\varepsilon_i).$$

Es gilt

$$EX_{i-1}\dot{\ell}(\varepsilon_i)u(\varepsilon_i) = EX_{i-1}E\dot{\ell}(\varepsilon_i)u(\varepsilon_i) = 0.$$

Also sind $X_{i-1}\dot{\ell}(\varepsilon_i)$ und $u(\varepsilon_i)$ orthogonal, und $D(\mathbb{R}\times U)$ ist eine orthogonale Summe aus den Vielfachen von $x\dot{\ell}(y-\vartheta x)$ und den Funktionen $u(y-\vartheta x)$ mit $u \in U$. Insbesondere zerfällt die LAN-Norm in eine Summe von Normen

$$\|D(t,u)\|_2^2 = t^2 EX^2 E\dot{\ell}^2(\varepsilon) + Eu^2(\varepsilon) = t^2\gamma(0)I + \|u\|_f^2,$$

wenn wir für die Norm auf $L_{2,0}(f)$ schreiben:

$$\|u\|_f^2 = \int u^2(x)f(x)\,dx = Eu^2(\varepsilon).$$

Das zur LAN-Norm gehörige innere Produkt zerfällt ebenfalls:

$$(D(t,u), D(t',u')) = tt'\gamma(0)I + (u,u')_f.$$

Adaptivität. Sei κ ein reellwertiges Funktional des Parameters (ϑ, f). Der kanonische Gradient von κ ist im Abschluß von $D(\mathbb{R}\times U)$, also von der Form

$$g_0(x,y) = t_0 x\dot{\ell}(y-\vartheta x) + u_0(y-\vartheta x)$$

mit $t_0 \in \mathbb{R}$, $u_0 \in U^-$ und

$$n^{1/2}(\kappa(\vartheta_{nt}, f_{nu}) - \kappa(\vartheta, f)) \to tt_0\gamma(0)I + (u,u_0)_f, \quad t \in \mathbb{R}, u \in U.$$

Hängt κ nicht von f ab, so hängt die linke Seite nicht von u ab, es muß also $u = 0$ gewählt werden. In den Modellen, in denen wir f *kennen*, reduziert sich das innere Produkt auf $(D(t,u), D(t',u')) = tt'\gamma(0)I$. Es ergibt sich also der gleiche kanonische Gradient wie im obigen Modell. Wir können also κ asymptotisch so gut schätzen, als ob wir f kennten. Solche Funktionale heißen *adaptiv*. In unserem Modell sind also alle Funktionale von ϑ und analog alle Funktionale von f adaptiv bezüglich des anderen Parameters.

Ein effizienter Schätzer für ϑ. Der kanonische Gradient für $\kappa(\vartheta, f) = \vartheta$ ergibt sich mit $u_0 = 0$ aus

$$n^{1/2}(\vartheta_{nt} - \vartheta) = t = tt_0\gamma(0)I.$$

Es ergibt sich $t_0 = \gamma(0)^{-1}I^{-1}$, also

$$g_0(x,y) = \gamma(0)^{-1}I^{-1}x\dot{\ell}(y-\vartheta x).$$

Die Varianz ist $\|g_0\|_2^2 = \gamma(0)^{-1}I^{-1}$. Die Varianz von $\hat{\vartheta}_{LS}$ ist $\sigma^2/\gamma(0)$. Es gilt

$$0 = E\varepsilon = \int y f(y)\, dy + \int (y+z) f(y+z)\, dy,$$

durch Ableitung nach $z = 0$ also

$$0 = \int f(y)\, dy + \int y f'(y)\, dy = 1 - E\varepsilon\dot{\ell}(\varepsilon),$$

mit der Cauchy-Ungleichung also $1 \leq \sigma^2 I$. Also ist $\hat{\vartheta}_{LS}$ nur effizient, wenn $\dot{\ell}(x)$ proportional zu x ist, also wenn ε normalverteilt ist.

Wäre f bekannt, so ergäbe sich ein effizienter Schätzer als Maximum-Likelihood-Schätzer, also als Maximum von $\sum \log f(X_i - \vartheta X_{i-1})$ bzw. als Lösung der Martingal-Schätzgleichung

$$k(\vartheta) = \frac{1}{n}\sum_{i=1}^n X_{i-1}\dot{\ell}(X_i - \vartheta X_{i-1}) = 0.$$

Ist $\dot{\ell}(x)$ proportional zu x, so ist das wieder der Kleinste-Quadrate-Schätzer.

Das *Newton-Verfahren* zur Bestimmung der Nullstelle einer Funktion $k(\vartheta)$ iteriert $\vartheta \to \vartheta - k(\vartheta)/k'(\vartheta)$. Wir wenden es auf den Startwert $\hat{\vartheta}_{LS}$ an, der gegen die Nullstelle konvergiert, brauchen deshalb nur einen Iterationsschritt. Die Ableitung

$$k'(\vartheta) = -\frac{1}{n}\sum_{i=1}^n X_{i-1}^2\ddot{\ell}(X_i - \vartheta X_{i-1})$$

ersetzen wir durch den Erwartungswert $-\gamma(0)E\ddot{\ell}(\varepsilon) = -\gamma(0)I$. Dann liefert das *Newton–Raphson-Verfahren* bei bekanntem f den Schätzer

$$\hat{\vartheta}_{LS} + \gamma(0)^{-1}I^{-1}\frac{1}{n}\sum_{i=1}^n X_{i-1}\dot{\ell}(X_i - \hat{\vartheta}_{LS}X_{i-1}),$$

der äquivalent zum Maximum-Likelihood-Schätzer ist. Da wir f nicht kennen, müssen wir $\gamma(0)$, I und $\dot{\ell}$ durch Schätzer ersetzen, also f durch einen Kernschätzer $\hat{f}$, der auf den Residuen $\hat{\varepsilon}_i = X_i - \hat{\vartheta}_{LS}X_{i-1}$ beruht, f' durch $\hat{f}'$, $\dot{\ell}$ durch $\hat{\dot{\ell}} = -\hat{f}'/\hat{f}$, I durch $\hat{I} = (1/n)\sum \hat{\dot{\ell}}^2(\hat{\varepsilon}_i)$ und $\gamma(0)$ durch den empirischen Schätzer $\hat{\gamma}(0) = (1/n)\sum X_{i-1}^2$. Man kann zeigen, siehe z.B. Koul und Schick (1997), daß

$$\hat{\vartheta} = \hat{\vartheta}_{LS} + \hat{\gamma}(0)^{-1}\hat{I}^{-1}\frac{1}{n}\sum_{i=1}^n X_{i-1}\hat{\dot{\ell}}(X_i - \hat{\vartheta}_{LS}X_{i-1})$$

die Einflußfunktion g_0 hat, also effizient ist.

Die folgenden beiden Tricks können beim Beweis helfen.

1. *Sample splitting:* Man verwendet eine (wachsenden) Anteil der Stichprobe für $\hat{\vartheta}$ und den Rest für $\hat{f}$.

2. *Diskretisierung:* Man ersetzt $\hat{\vartheta}$ durch den nächstliegenden Punkt eines Gitters, dessen Schrittweite von der Ordnung $n^{-1/2}$ ist.

Ein effizienter Schätzer für $Ek(\varepsilon)$. Sei $k \in L_2(f)$. Wäre ϑ bekannt, hätten wir für $Ek(\varepsilon)$ den empirischen Schätzer $(1/n)\sum k(\varepsilon_i)$, der auf i.i.d. Beobachtungen beruht. Wegen $E\varepsilon = 0$ ergäbe sich ein verbesserter Schätzer

$$\frac{1}{n}\sum_{i=1}^{n} k(\varepsilon_i) - \frac{\sum \varepsilon_i k(\varepsilon_i)}{\sum \varepsilon_i^2}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i$$

mit Einflußfunktion

$$k^0(z) = k(z) - Ek(\varepsilon) - \frac{E\varepsilon k(\varepsilon)}{E\varepsilon^2} z.$$

Für i.i.d. Beobachtungen $\varepsilon_1, \dots, \varepsilon_n$ mit $E\varepsilon = 0$ ist das lokale Modell durch $f_{nu}(x) = f(x)(1 + n^{-1/2}u(x))$ gegeben, und wegen $E\varepsilon = 0$ muß $E\varepsilon u(\varepsilon) = 0$ gelten, also ist U der lokale Parameter-Raum. Der kanonische Gradient von $Ek(\varepsilon)$ ergibt sich durch Projektion der Einflußfunktion $k - Ek(\varepsilon)$ des empirischen Schätzers auf U^- als $u_0(x) = k^0(x)$. Also ist der verbesserte Schätzer effizient.

Im AR(1)-Modell ergibt sich wegen der Adaptivität der kanonische Gradient von $Ek(\varepsilon)$ genauso:

$$n^{1/2}\Big(\int k(x) f_{nu}(x)\,dx - \int k(x) f(x)\,dx\Big) = (k, u)_f = (u_0, u)_f, \quad u \in U.$$

Die Einflußfunktion ist jetzt $u_0(y - \vartheta x)$. Weil wir ϑ nicht kennen, schätzen wir $Ek(\varepsilon)$ mit

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^{n} k(\hat{\varepsilon}_i) - \frac{\sum \hat{\varepsilon}_i k(\hat{\varepsilon}_i)}{\sum \hat{\varepsilon}_i^2}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i.$$

Falls k geeignet differenzierbar ist, erhalten wir mit $\hat{\varepsilon}_i - \varepsilon_i = -(\hat{\vartheta}_{LS} - \vartheta)X_{i-1}$:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} k(\hat{\varepsilon}_i) &= \frac{1}{n}\sum_{i=1}^{n} k(\varepsilon_i) - (\hat{\vartheta}_{LS} - \vartheta)\frac{1}{n}\sum_{i=1}^{n} X_{i-1}k'(\varepsilon_i) + o_p(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^{n} k(\varepsilon_i) + o_p(n^{-1/2});\end{aligned}$$

letzteres wegen $n^{1/2}(\hat{\vartheta}_{LS} - \vartheta) = O_p(1)$ und $EX_{i-1}k'(\varepsilon_i) = 0$. Die Voraussetzung $E\varepsilon = 0$ hat also den Effekt, daß die Ersetzung von ε_i durch $\hat{\varepsilon}_i$ den Schätzer für $Ek(\varepsilon)$ nicht verändert. Er hat also wieder die Einflußfunktion $u_0(y - \vartheta x)$ und ist effizient. Insbesondere haben wir keinen effizienten Schätzer für ϑ gebraucht.

Effiziente Schätzer für bedingte Erwartungswerte. Es gilt

$$P_{\mathcal{M}_n} X_{n+1} = E(X_{n+1}|X_n) = \vartheta X_n.$$

Ein Schätzer für den Prädiktor ist also $\hat{\vartheta} X_n$. Der Prädiktor ist zufällig. Damit wir asymptotische Verteilungen und Effizienz studieren können, betrachten wir das Problem, bedingte Erwartungswerte gegeben einen festen Wert $X_n = w$ zu schätzen. Damit das Problem nichttrivial wird, betrachten wir bedingte Erwartungswerte von Funktionen $k(X_{n+1})$ statt von X_{n+1}. Für eine reellwertige Markov-Kette mit Übergangsverteilung $Q(x, dy)$ und Übergangsdichte $q(x, y)$ gilt

$$\kappa = E(k(X_{n+1})|X_n = w) = \int k(y) q(w, y)\, dy.$$

Ist $p(x)$ die stationäre Dichte von X_0 und $p_2(x, y)$ die von (X_0, X_1), so gilt $q(x, y) = p_2(x, y)/p(x)$. Sind $\hat{p}$ und $\hat{p}_2$ (Kern-)Schätzer für p und p_2, so erhalten wir als Schätzer für κ den nichtparametrischen Schätzer

$$\hat{\kappa} = \frac{1}{\hat{p}(w)} \int k(y) \hat{p}_2(w, y)\, dy.$$

Seine Konvergenzrate hängt von der Bandweite ab. Insbesondere ist sie höchstens so schnell wie die von $\hat{p}(w)$, also langsamer als $n^{-1/2}$.

Im AR(1)-Modell gilt

$$E(k(X_{n+1})|X_n = w) = E(k(\vartheta X_n + \varepsilon_{n+1})|X_n = w) = Ek(\varepsilon + \vartheta w).$$

Der bedingte Erwartungswert ist also ein gewöhnlicher. Wäre ϑ bekannt, ergäbe sich ein effizienter Schätzer als

$$\frac{1}{n} \sum_{i=1}^{n} k(\varepsilon_i + \vartheta w) - \frac{\sum \varepsilon_i k(\varepsilon_i + \vartheta w)}{\sum \varepsilon_i^2} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i.$$

Wir haben beim Schätzen von $Ek(\varepsilon)$ gesehen, daß sich asymptotisch nichts ändert, wenn ε_i durch $\hat{\varepsilon}_i$ ersetzt wird. Aber $Ek(\varepsilon + \vartheta w)$ hängt auch von ϑ

ab. Es ist übersichtlicher, allgemeiner das Schätzen von $Ek_\vartheta(\varepsilon)$ zu betrachten. Sei k_ϑ geeignet in ϑ differenzierbar, mit Ableitung $\dot{k}_\vartheta$. Wir schreiben $Ek_\vartheta(\varepsilon) = \int k_\vartheta(x) f(x)\, dx$ und erhalten

$$\begin{aligned} &n^{1/2}\Big(\int k_{\vartheta_{nt}}(x) f_{nu}(x)\, dx - \int k_\vartheta(x) f(x)\, dx\Big) \\ &\to tE\dot{k}_\vartheta(\varepsilon) + (u, k_\vartheta)_f, \quad t \in \mathbb{R}, u \in U. \end{aligned}$$

Der kanonische Gradient $g_0(x,y) = t_0 x \dot{\ell}(y - \vartheta x) + u_0(y - \vartheta x)$ ergibt sich also als Lösung von

$$tE\dot{k}_\vartheta(\varepsilon) + (u, k_\vartheta)_f = tt_0\gamma(0)I + (u, u_0)_f, \quad t \in \mathbb{R}, u \in U.$$

Das liefert mit Projektion von k_ϑ auf U^-

$$t_0 = \frac{E\dot{k}_\vartheta(\varepsilon)}{\gamma(0)I}, \quad u_0(x) = k_\vartheta^0(x) - Ek_\vartheta^0(\varepsilon) - \frac{E\varepsilon k_\vartheta(\varepsilon)}{E\varepsilon^2} x.$$

Als Schätzer für $\kappa = Ek_\vartheta(\varepsilon)$ nehmen wir

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^n k_{\hat{\vartheta}}(\hat{\varepsilon}_i) - \frac{\sum \hat{\varepsilon}_i k_{\hat{\vartheta}}(\hat{\varepsilon}_i)}{\sum \hat{\varepsilon}_i^2} \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i.$$

Wir wissen schon, daß sich asymptotisch nichts ändert, wenn wir ε_i durch $\hat{\varepsilon}_i$ ersetzen. Außerdem gilt

$$\frac{1}{n}\sum_{i=1}^n \varepsilon_i k_{\hat{\vartheta}}(\varepsilon_i) \to E\varepsilon k_\vartheta(\varepsilon).$$

Wir brauchen nur noch die stochastische Entwicklung

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n k_{\hat{\vartheta}}(\varepsilon_i) &= \frac{1}{n}\sum_{i=1}^n k_\vartheta(\varepsilon_i) + (\hat{\vartheta} - \vartheta)\frac{1}{n}\sum_{i=1}^n \dot{k}_\vartheta(\varepsilon_i) + o_p(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^n k_\vartheta(\varepsilon_i) + (\hat{\vartheta} - \vartheta)E\dot{k}_\vartheta(\varepsilon) + o_p(n^{-1/2}). \end{aligned}$$

Insgesamt erhalten wir

$$\hat{\kappa} - \kappa = \frac{1}{n}\sum_{i=1}^n u_0(\varepsilon_i) + (\hat{\vartheta} - \vartheta)E\dot{k}_\vartheta(\varepsilon) + o_p(n^{-1/2}).$$

Also ist $\hat{\kappa}$ genau dann effizient, wenn $\hat{\vartheta}$ effizient ist.

Von-Mises-Statistiken. Als Vorbereitung für das Schätzen bedingter Erwartungswerte höherer Ordnung betrachten wir das Schätzen von Erwartungswerten $\kappa(f) = Eh(\varepsilon_1, \varepsilon_2)$ aufgrund von unabhängigen Beobachtungen $\varepsilon_1, \dots, \varepsilon_n$ mit Dichte f. Der naheliegende Schätzer ist der empirische Schätzer, die von-Mises-Statistik

$$\hat{\kappa} = \frac{1}{n^2} \sum_{i,j=1}^{n} h(\varepsilon_i, \varepsilon_j).$$

Wir haben schon in Kapitel 4 gesehen, daß mit $f_{nu}(x) = f(x)(1+n^{-1/2}u(x))$ für beschränktes $u \in L_{2,0}(f)$ LAN gilt:

$$\Lambda_{nu} = n^{-1/2} \sum_{i=1}^{n} u(\varepsilon_i) - \frac{1}{2}\|u\|_f^2 + o_p(1).$$

Ist $h \in L_2(f \otimes f)$, so gilt

$$\begin{aligned}
&n^{1/2}(\kappa(f_{nu}) - \kappa(f)) \\
&= n^{1/2}\Big(\iint h(x,y) f_{nu}(x) f_{nu}(y)\, dx\, dy - \iint f(x)f(y)\, dx\, dy\Big) \\
&= \iint h(x,y)\big(u(x) + u(y) + n^{-1/2}u(x)u(y)\big) f(x)f(y)\, dx\, dy \\
&= (u, h_1 + h_2)_f + O(n^{-1/2})
\end{aligned}$$

mit

$$\begin{aligned}
h_1(x) &= \int h(x,y)f(y)\, dy = E(h(\varepsilon_1, \varepsilon_2)|\varepsilon_1 = x), \\
h_2(x) &= \int h(y,x)f(y)\, dy = E(h(\varepsilon_1, \varepsilon_2)|\varepsilon_2 = x).
\end{aligned}$$

Der kanonische Gradient ergibt sich durch Projektion von h_1+h_2 auf $L_{2,0}(f)$:

$$u_0 = \bar{h}_1 + \bar{h}_2, \quad \bar{h}_r = h_r - Eh_r(\varepsilon).$$

Wir zeigen, daß die von-Mises-Statistik diese Einflußfunktion hat. Wie bei der Hoeffding-Zerlegung setzen wir

$$m(x,y) = \bar{h}(y,x) - \bar{h}_1(x) - \bar{h}_2(y), \quad \bar{h} = h - Eh(\varepsilon_1, \varepsilon_2).$$

Dann gilt

$$\begin{aligned}
\hat{\kappa} = \frac{1}{n^2} \sum_{i,j=1}^{n} h(\varepsilon_i, \varepsilon_j) &= \kappa(f) + \frac{1}{n} \sum_{i=1}^{n} (\bar{h}_1(\varepsilon_i) + \bar{h}_2(\varepsilon_i)) + \frac{1}{n^2} \sum_{i,j=1}^{n} m(\varepsilon_i, \varepsilon_j) \\
&= \kappa(f) + \frac{1}{n} \sum_{i=1}^{n} (\bar{h}_1(\varepsilon_i) + \bar{h}_2(\varepsilon_i)) + O_p(n^{-1}).
\end{aligned}$$

Also ist die von-Mises-Statistik effizient.

Schätzer für bedingte Erwartungswerte höherer Ordnung. Sei $k \in L_2(f)$. Es gilt

$$\begin{aligned} E(k(X_{n+2})|X_n = w) &= E(k(\vartheta X_{n+1} + \varepsilon_{n+2})|X_n = w) \\ &= E(k(\vartheta^2 X_n + \vartheta\varepsilon_{n+1} + \varepsilon_{n+2})|X_n = w) \\ &= E(k(\vartheta\varepsilon_{n+1} + \varepsilon_{n+2} + \vartheta^2 w) = Eh_\vartheta(\varepsilon_1, \varepsilon_2) \end{aligned}$$

für $h_\vartheta(x, y) = k(\vartheta x + y + \vartheta^2 w)$. Wir betrachten das allgemeinere Problem des Schätzens von

$$\kappa(\vartheta, f) = Eh_\vartheta(\varepsilon_1, \varepsilon_2) = \iint h_\vartheta(x, y) f(x) f(y)\, dx\, dy.$$

Wir erhalten

$$\begin{aligned} &n^{1/2}(\kappa(\vartheta_{nt}, f_{nu}) - \kappa(\vartheta, f)) \\ &= n^{1/2}(\kappa(\vartheta, f_{nu}) - \kappa(\vartheta, f)) + n^{1/2}(\kappa(\vartheta_{nt}, f) - \kappa(\vartheta, f)) + o(1) \\ &\to (u, h_{\vartheta 1} + h_{\vartheta 2})_f + tE\dot{h}_\vartheta(\varepsilon_1, \varepsilon_2), \quad t \in \mathbb{R}, u \in U, \end{aligned}$$

wobei $\dot{h}_\vartheta(x, y)$ die Ableitung von $h_\vartheta(x, y)$ nach ϑ bezeichnet, Der kanonische Gradient ergibt sich durch Projektion von $h_{\vartheta 1} + h_{\vartheta 2}$ auf U^- mit

$$u_0 = h^0_{\vartheta 1} + h^0_{\vartheta 2}, \quad t_0 = \frac{E\dot{h}_\vartheta(\varepsilon_1, \varepsilon_2)}{\gamma(0)I}.$$

Wir schätzen $Eh_\vartheta(\varepsilon_1, \varepsilon_2)$ mit

$$\hat{\kappa} = \frac{1}{n^2} \sum_{i,j=1}^n h_{\hat{\vartheta}}(\hat{\varepsilon}_i, \hat{\varepsilon}_j).$$

Wir haben gesehen, daß sich asymptotisch nichts ändert, wenn wir ε_k durch $\hat{\varepsilon}_k$ ersetzen. Wir erhalten die stochastische Entwicklung

$$\begin{aligned} \hat{\kappa} - \kappa(\vartheta, f) = \frac{1}{n} &\sum_{i=1}^n (\bar{h}_{\vartheta 1}(\varepsilon_i) + \bar{h}_{\vartheta 2}(\varepsilon_i)) \\ &+ (\hat{\vartheta} - \vartheta)\frac{1}{n^2} \sum_{i,j=1}^n \dot{h}_\vartheta(\varepsilon_i, \varepsilon_j) + o_p(n^{-1/2}). \end{aligned}$$

Es gilt

$$\frac{1}{n^2}\sum_{i,j=1}^n \dot h_\vartheta(\varepsilon_i,\varepsilon_j) = E\dot h_\vartheta(\varepsilon_1,\varepsilon_2) + o_p(1).$$

Falls $\hat\vartheta$ effizient ist, liefert der zweite Term der Entwicklung von $\hat\kappa$ den richtigen Anteil der effizienten Einflußfunktion. Im ersten Term müßte aber $h^0_{\vartheta 1} + h^0_{\vartheta 2}$ statt $\bar h_{\vartheta 1} + \bar h_{\vartheta 2}$ stehen. Der Schätzer $\hat\kappa$ ist also nicht effizient. Wir haben die Information $E\varepsilon = 0$ noch nicht verwendet. Das können wir jetzt nachholen. Es gilt

$$E\varepsilon \bar h_{\vartheta 1}(\varepsilon) = E\varepsilon h_{\vartheta 1}(\varepsilon) = E\varepsilon_1 h_\vartheta(\varepsilon_1,\varepsilon_2),$$
$$E\varepsilon \bar h_{\vartheta 2}(\varepsilon) = E\varepsilon h_{\vartheta 2}(\varepsilon) = E\varepsilon_2 h_\vartheta(\varepsilon_1,\varepsilon_2).$$

Also korrigieren wir $\hat\kappa$ wie folgt:

$$\hat\kappa - \frac{\frac{1}{n^2}\sum_{i,j=1}^n(\hat\varepsilon_i + \hat\varepsilon_j)h_{\hat\vartheta}(\hat\varepsilon_i,\hat\varepsilon_j)}{\frac{1}{n}\sum_{i=1}^n \hat\varepsilon_i^2}\,\frac{1}{n}\sum_{i=1}^n \hat\varepsilon_i.$$

Schätzer für $Ek(X)$. Sei $k \in L_2(\pi)$. Für nichtparametrische Markov-Ketten ist der empirische Schätzer $(1/n)\sum k(X_i)$ für $\pi k = Ek(X)$ asymptotisch linear mit Einflußfunktion Ak und effizient; siehe Sätze 10 und 12.

Es gelte die Nebenbedingung $\pi h = Eh(X) = 0$ für ein $h \in L_2(\pi)$. Dann ergibt sich für jedes $c \in \mathbb{R}$ ein neuer Schätzer für $Ek(X)$,

$$\frac{1}{n}\sum_{i=1}^n (k(X_i) - ch(X_i)).$$

Er hat die Einflußfunktion $Ak - cAh$, also nach Satz 7 die asymptotische Varianz $\pi \otimes Q(Ak - cAh)^2$. Sie wird minimiert für

$$c = \pi \otimes QAk\,Ah/\pi \otimes Q(Ah)^2 = a/b.$$

Es gilt wie in Kapitel 9 mit $Eh(X) = 0$:

$$a = \pi \otimes QAk\,Ah = Ek(X)h(X) + \sum_{s=1}^\infty Ek(X_0)h(X_s) + \sum_{s=1}^\infty Ek(X_s)h(X_0).$$

Ein empirischer Schätzer für a ist

$$\hat a = \frac{1}{n}\sum_{i=1}^n k(X_i)h(X_i) + \sum_{s=1}^m \frac{1}{n-s}\sum_{i=1}^{n-s}\big(k(X_i)h(X_{i+s}) + k(X_{i+s})h(X_i)\big).$$

Ein empirischer Schätzer für b ist analog

$$\hat{b} = \frac{1}{n}\sum_{i=1}^{n} h^2(X_i) + 2\sum_{s=1}^{m}\frac{1}{n-s}\sum_{i=1}^{n-s} h(X_i)h(X_{i+s}).$$

Dabei wächst m langsam mit n. Es ergibt sich der Schätzer

$$\hat{\kappa} = \frac{1}{n}\sum_{i=1}^{n} k(X_i) - \frac{\hat{a}}{\hat{b}}\frac{1}{n}\sum_{i=1}^{n} h(X_i)$$

mit Einflußfunktion $g_0 = Ak - (a/b)Ah$ und asymptotischer Varianz

$$\begin{aligned}\pi \otimes Qg_0^2 &= \pi \otimes Q(Ak)^2 - 2\frac{a}{b}a + \frac{a^2}{b^2}b \\ &= \pi \otimes Q(Ak)^2 - \frac{(\pi \otimes QAk\,Ah)^2}{\pi \otimes Q(Ah)^2}.\end{aligned}$$

Das lokale Modell ist $Q_{nu}(x, dy) = Q(x, dy)(1 + n^{-1/2}u(x, y))$. Dabei ist u beschränkt, $Qu = 0$ und wegen der Nebenbedingung auch

$$0 = n^{1/2}(\pi_{nu}h - \pi h) \to \pi \otimes QuAh.$$

Der kanonische Gradient für $\pi k = Ek(X)$ ergibt sich aus

$$n^{1/2}(\pi_{nu}k - \pi k) \to \pi \otimes QuAk$$

als Projektion von Ak auf den Raum der u, also wieder als g_0. Also ist $\hat{\kappa}$ effizient unter der Nebenbedingung $\pi h = 0$.

Im AR(1)-Modell können wir schreiben:

$$Ek(X_t) = Ek(\vartheta X_{t-1} + \varepsilon_t).$$

Als Schätzer bietet sich eine von-Mises-Statistik an:

$$\frac{1}{n^2}\sum_{i,j=1}^{n} k(\hat{\vartheta}X_i + \hat{\varepsilon}_j).$$

Wegen $E\varepsilon = EX = 0$ kann er noch verbessert werden. Er ist dann noch nicht effizient, denn X_{t-1} kann wiederum durch $\vartheta X_{t-2} + \varepsilon_{t-1}$ ersetzt werden. Durch Iteration:

$$Ek(X_t) = Ek\Big(\sum_{j=0}^{\infty} \vartheta^j \varepsilon_{t-j}\Big).$$

Das kann durch eine von-Mises-Statistik wachsender Ordnung geschätzt werden:

$$\hat{\kappa} = \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} k\Big(\sum_{j=1}^{m} \hat{\vartheta}^j \hat{\varepsilon}_{i_j}\Big).$$

Dabei wächst m langsam mit n. Der Schätzer kann wegen $E\varepsilon = 0$ noch verbessert werden und ist dann effizient, wenn $\hat{\vartheta}$ effizient ist.

Schätzer für die stationäre Dichte. Wir haben schon erwähnt, daß die Dichte p der invarianten Verteilung π einer reellwertigen Markov-Kette durch (Kern-)Schätzer $\hat{p}$ geschätzt werden kann, deren Rate aber schlechter als $n^{-1/2}$ ist. Im AR(1)-Modell gilt nach Definition der invarianten Verteilung

$$p(y) = \int f(y - \vartheta x) p(x)\, dx.$$

Als Schätzer für $p(y)$ ergibt sich die lokale von-Mises-Statistik

$$\hat{p}^{(0)}(y) = \frac{1}{n} \sum_{j=1}^{n} \hat{f}(y - \hat{\vartheta} X_j) = \frac{1}{n^2} \sum_{i,j=1}^{n} K_b(y - \hat{\varepsilon}_i - \hat{\vartheta} X_j)$$

und ein Plug-in-Schätzer

$$\hat{p}^{(1)}(y) = \int \hat{f}(y - \vartheta x) \hat{p}(x)\, dx.$$

Das läßt sich wie beim Schätzen von $Ek(X)$ iterieren,

$$\begin{aligned} p(y) &= \iint f(y - \vartheta x) f(x - \vartheta w) p(w)\, dx\, dw \\ &= \iint f(y - \vartheta x - \vartheta^2 w) f(x) p(w)\, dx\, dw \end{aligned}$$

und schließlich

$$p(y) = \int f\Big(y - \sum_{j=1}^{\infty} \vartheta^j x_j\Big) \prod_{j=1}^{\infty} f(x_j)\, dx_j = Ef\Big(y - \sum_{j=1}^{\infty} \vartheta^j \varepsilon_j\Big).$$

Die Dichte kann also durch eine lokale von-Mises-Statistik wachsender Ordnung geschätzt werden:

$$\begin{aligned} \hat{p}^{(m)}(y) &= \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} \hat{f}\Big(y - \sum_{j=1}^{m} \hat{\vartheta}^j \hat{\varepsilon}_{i_j}\Big) \\ &= \frac{1}{n^m} \sum_{i_1,\dots,i_m=1}^{n} \frac{1}{n} \sum_{i=1}^{n} K_b\Big(y - \hat{\varepsilon}_i - \sum_{j=1}^{m} \hat{\vartheta}^j \hat{\varepsilon}_{i_j}\Big). \end{aligned}$$

Dabei wächst m langsam mit n. Der Schätzer kann wegen $E\varepsilon = 0$ noch verbessert werden und ist dann effizient, wenn $\hat{\vartheta}$ effizient ist.

Literatur

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods.* Second Edition. Springer Series in Statistics. Springer-Verlag, New York.

Durrett, R. (1991). *Probability. Theory and Examples.* The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole, Pacific Grove.

Kartashov, N. V. (1985). Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theory Probab. Math. Statist.* 30, 71–89.

Koul, H. L. and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* 3, 247–277.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability.* Communications and Control Engineering Series. Springer-Verlag, London.

Revuz, D. (1984). *Markov Chains.* Second Edition. North-Holland Mathematical Library 11. North-Holland, Amsterdam.