

JOHANNES GUTENBERG-UNIVERSITÄT MAINZ

**Parameterschätzung in zeitdiskreten
ergodischen Markov-Prozessen
am Beispiel des Cox-Ingersoll-Ross-Modells**

Diplomarbeit

vorgelegt dem Fachbereich Physik, Mathematik und Informatik
im Oktober 2005

von

Matthias Hammer

Inhaltsverzeichnis

Einleitung	v
1 Das Cox-Ingersoll-Ross-Modell	1
1.1 Einführung und grundlegende Eigenschaften des Modells	1
1.2 Diskretisierung des Modells	9
2 Parameterschätzung in zeitdiskreten Markovprozessen	11
2.1 Problemstellung und allgemeine Voraussetzungen	11
2.2 Schätzgleichungen und Martingal-Schätzfunktionen	14
2.3 Existenz und Asymptotik von Lösungen	18
2.4 Likelihood-Schätzung und LAN	41
2.5 Bemerkungen zu optimalen Schätzfunktionen	48
3 Parameterschätzung im diskretisierten CIR-Modell	55
3.1 Vorbereitungen	55
3.2 Schätzung mittels quadratischer Schätzfunktionen	56
3.2.1 Conditional Least Squares-Schätzung	58
3.2.2 Schätzung mittels optimaler quadratischer Schätzfunktion	75
3.3 Likelihood-Schätzung und LAN im CIR-Modell	79
A Maßtheoretische Hilfsresultate	97
Literaturverzeichnis	101

Einleitung

In dieser Arbeit beschäftigen wir uns mit der Schätzung von Parametern in zeitdiskreten ergodischen Markovprozessen im allgemeinen und im sogenannten *Cox-Ingersoll-Ross-Modell* im besonderen. Betrachten wir die eindimensionale stochastische Differentialgleichung

$$dY_t = (a + bY_t) dt + \sigma \sqrt{Y_t^+} dW_t, \quad (1)$$

mit den Parametern $a, b \in \mathbb{R}$ und $\sigma > 0$, wobei W eine Standard-Brownsche Bewegung und $x^+ := \max(x, 0)$ der Positivteil ist. Gleichung (1) hat für jede vorgegebene Startverteilung eine eindeutige und globale starke Lösung $(Y_t)_{t \geq 0}$. Für $a > 0, b < 0$ und eine auf $\mathbb{R}_{\geq 0}$ konzentrierte Startverteilung erhält man das 1985 von Cox, Ingersoll und Ross zur Beschreibung der Entwicklung von Zinsraten vorgeschlagene Modell, siehe [CIR]. Wir bezeichnen es kurz als **CIR**-Modell und die entsprechende Lösung von (1) als **CIR**-Prozeß. Diese ist ein Diffusionsprozeß, d.h. ein Markov-Prozeß mit stetigen Pfaden. Vom Standpunkt des Anwenders aus hat das Modell unter anderem die Vorzüge, daß die Lösung $(Y_t)_{t \geq 0}$ stets nichtnegativ bleibt und für $t \rightarrow \infty$ schwach gegen eine Gammaverteilung konvergiert. Ziel dieser Arbeit ist es nun, die Parameter a, b, σ^2 aufgrund von Beobachtungen des Prozesses zu äquidistanten diskreten Zeitpunkten $n \cdot \Delta$ der Schrittweite $\Delta > 0$ zu schätzen. Man betrachtet also den Prozeß $(X_n)_{n \in \mathbb{N}_0}$ mit

$$X_n := Y_{n \cdot \Delta}, \quad n \in \mathbb{N}_0,$$

den wir als *diskretisierten CIR*-Prozeß bezeichnen. Dieses Beispiel einer diskret beobachteten Diffusion führt uns zum allgemeinen Problem der Parameterschätzung in zeitdiskreten Markov-Prozessen. Die Analyse dieses Problems bildet den Schwerpunkt der Arbeit, die wie folgt aufgebaut ist:

Kapitel 1 gibt einen kurzen Überblick über die wichtigsten Eigenschaften des **CIR**-Prozesses, wobei die meisten einschlägigen Resultate ohne Beweis aus [IW] entnommen sind. Eine Ausnahme bildet insbesondere Satz 1.1.4, wo wir eine explizite Darstellung für die n -ten Momente des Prozesses beweisen.

In Kapitel 2 beschäftigen wir uns dann mit dem Problem der Parameterschätzung im allgemeinen Kontext. Ausgangspunkt ist ein zeitdiskreter stochastischer Prozeß $(X_n)_{n \in \mathbb{N}_0}$, dessen Verteilung von einem d -dimensionalen Parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ abhängt. Die grundlegende Annahme ist, daß $(X_n)_n$ unter jedem $\theta \in \Theta$ ein homogener und ergodischer Markovprozeß ist, zum Beispiel ein zu diskreten äquidistanten Zeitpunkten beobachteter Diffusionsprozeß. Zur Schätzung von θ stützen wir uns auf die Methode der *Martingal-Schätzfunktionen* und *-Schätzgleichungen*, die in ihrer allgemeinen Form insbesondere von Bibby und Sørensen, beginnend mit [BS], eingehend untersucht wurde. Eine Schätzfunktion

$$G_n(\theta) \equiv G_n(X_0, \dots, X_n; \theta)$$

ist eine \mathbb{R}^d -wertige Funktion des Parameters θ und der ersten n Beobachtungen des Prozesses. Unter geeigneten Voraussetzungen erhält man einen Schätzer für den Parameter θ durch eine Lösung der Gleichung

$$G_n(\theta) = 0. \tag{2}$$

Bildet die Folge $(G_n(\theta))_n$ unter jedem $\theta \in \Theta$ ein Martingal, so spricht man von einer Folge von Martingal-Schätzfunktionen. Nach einem Überblick über die benötigten Voraussetzungen und Hilfssätze wird in Abschnitt 2.3, der wesentlich auf [Sør] aufbaut, der Begriff der (stark) konsistenten Lösung einer Schätzgleichung definiert und die Frage nach der Existenz und Asymptotik solcher Lösungen gestellt. Lemma 2.3.3 enthält ein wichtiges technisches Hilfsresultat über lokal dominiert integrierbare Funktionen. Im Anschluß an [Sør] werden dann vier Bedingungen formuliert, die in unterschiedlicher Kombination die Existenz sowie asymptotische Normalität von Lösungen der Schätzgleichung (2) sichern. Theorem 2.3.5 enthält zunächst ein Resultat zur asymptotischen Normalität einer vorgelegten konsistenten Lösung von (2). In Satz 2.3.6 beweisen wir dann „für schließlich alle n “ die Existenz von Lösungen der Schätzgleichung (2) in beliebig kleinen Umgebungen eines jeden Punktes $\theta \in \Theta$. Dieses Resultat liefert jedoch noch nicht die Existenz einer konsistenten Schätzfolge, da unsere Bedingungen lokaler Natur sind und somit die Eindeutigkeit der Lösung von (2) nicht sichergestellt ist. Damit steht man vor dem Problem, aus der Lösungsmenge der Gleichung (2) eine Auswahl derart zu treffen, daß die resultierende Schätzfolge (meßbar und) konsistent wird. Zur Lösung dieses Problems folgen wir einer Idee von [Leh] und [Pfa]: nimmt man als zusätzliche Voraussetzung die Existenz einer konsistenten Folge von Hilfsschätzern an, so kann unter deren Verwendung eine konsistente Lösung von (2) konstruiert werden. Diese Konstruktion wird in Theorem 2.3.8, dem Hauptresultat dieses Abschnitts, durchgeführt. Der anschließende Abschnitt 2.4 enthält eine Anwendung der bisher erzielten Resultate auf das Problem der *Likelihood-Schätzung*. Es stellt sich heraus, daß praktisch dieselben Bedingungen, die die Existenz einer konsistenten Lösung der Likelihood-Gleichung liefern, auch lokal-asymptotische Normalität, kurz: **LAN**, an jeder Stelle $\theta \in \Theta$ sicherstellen; dies ist der Inhalt von Theorem 2.4.2. Abschnitt 2.5 enthält eine

kurze und elementare Diskussion asymptotisch optimaler Schätzfunktionen. Diese sind dadurch ausgezeichnet, daß die als Lösung von (2) definierte Schätzfolge die kleinstmögliche asymptotische Varianz innerhalb einer gewissen Klasse von Schätzfunktionen hat. In dem der Cramér-Rao-Schranke „nachempfundenen“ Lemma 2.5.1 wird eine einfache hinreichende Bedingung für Optimalität in diesem Sinne angegeben. Als Beispiel untersuchen wir anschließend die Klasse der *quadratischen* Schätzfunktionen; bei diesen handelt es sich um gewichtete Kombinationen der bedingten Erwartung und der bedingten Varianz. In Satz 2.5.2 werden optimale Gewichte für diese Klasse berechnet.

In Kapitel 3 kehren wir zum diskretisierten **CIR**-Prozeß zurück und untersuchen zunächst quadratische Martingal-Schätzfunktionen. Im Unterabschnitt 3.2.1, der sich an der Arbeit [OR] von Overbeck und Rydén orientiert, wird zunächst unter Verwendung des Conditional Least Squares-Ansatzes eine Schätzfolge für den Driftkoeffizienten (a, b) konstruiert. Anschließend werden zwei Schätzfolgen für den Diffusionskoeffizienten σ^2 vorgeschlagen. Daraus ergeben sich zwei Schätzfolgen $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ für (a, b, σ^2) , die sich als stark konsistente Lösungen quadratischer Martingal-Schätzgleichungen herausstellen; dies ist die Aussage von Theorem 3.2.6. In Theorem 3.2.8 zeigen wir asymptotische Normalität für beide Schätzfolgen unter Verwendung der allgemeinen Ergebnisse aus Kapitel 2. Da die asymptotischen Varianzen der beiden Schätzfolgen auf analytischem Wege nicht verglichen werden kann, wird in Unterabschnitt 3.2.2 die Frage nach der optimalen quadratischen Schätzfunktion im Sinne von Abschnitt 2.5 für das **CIR**-Modell aufgeworfen. In Theorem 3.2.9 wird gezeigt, daß ein solches Element sowie eine zugehörige Schätzfolge $(\hat{\theta}_n^*)_n$, die die kleinstmögliche asymptotische Varianz in dieser Klasse realisiert und damit „besser“ als $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ ist, tatsächlich existieren. Der letzte Abschnitt der Arbeit befaßt sich dann mit Likelihood-Schätzung und lokal-asymptotischer Normalität im diskretisierten **CIR**-Modell. Die Ergebnisse aus Kapitel 2 werden verwendet, um die Existenz einer eindeutigen stark konsistenten Lösung der Likelihood-Gleichung sowie **LAN** an jeder Stelle (a, b, σ^2) des Parameterraumes zu zeigen. Der Nachweis der **LAN**-Bedingung (mit Ausnahme der Invertierbarkeit der Fisher-Information) findet sich ebenfalls in der Arbeit [OR]. Die Autoren beschränken sich allerdings auf das Submodell derjenigen Parameterwerte (a, b, σ^2) mit $2a/\sigma^2 > 1$; diese Einschränkung stellt sich jedoch (anders als bei Beobachtung in stetiger Zeit, siehe [Ove]) als unnötig heraus.

Die Arbeit schließt mit einem Anhang, in dem wir einige maßtheoretische Hilfsmittel zusammengestellt haben. Insbesondere formulieren und beweisen wir eine einfache Version des „Measurable Selection Theorem“ für mengenwertige Abbildungen.

Kapitel 1

Das Cox-Ingersoll-Ross-Modell

1.1 Einführung und grundlegende Eigenschaften des Modells

Betrachte die eindimensionale stochastische Differentialgleichung

$$\begin{aligned}dY_t &= (a + bY_t) dt + \sigma \sqrt{Y_t^+} dW_t, \\ Y_0 &\sim \nu\end{aligned}\tag{1.1.1}$$

mit einer eindimensionalen Standard-Brownschen Bewegung $(W_t)_{t \geq 0}$, den Parametern $a, \sigma > 0$ und $b < 0$ sowie der Startverteilung ν . In [CIR] schlugen Cox, Ingersoll und Ross vor, die Dynamik von Zinsraten durch die Lösung $(Y_t)_{t \geq 0}$ von (1.1.1) zu beschreiben; daher ist das Modell als *Cox-Ingersoll-Ross-Modell* (kurz: **CIR-Modell**) bekannt.

Nach [IW], S. 235 f. gibt es für jedes auf $\mathbb{R}_{\geq 0}$ konzentrierte Wahrscheinlichkeitsmaß ν eine eindeutige und globale starke Lösung von $(Y_t)_{t \geq 0}$ von (1.1.1). Diese ist nichtnegativ, $Y_t \geq 0$ für alle $t \geq 0$, also kann Y_t^+ in (1.1.1) durch Y_t ersetzt werden. Da die Differentialgleichung (1.1.1) vom Markov-Typ ist, ist $(Y_t)_{t \geq 0}$ ein Markovprozeß mit stetigen Pfaden, d.h. ein Diffusionsprozeß; der zugehörige Generator ist nach [IW], S. 236 gegeben durch

$$Lf(x) = \frac{\sigma^2}{2} x \frac{d^2}{dx^2} f(x) + (a + bx) \frac{d}{dx} f(x).$$

Da wir uns nur für die Verteilungen des Prozesses interessieren, können wir davon ausgehen, daß $(Y_t)_{t \geq 0}$ als kanonischer Prozeß auf dem kanonischen Pfadraum $C([0, \infty))$ vorliegt: sei $(Q_t)_{t \geq 0}$ die Markov-Halbgruppe von $(Y_t)_{t \geq 0}$ und $\Omega := C([0, \infty))$ der kanonische Pfadraum versehen mit der von den Koordinatenprojektionen

$$Y_t : \Omega = C([0, \infty)) \ni \omega \mapsto \omega(t) = Y_t(\omega)$$

erzeugten σ -Algebra \mathcal{A} . Dann gibt es eine Familie von Wahrscheinlichkeitsmaßen $\{P^x : x \geq 0\}$ auf (Ω, \mathcal{A}) , so daß der kanonische Prozeß $(Y_t)_{t \geq 0}$ unter P^x ein Markovprozeß (bezüglich der kanonischen Filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, $\mathcal{F}_t := \sigma(Y_s : 0 \leq s \leq t)$) mit Halbgruppe $(Q_t)_{t \geq 0}$ und Startverteilung δ_x ist. Schreibe E^x für Erwartungswerte unter dem Wahrscheinlichkeitsmaß P^x .

In [IW], S. 236 wird mit Hilfe der Itô-Formel die Laplace-Transformierte des **CIR**-Prozesses berechnet. Sie ist gegeben durch

$$\begin{aligned} \psi_{Y_t}(\lambda; x) &:= E^x [e^{-\lambda Y_t}] \\ &= \left(\frac{c_t}{c_t + \lambda} \right)^{q+1} \cdot \exp \left(-x e^{bt} \frac{\lambda c_t}{c_t + \lambda} \right), \quad x \geq 0, t > 0, \end{aligned} \quad (1.1.2)$$

wobei

$$q := 2a/\sigma^2 - 1 > -1 \quad (1.1.3)$$

und

$$c_t := -\frac{2b}{\sigma^2(1 - e^{bt})} > 0. \quad (1.1.4)$$

Die Darstellung (1.1.2) für die Laplace-Transformierte kann verwendet werden, um zu zeigen, daß die Halbgruppe $(Q_t)_{t \geq 0}$ des **CIR**-Prozesses durch eine Lebesgue-Dichte gegeben ist. Dazu definieren wir für $t > 0$ und $x, y \geq 0$

$$p_t(x, y) = p_t(x, y; a, b, \sigma^2) := c_t^{q+1} y^q \exp(-c_t y - c_t e^{bt} x) \cdot \sum_{j=0}^{\infty} \frac{(c_t^2 y x e^{bt})^j}{j! \Gamma(q + j + 1)} \quad (1.1.5)$$

mit q und c_t wie in (1.1.3) und (1.1.4). Offensichtlich ist p_t stetig in (x, y) für jedes feste $t > 0$.

Lemma 1.1.1 *Es gilt*

$$\forall t > 0, x \geq 0 : Q_t(x, dy) = \mathcal{L}(Y_t | P^x)(dy) = p_t(x, y) dy.$$

Beweis Seien $t > 0$, $x \geq 0$ und $\lambda \geq 0$ beliebig. Dann haben wir unter Verwendung der

Substitution $z := (\lambda + c_t)y$

$$\begin{aligned}
\int_0^\infty e^{-\lambda y} p_t(x, y) dy &= c_t^{q+1} \exp(-c_t e^{bt} x) \int_0^\infty y^q e^{-(\lambda+c_t)y} \sum_{j=0}^\infty \frac{(c_t^2 y x e^{bt})^j}{j! \Gamma(q+j+1)} dy \\
&= c_t^{q+1} \exp(-c_t e^{bt} x) \sum_{j=0}^\infty \frac{(c_t^2 x e^{bt})^j}{j! \Gamma(q+j+1)} \int_0^\infty y^{q+j} e^{-(\lambda+c_t)y} dy \\
&= c_t^{q+1} \exp(-c_t e^{bt} x) \sum_{j=0}^\infty \frac{(c_t^2 x e^{bt})^j}{j! \Gamma(q+j+1)} \int_0^\infty \left(\frac{z}{\lambda+c_t}\right)^{q+j} e^{-z} \frac{dz}{\lambda+c_t} \\
&= \left(\frac{c_t}{\lambda+c_t}\right)^{q+1} \exp(-c_t e^{bt} x) \sum_{j=0}^\infty \frac{1}{j!} \left(\frac{c_t^2 x e^{bt}}{\lambda+c_t}\right)^j \frac{1}{\Gamma(q+j+1)} \underbrace{\int_0^\infty z^{q+j} e^{-z} dz}_{=\Gamma(q+j+1)} \\
&= \left(\frac{c_t}{\lambda+c_t}\right)^{q+1} \exp(-c_t e^{bt} x) \exp\left(\frac{c_t^2 x e^{bt}}{\lambda+c_t}\right) \\
&= \psi_{Y_t}(\lambda; x).
\end{aligned}$$

Da ein Wahrscheinlichkeitsmaß durch seine Laplace-Transformierte eindeutig festgelegt ist, folgt die Behauptung. Die Vertauschung von Integration und Grenzübergang in der zweiten Zeile kann dabei wie folgt gerechtfertigt werden: definiere eine Konstante $C_0 = C_0(t; x)$ durch

$$0 < C_0 := \sum_{j=0}^\infty \frac{(2c_t x e^{bt})^j}{\Gamma(q+j+1)} < \infty. \quad (1.1.6)$$

Dann folgt für beliebiges $y \geq 0$ und $n \in \mathbb{N}_0$

$$0 \leq \sum_{j=0}^n \frac{(c_t^2 y x e^{bt})^j}{j! \Gamma(q+j+1)} = \sum_{j=0}^n \underbrace{\frac{(2c_t x e^{bt})^j}{\Gamma(q+j+1)}}_{\leq C_0} \cdot \frac{(c_t y/2)^j}{j!} \leq C_0 \sum_{j=0}^\infty \frac{(c_t y/2)^j}{j!} = C_0 e^{c_t y/2}.$$

Nun ist wegen $c_t > 0$ und $q > -1$ für jedes $\lambda \geq 0$ die Funktion

$$y \mapsto y^q e^{-(\lambda+c_t)y} \cdot C_0 e^{c_t y/2} = C_0 y^q e^{-(\lambda+c_t/2)y}$$

integrierbar auf $[0, \infty)$, also folgt mit dominierter Konvergenz

$$\int_0^\infty y^q e^{-(\lambda+c_t)y} \sum_{j=0}^\infty \frac{(c_t^2 y x e^{bt})^j}{j! \Gamma(q+j+1)} dy = \sum_{j=0}^\infty \frac{(c_t^2 x e^{bt})^j}{j! \Gamma(q+j+1)} \int_0^\infty y^{q+j} e^{-(\lambda+c_t)y} dy.$$

■

Bemerkung Insbesondere folgt aus den obigen Überlegungen auch, daß

$$\begin{aligned}
p_t(x, y) &= c_t^{q+1} y^q \exp(-c_t y - c_t e^{bt} x) \cdot \sum_{j=0}^{\infty} \frac{(c_t^2 y x e^{bt})^j}{j! \Gamma(q + j + 1)} \\
&\leq c_t^{q+1} y^q \exp(-c_t y - c_t e^{bt} x) \cdot C_0(t; x) e^{c_t y/2} \\
&= \underbrace{c_t^{q+1} C_0(t; x) \exp(-c_t e^{bt} x)}_{=: C(t; x)} \cdot y^q e^{-c_t y/2}
\end{aligned} \tag{1.1.7}$$

Dabei ist für festes $t > 0$ die Funktion $x \mapsto C(t; x)$ offenbar stetig. Mit dominierter Konvergenz folgt sofort, daß für jede Borelmenge $A \subseteq \mathbb{R}_{\geq 0}$ und festes $t > 0$ die Funktion

$$x \mapsto Q_t(x, A)$$

stetig ist. ◆

Lassen wir im Ausdruck (1.1.2) für die Laplace-Transformierte $t \rightarrow \infty$ streben, so ergibt sich wegen $c_t \xrightarrow{t \rightarrow \infty} -2b/\sigma^2$

$$\forall x \geq 0, \lambda \geq 0 : \psi_{Y_t}(x; \lambda) \xrightarrow{t \rightarrow \infty} \left(\frac{-2b/\sigma^2}{-2b/\sigma^2 + \lambda} \right)^{q+1}. \tag{1.1.8}$$

Dies ist die Laplace-Transformierte einer Gamma-Verteilung $\Gamma(\alpha, r)$ mit Parametern

$$\alpha := q + 1 = \frac{2a}{\sigma^2}, \quad r := -\frac{2b}{\sigma^2}, \tag{1.1.9}$$

d.h. einer Verteilung mit der Lebesgue-Dichte

$$x \mapsto \mathbf{1}_{(0, \infty)}(x) \frac{r^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-rx}.$$

Mit dem Stetigkeitssatz für die Laplace-Transformierte folgt, daß der Prozeß $(Y_t)_{t \geq 0}$ unter jedem P^x , $x \geq 0$, schwach gegen $\mu(dx) := \Gamma(\alpha, r)(dx)$ konvergiert:

$$\forall x \geq 0 : Y_t \xrightarrow[t \rightarrow \infty]{\mathcal{L}} \mu \quad \text{unter } P^x. \tag{1.1.10}$$

Als Konsequenz hieraus ergibt sich:

Lemma 1.1.2 *Der CIR-Prozeß besitzt ein eindeutig bestimmtes invariantes Wahrscheinlichkeitsmaß, welches durch μ gegeben ist. Für jede Borelmenge $A \subseteq \mathbb{R}_{\geq 0}$ gilt also*

$$\forall t \geq 0 : \mu(A) = \int_{\mathbb{R}_{\geq 0}} \mu(dx) Q_t(x, A). \tag{1.1.11}$$

Beweis Für $t = 0$ ist die Aussage klar. Sei $t > 0$ beliebig. Es genügt, (1.1.11) für Intervalle $A = [0, a]$, $a > 0$ nachzuweisen. Diese sind μ -randlos, da μ als Gammaverteilung keine Punktmassen hat. Nach dem „Portmanteau-Theorem“ folgt also aus (1.1.10)

$$\forall x \geq 0 : Q_s(x, A) \xrightarrow{s \rightarrow \infty} \mu(A). \quad (1.1.12)$$

Für ein beliebiges $x \geq 0$ ergibt sich

$$\mu(A) = \lim_{s \rightarrow \infty} Q_{t+s}(x, A) = \lim_{s \rightarrow \infty} \int_{\mathbb{R}_{\geq 0}} Q_s(x, dy) Q_t(y, A) = \int_{\mathbb{R}_{\geq 0}} \mu(dx) Q_t(x, A),$$

wobei beim letzten Schritt die schwache Konvergenz (1.1.10), die Stetigkeit von $y \mapsto Q_t(y, A)$ sowie $Q_s(x, \cdot) = \mathcal{L}(Y_s | P^x)$ benutzt wurde. Somit ist μ invariant für $(Q_t)_{t \geq 0}$. Ist nun ν ein beliebiges invariantes Maß, so folgt wegen (1.1.12) mit dominierter Konvergenz

$$\nu(A) = \int_{\mathbb{R}_{\geq 0}} \nu(dx) Q_t(x, A) \xrightarrow{t \rightarrow \infty} \int_{\mathbb{R}_{\geq 0}} \nu(dx) \mu(A) = \mu(A),$$

das zeigt die Eindeutigkeit von μ . ■

Wir halten fest, daß eine gemäß μ verteilte Zufallsvariable Z endliche Momente beliebiger Ordnung hat; insbesondere gilt

$$E[Z] = -\frac{a}{b}, \quad \text{Var}[Z] = \frac{a\sigma^2}{2b^2}. \quad (1.1.13)$$

Wir wollen nun die Darstellung (1.1.2) der Laplace-Transformierten verwenden, um zu zeigen, daß der **CIR**-Prozess unter jedem P^x , $x \geq 0$ endliche Momente beliebig hoher Ordnung hat. Dazu benötigen wir die Ableitungen der Laplace-Transformierten nach λ :

Lemma 1.1.3 *Seien $t > 0, x \geq 0$ beliebig. Dann existieren die Ableitungen beliebig hoher Ordnung der Funktion $\lambda \mapsto \psi_{Y_t}(\lambda; x)$ auf $[0, \infty)$ und sind gegeben durch*

$$\begin{aligned} \frac{d^n}{d\lambda^n} \psi_{X(t)}(\lambda; x) &= \psi_{X(t)}(\lambda; x) \cdot \left(\frac{c_t}{c_t + \lambda} \right)^{2n} \cdot p_n(\lambda; x; t) \\ &= \left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp\left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda}\right) \cdot p_n(\lambda; x; t), \quad \lambda \geq 0 \end{aligned} \quad (1.1.14)$$

mit

$$p_n(\lambda; x; t) := (-1)^n \sum_{k=0}^n \binom{n}{k} \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n-k} \prod_{j=1}^k (q + n - k + j), \quad (1.1.15)$$

und der Konvention $\prod_{j=1}^0 \cdots = 1$.

Beweis Wir führen Induktion nach n . Für $n = 0$ ist $p_0(\lambda; x; t) \equiv 1$ und Gleichung (1.1.14) gilt trivialerweise.

Ist nun die Behauptung für ein n schon gezeigt, so haben wir nach Induktionsvoraussetzung

$$\begin{aligned} \frac{d^{n+1}}{d\lambda^{n+1}} \psi_{Y_t}(\lambda; x) &= \frac{d}{d\lambda} \left[\left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \cdot p_n(\lambda; x; t) \right] \\ &= \frac{d}{d\lambda} \left[\left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \right] \cdot p_n(\lambda; x; t) \\ &\quad + \left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \cdot \frac{d}{d\lambda} [p_n(\lambda; x; t)]. \end{aligned} \quad (1.1.16)$$

Nun ist

$$\begin{aligned} &\frac{d}{d\lambda} \left[\left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \right] \\ &= (q+1+2n) \left(\frac{c_t}{c_t + \lambda} \right)^{q+2n} \cdot \frac{-c_t}{(c_t + \lambda)^2} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \\ &\quad + \left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot (-xe^{bt}) \cdot \frac{c_t(c_t + \lambda) - \lambda c_t}{(c_t + \lambda)^2} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \\ &= - \left(\frac{c_t}{c_t + \lambda} \right)^{q+3+2n} \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \cdot \left[(q+2n+1) \frac{c_t + \lambda}{c_t^2} + xe^{bt} \right] \end{aligned} \quad (1.1.17)$$

und

$$\frac{d}{d\lambda} p_n(\lambda; x; t) = (-1)^n \sum_{k=1}^n \binom{n}{k} \cdot k \cdot \left(\frac{c_t + \lambda}{c_t^2} \right)^{k-1} \cdot \frac{1}{c_t^2} \cdot (xe^{bt})^{n-k} \prod_{j=1}^k (q+n-k+j). \quad (1.1.18)$$

Einsetzen von (1.1.17) in (1.1.16) ergibt

$$\begin{aligned} &\frac{d^{n+1}}{d\lambda^{n+1}} \psi_{X(t)}(\lambda; x) \\ &= - \left(\frac{c_t}{c_t + \lambda} \right)^{q+3+2n} \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \cdot \left[(q+2n+1) \frac{c_t + \lambda}{c_t^2} + xe^{bt} \right] \cdot p_n(\lambda; x; t) \\ &\quad + \left(\frac{c_t}{c_t + \lambda} \right)^{q+1+2n} \cdot \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \cdot \frac{d}{d\lambda} p_n(\lambda; x; t) \\ &= \left(\frac{c_t}{c_t + \lambda} \right)^{q+3+2n} \exp \left(-xe^{bt} \frac{\lambda c_t}{c_t + \lambda} \right) \times \\ &\quad \times \left[- \left((q+2n+1) \frac{c_t + \lambda}{c_t^2} + xe^{bt} \right) \cdot p_n(\lambda; x; t) + \left(\frac{c_t + \lambda}{c_t} \right)^2 \cdot \frac{d}{d\lambda} p_n(\lambda; x; t) \right]. \end{aligned} \quad (1.1.19)$$

Die große eckige Klammer in (1.1.19) berechnet sich unter Verwendung von (1.1.18) zu

$$\begin{aligned}
& \left[- \left((q + 2n + 1) \frac{c_t + \lambda}{c_t^2} + xe^{bt} \right) \cdot p_n(\lambda; x; t) + \left(\frac{c_t + \lambda}{c_t} \right)^2 \cdot \frac{d}{d\lambda} p_n(\lambda; x; t) \right] \\
&= - \left((q + 2n + 1) \frac{c_t + \lambda}{c_t^2} + xe^{bt} \right) \cdot (-1)^n \sum_{k=0}^n \binom{n}{k} \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n-k} \prod_{j=1}^k (q + n - k + j) \\
&\quad + \left(\frac{c_t + \lambda}{c_t} \right)^2 \cdot (-1)^n \sum_{k=1}^n \binom{n}{k} \cdot k \cdot \left(\frac{c_t + \lambda}{c_t^2} \right)^{k-1} \cdot \frac{1}{c_t^2} \cdot (xe^{bt})^{n-k} \prod_{j=1}^k (q + n - k + j) \\
&\stackrel{k=i}{=} (-1)^{n+1} (q + 2n + 1) \sum_{i=0}^n \binom{n}{i} \left(\frac{c_t + \lambda}{c_t^2} \right)^{i+1} (xe^{bt})^{n-i} \prod_{j=1}^i (q + n - i + j) \\
&\quad + (-1)^{n+1} \sum_{k=0}^n \binom{n}{k} \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^k (q + n - k + j) \\
&\quad + (-1)^n \sum_{i=1}^n \binom{n}{i} \cdot i \cdot \left(\frac{c_t + \lambda}{c_t^2} \right)^{i+1} \cdot (xe^{bt})^{n-i} \prod_{j=1}^i (q + n - i + j) \\
&\stackrel{i+1=k}{=} (-1)^{n+1} (q + 2n + 1) \sum_{k=1}^{n+1} \binom{n}{k-1} \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^{k-1} (q + n + 1 - k + j) \\
&\quad + (-1)^{n+1} \sum_{k=0}^n \binom{n}{k} \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n+1-k} \underbrace{\prod_{j=1}^k (q + n - k + j)}_{=(q+n-k+1) \cdot \prod_{j=1}^{k-1} (q+n+1-k+j)} \\
&\quad - (-1)^{n+1} \sum_{k=2}^{n+1} \binom{n}{k-1} \cdot (k-1) \cdot \left(\frac{c_t + \lambda}{c_t^2} \right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^{k-1} (q + n + 1 - k + j)
\end{aligned}$$

$$\begin{aligned}
&= (-1)^{n+1}(q+2n+1) \cdot \left[\underbrace{\frac{c_t + \lambda}{c_t^2} (xe^{bt})^n}_{k=1} + \underbrace{\left(\frac{c_t + \lambda}{c_t^2}\right)^{n+1} \prod_{j=1}^n (q+j)}_{k=n+1} \right] \\
&+ (-1)^{n+1} \left[\underbrace{(xe^{bt})^{n+1}}_{k=0} + \underbrace{n \frac{c_t + \lambda}{c_t^2} (xe^{bt})^n (q+n)}_{k=1} \right] \\
&- (-1)^{n+1} n \underbrace{\left(\frac{c_t + \lambda}{c_t^2}\right)^{n+1} \prod_{j=1}^n (q+j)}_{k=n+1} \\
&+ (-1)^{n+1} \sum_{k=2}^n \left(\frac{c_t + \lambda}{c_t^2}\right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^{k-1} (q+n+1-k+j) \times \\
&\quad \times \underbrace{\left[(q+2n+1) \binom{n}{k-1} + \binom{n}{k} (q+n-k+1) - \binom{n}{k-1} (k-1) \right]}_{= \binom{n+1}{k} (q+n+1)} \\
&= (-1)^{n+1} \left[(xe^{bt})^{n+1} + \frac{c_t + \lambda}{c_t^2} (xe^{bt})^n \underbrace{\left((q+2n+1) + n(q+n) \right)}_{=(n+1)(q+n+1)} \right] \\
&\quad + \left(\frac{c_t + \lambda}{c_t^2}\right)^{n+1} \prod_{j=1}^n (q+j) \cdot \underbrace{\left((q+2n+1) - n \right)}_{=q+n+1} \\
&\quad + \sum_{k=2}^n \binom{n+1}{k} \left(\frac{c_t + \lambda}{c_t^2}\right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^k (q+n+1-k+j) \Big] \\
&= (-1)^{n+1} \sum_{k=0}^{n+1} \binom{n+1}{k} \left(\frac{c_t + \lambda}{c_t^2}\right)^k (xe^{bt})^{n+1-k} \prod_{j=1}^k (q+n+1-k+j) \\
&= p_{n+1}(\lambda; x; t).
\end{aligned}$$

Damit ist die Behauptung bewiesen. ■

Als einfache Konsequenz erhalten wir

Satz 1.1.4 *Unter jedem P^x , $x \geq 0$, hat der **CIR**-Prozeß endliche Momente beliebig hoher Ordnung, und es gilt*

$$E^x [Y_t^n] = \sum_{k=0}^n \binom{n}{k} (xe^{bt})^{n-k} \left(\frac{e^{bt}-1}{2b}\right)^k \prod_{j=1}^k (2a + \sigma^2(n-k-1+j)).$$

Beweis Es ist unter Verwendung von Lemma 1.1.3

$$\begin{aligned}
E^x [Y_t^n] &= (-1)^n \left. \frac{d^n \psi_{Y_t}(\lambda; x)}{d\lambda^n} \right|_{\lambda=0} \\
&= (-1)^n p_n(0; x; t) \\
&= \sum_{k=0}^n \binom{n}{k} c_t^{-k} (xe^{bt})^{n-k} \prod_{j=1}^k (q + n - k + j) \\
&= \sum_{k=0}^n \binom{n}{k} (xe^{bt})^{n-k} \left(\frac{e^{bt} - 1}{2b} \right)^k (\sigma^2)^k \prod_{j=1}^k \left(\frac{2a}{\sigma^2} - 1 + n - k + j \right) \\
&= \sum_{k=0}^n \binom{n}{k} (xe^{bt})^{n-k} \left(\frac{e^{bt} - 1}{2b} \right)^k \prod_{j=1}^k (2a + \sigma^2(n - k - 1 + j)).
\end{aligned}$$

■

1.2 Diskretisierung des Modells

Ziel dieser Arbeit ist es, die Parameter a , b und σ^2 des Modells aus (1.1.1) aufgrund von diskreten Beobachtungen des Prozesses $(Y_t)_{t \geq 0}$ zu schätzen. Dazu verwenden wir die Notation

$$\theta := (a, b, \sigma^2) \in \Theta := \mathbb{R}_{>0} \times \mathbb{R}_{<0} \times \mathbb{R}_{>0}.$$

Wir gehen nun davon aus, daß der Prozeß zu bestimmten äquidistanten Zeitpunkten $n \cdot \Delta$, $n \in \mathbb{N}_0$, beobachtet werden kann, wobei die Schrittweite $\Delta > 0$ als bekannt angenommen wird. Definiere den *diskretisierten CIR*-Prozeß durch

$$X_n := Y_{\Delta \cdot n}, \quad n \in \mathbb{N}_0.$$

Dann ist offenbar $(X_n)_{n \in \mathbb{N}_0}$ für jedes $\theta \in \Theta$ ein zeitdiskreter homogener Markov-Prozeß mit Übergangswahrscheinlichkeit

$$Q_\theta(x, dy) = p(x, y; \theta) dy, \quad x \geq 0, \theta \in \Theta, \quad (1.2.1)$$

wobei die Übergangsdichte p durch

$$p(x, y; \theta) := p_\Delta(x, y; a, b, \sigma^2), \quad \theta = (a, b, \sigma^2) \in \Theta \quad (1.2.2)$$

mit $p_\bullet(x, y; a, b, \sigma^2)$ aus (1.1.5) gegeben ist. Für $x, y \in E$ und $\theta \in \Theta$ erhält man aus (1.1.5)

$$p(x, y; \theta) = c(\theta)^{q(\theta)+1} y^{q(\theta)} \exp(-c(\theta)y - c(\theta)e^{b\Delta}x) \cdot \sum_{j=0}^{\infty} \frac{(c(\theta)^2 y x e^{b\Delta})^j}{j! \Gamma(q(\theta) + j + 1)} \quad (1.2.3)$$

mit

$$q(\theta) := \frac{2a}{\sigma^2} - 1 > -1, \quad c(\theta) := -\frac{2b}{\sigma^2(1 - e^{b\Delta})} > 0, \quad \theta \in \Theta.$$

Der diskretisierte Prozeß hat unter $\theta \in \Theta$ das eindeutige invariante Wahrscheinlichkeitsmaß

$$\mu_\theta(dx) = \Gamma(\alpha(\theta), r(\theta))(dx)$$

mit $\alpha(\cdot), r(\cdot)$ aus (1.1.9), aufgefaßt als Funktionen von θ .

Im folgenden Kapitel soll nun das Problem der Parameterschätzung in zeitdiskreten homogenen Markovprozessen zunächst in allgemeiner Form untersucht werden. Auf den **CIR**-Prozeß kommen wir in Kapitel 3 zurück.

Kapitel 2

Parameterschätzung in zeitdiskreten Markovprozessen

2.1 Problemstellung und allgemeine Voraussetzungen

Sei E ein vollständiger separabler metrischer Raum und $\mathcal{B}(E)$ seine Borelsche σ -Algebra. Fasse die Elemente $x \in E$ auf als mögliche Zustände eines Systems, dessen Entwicklung in der Zeit durch eine Wahrscheinlichkeitsverteilung beschrieben wird. Wir nehmen an, daß das System zu den diskreten Zeitpunkten $n = 0, 1, \dots$ beobachtet werden kann. Die Zustände des Systems zu den Zeitpunkten $n \in \mathbb{N}_0$ sind also durch einen stochastischen Prozeß $(X_n)_{n \in \mathbb{N}_0}$, definiert auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) und mit Werten in $(E, \mathcal{B}(E))$, gegeben. Ist das zugrundeliegende Wahrscheinlichkeitsmaß P unbekannt, aber Element einer durch einen d -dimensionalen Parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ beschriebenen Familie $\{P_\theta : \theta \in \Theta\}$ von Wahrscheinlichkeitsmaßen, so liegt ein parametrisches statistisches Modell

$$(\Omega, \mathcal{A}, \{P_\theta : \theta \in \Theta\}) \tag{2.1.1}$$

vor. Wir gehen im gesamten Kapitel davon aus, daß $\Theta \subseteq \mathbb{R}^d$ offen ist. Ziel ist es nun, den Parameter $\theta \in \Theta$ aufgrund einer Stichprobe (X_0, X_1, \dots, X_n) von Beobachtungen des Prozesses zu schätzen. Dieses Problem soll im folgenden untersucht werden unter der Annahme, daß $(X_n)_n$ unter jedem P_θ , $\theta \in \Theta$, ein (zeitdiskreter) homogener Markovprozess mit Zustandsraum E und Übergangswahrscheinlichkeit $Q_\theta(\cdot, \cdot) : E \times \mathcal{B}(E) \rightarrow [0, 1]$ ist; dies schließt insbesondere den Fall eines zu diskreten äquidistanten Zeitpunkten beobachteten Diffusionsprozesses wie in Abschnitt 1.2 ein. Es gelte also

$$\forall A \in \mathcal{B}(E) \text{ fest} : E \ni x \mapsto Q_\theta(x, A) \text{ ist meßbar,}$$

$$\forall x \in E \text{ fest} : \mathcal{B}(E) \ni A \mapsto Q_\theta(x, A) \text{ ist ein Wahrscheinlichkeitsmaß,}$$

und

$$\forall A \in \mathcal{B}(E), n \in \mathbb{N}_0 : Q_\theta(X_n, A) = P_\theta(X_{n+1} \in A \mid \mathcal{F}_n),$$

wobei

$$\mathcal{F}_n := \sigma(X_0, X_1, \dots, X_n)$$

die σ -Algebra der Vergangenheit bis zur Zeit n ist. Die Potenzen von Q_θ , d.h. die n -Schritt-Übergangswahrscheinlichkeiten, werden mit Q_θ^n , $n \in \mathbb{N}$ bezeichnet.

Da für unsere Überlegungen nur die Verteilungen $\mathcal{L}(X_0, X_1, \dots, X_n \mid P_\theta)$ des Prozesses, nicht aber der Grundraum Ω eine Rolle spielen, können wir davon ausgehen, daß $(X_n)_n$ in der folgenden Form gegeben ist: sei (Ω, \mathcal{A}) der kanonische Pfadraum

$$(\Omega, \mathcal{A}) := (\times_{n=0}^\infty E, \otimes_{n=0}^\infty \mathcal{B}(E)).$$

Bekanntlich (siehe z.B. [Rev], S. 17 f.) gibt es dann für jedes Q_θ , $\theta \in \Theta$ und jedes vorgegebene Wahrscheinlichkeitsmaß ν auf $\mathcal{B}(E)$ eine Familie $\{P_\theta^x\}_{x \in E}$ von Wahrscheinlichkeitsmaßen auf (Ω, \mathcal{A}) , so daß für alle $A \in \mathcal{A}$ die Abbildung $x \mapsto P_\theta^x(A)$ $\mathcal{B}(E)$ -meßbar und der kanonische Prozeß $(X_n)_n$ der Koordinatenprojektionen

$$X_n : \Omega \ni \omega = (\omega_1, \omega_2, \dots) \mapsto \omega_n$$

unter P_θ^ν ein homogener Markovprozeß mit Übergangskern Q_θ und Startverteilung ν ist, wobei

$$P_\theta^\nu(A) := \int_E P_\theta^x(A) \nu(dx), \quad A \in \mathcal{A}.$$

Die Startverteilung ν nehmen wir entweder als bekannt oder als ebenfalls durch Θ parametrisiert an: $\nu \in \{\nu_\theta : \theta \in \Theta\}$; z.B. könnte ν_θ das zu Q_θ gehörige invariante Maß sein, falls ein solches existiert.

Im folgenden sagen wir, eine Aussage gelte „ P_θ -fast sicher“ bzw. „unter P_θ “ etc., wenn sie P_θ^ν -fast sicher bzw. unter P_θ^ν für jede mögliche Startverteilung ν gilt. Ist $Z : \Omega \rightarrow \mathbb{R}$ \mathcal{A} -meßbar, so schreiben wir $E_\theta^\nu[Z]$ für den Erwartungswert der Zufallsvariablen Z unter P_θ^ν und analog für bedingte Erwartungen.

Wir nehmen nun weiterhin an, daß für alle $\theta \in \Theta$ zu Q_θ ein invariantes Wahrscheinlichkeitsmaß μ_θ auf $\mathcal{B}(E)$ existiert, d.h. es gilt

$$\forall A \in \mathcal{B}(E) : \mu_\theta(A) = \int_E \mu_\theta(dx) Q_\theta(x, A).$$

Für $\nu = \mu_\theta$ schreiben wir dann kurz P_{μ_θ} und E_{μ_θ} anstelle von $P_\theta^{\mu_\theta}$ und $E_\theta^{\mu_\theta}$. Für jedes $\theta \in \Theta$ ist $(X_n)_n$ unter P_{μ_θ} ein stationärer stochastischer Prozeß mit $\mathcal{L}(X_n \mid P_{\mu_\theta}) = \mu_\theta$ für alle $n \in \mathbb{N}_0$ (siehe z.B. [Rev], S. 60). Wir definieren ein Wahrscheinlichkeitsmaß $\tilde{\mu}_\theta$ auf $(E^2, \mathcal{B}(E^2))$ durch

$$\tilde{\mu}_\theta := \mathcal{L}(X_0, X_1 \mid P_{\mu_\theta}). \quad (2.1.2)$$

Integrale bezüglich $\tilde{\mu}_\theta$ werden im folgenden als $\tilde{\mu}_\theta(f) := \int_{E^2} f(x, y) d\tilde{\mu}_\theta(x, y)$ notiert, falls z.B. $f \in L^1(\tilde{\mu}_\theta)$. Analoges gilt für Integrale bezüglich μ_θ .

Wir setzen nun voraus, daß für jedes $\theta \in \Theta$ die folgende *Ergodizitätsannahme* erfüllt ist:

Annahme 2.1.1 (Ergodizität) *Für jedes $f \in L^1(\tilde{\mu}_\theta)$ gilt ein Starkes Gesetz der großen Zahlen*

$$\frac{1}{n} \sum_{k=1}^n f(X_{k-1}, X_k) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(f) = E_{\mu_\theta}[f(X_0, X_1)] \quad P_\theta\text{-f.s.} \quad (2.1.3)$$

Nach [MT], Theorem 17.1.7 ist Annahme 2.1.1 äquivalent dazu, daß der Prozeß $(X_n)_n$ für jedes $\theta \in \Theta$ positiv Harris-rekurrent ist; für die Definition dieses Begriffs siehe z.B. [MT], Kapitel 9 und 10. Als Spezialfall von (2.1.3) gilt natürlich insbesondere

$$\frac{1}{n} \sum_{k=1}^n f(X_k) \xrightarrow{n \rightarrow \infty} \mu_\theta(f) = E_{\mu_\theta}[f(X_0)] \quad P_\theta\text{-f.s.}$$

für jedes $f \in L^1(\mu_\theta)$.

Der Nachweis der Ergodizitätsannahme 2.1.1 via positiver Harris-Rekurrenz ist mitunter (so z.B. auch für das diskretisierte **CIR**-Modell) aufwendig. Hilfreich ist daher folgende einfache hinreichende Bedingung für Annahme 2.1.1, die in [Bil1], Theorem 1.1, bewiesen wird:

Satz 2.1.2 (Billingsley) *Für jedes $\theta \in \Theta$ sei das invariante Wahrscheinlichkeitsmaß μ_θ zu Q_θ eindeutig bestimmt, und für jedes $x \in E$ sei $Q_\theta(x, dy)$ absolutstetig bezüglich $\mu_\theta(dy)$:*

$$\forall x \in E : Q_\theta(x, dy) \ll \mu_\theta(dy).$$

Dann ist für jedes $\theta \in \Theta$ die Ergodizitätsannahme 2.1.1 erfüllt.

Zusammen mit unseren übrigen Voraussetzungen stellt die Ergodizitätsannahme 2.1.1 auch sicher, daß für jedes $\theta \in \Theta$ die folgende Version des Zentralen Grenzwertsatzes für Martingale gilt (vgl. [Bil2]):

Satz 2.1.3 *Sei $f \in L^2(\tilde{\mu}_\theta)$ so, daß*

$$\forall n \in \mathbb{N} : E_{\mu_\theta}[f(X_{n-1}, X_n) | X_{n-1}] = 0 \quad P_{\mu_\theta}\text{-f.s.} \quad (2.1.4)$$

Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n f(X_{k-1}, X_k) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\mu}_\theta(f^2)) \quad \text{unter } P_\theta. \quad (2.1.5)$$

Beweisskizze Da $(X_n)_n$ positiv Harris-rekurrent ist, genügt es nach [MT], Proposition 17.1.6, die schwache Konvergenz (2.1.5) für den Fall $\nu = \mu_\theta$, also unter $P'_\theta = P_{\mu_\theta}$, nachzuweisen. Nun ist unter P_{μ_θ} der Prozeß $(f(X_{n-1}, X_n))_{n \in \mathbb{N}}$ stationär und ergodisch mit

$$\forall n \in \mathbb{N} : E_{\mu_\theta} [f(X_{n-1}, X_n) | \mathcal{F}_{n-1}] = E_{\mu_\theta} [f(X_{n-1}, X_n) | X_{n-1}] = 0$$

und

$$E_{\mu_\theta} [f(X_0, X_1)^2] = \tilde{\mu}_\theta(f^2) < \infty.$$

Nach [Bil2] gilt somit

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n g(X_{k-1}, X_k) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, E_{\mu_\theta} [g(X_0, X_1)^2]) = \mathcal{N}(0, \tilde{\mu}_\theta(g^2)) \quad \text{unter } P_{\mu_\theta}.$$

Das zeigt die Behauptung.

Bemerkung Wegen $E_{\mu_\theta} [f(X_{n-1}, X_n) | \mathcal{F}_{n-1}] = E_{\mu_\theta} [f(X_{n-1}, X_n) | X_{n-1}]$ für alle $n \in \mathbb{N}$ bedeutet (2.1.4) gerade, daß die Folge der Partialsummen $(\sum_{k=1}^n f(X_{k-1}, X_k))_{n \in \mathbb{N}}$ ein Martingal unter P_{μ_θ} ist. Die Konvergenzaussage (2.1.5) gilt jedoch für beliebige Startverteilungen. Ferner ist für jedes $n \in \mathbb{N}$ nach der Markov-Eigenschaft

$$E_{\mu_\theta} [f(X_{n-1}, X_n) | X_{n-1}] = E_\theta^{X_{n-1}} [f(X_0, X_1)] = \int_E Q_\theta(X_{n-1}, dy) f(X_{n-1}, y) \quad P_{\mu_\theta}\text{-f.s.}$$

Daher und wegen $\mathcal{L}(X_{n-1} | P_{\mu_\theta}) = \mu_\theta$ ist Bedingung (2.1.4) äquivalent zu

$$\int_E Q_\theta(x, dy) f(x, y) = 0 \quad \text{für } \mu_\theta\text{-fast alle } x \in E. \quad (2.1.6)$$

◆

2.2 Schätzgleichungen und Martingal-Schätzfunktionen

Betrachte ein statistisches Modell $(\Omega, \mathcal{A}, \{P_\theta : \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^d$ offen, wie in (2.1.1) mit einem Markovprozess $(X_n)_{n \in \mathbb{N}_0}$ unter den Voraussetzungen aus Abschnitt 2.1. Unser Ziel ist die Schätzung des Parameters $\theta \in \Theta$ aus (2.1.1); insbesondere wollen wir Schätzfolgen mit guten asymptotischen Eigenschaften wie (starke) Konsistenz und asymptotische Normalität finden. Der verwendete Ansatz ist dabei die Methode der Martingal-Schätzfunktionen. Eine *Schätzfunktion* ist eine \mathbb{R}^d -wertige Funktion des Parameters θ und der ersten n Beobachtungen des Prozesses:

$$G_n : \Omega \times \Theta \rightarrow \mathbb{R}^d, \quad G_n(\omega; \theta) = \tilde{G}_n(X_0(\omega), \dots, X_n(\omega); \theta), \quad n \in \mathbb{N}_0, \quad (2.2.1)$$

wobei $\tilde{G}_n : E^{n+1} \times \Theta \rightarrow \mathbb{R}^d$ eine $\mathcal{B}(E^{n+1}) \otimes \mathcal{B}(\Theta)$ -meßbare Funktion ist. Offensichtlich ist dann G_n eine $\mathcal{F}_n \otimes \mathcal{B}(\Theta)$ -meßbare Funktion. Die Komponenten von G_n bezeichnen wir mit $G_{n,i}$, $i = 1, \dots, d$. Der Ansatz besteht nun darin, eine Schätzfolge für den unbekannt Parameter zu konstruieren, indem man für festes $\omega \in \Omega$ nach Lösungen der *Schätzgleichung*

$$G_n(\omega; \theta) = 0, \quad \theta \in \Theta \quad (2.2.2)$$

sucht. Gesucht ist also eine Folge \mathcal{F}_n -meßbarer Funktionen $(\hat{\theta}_n)_{n \in \mathbb{N}_0}$, so daß (zumindest in einem geeigneten asymptotischen Sinne) $G_n(\omega; \hat{\theta}_n(\omega)) = 0$ gilt. Hinreichende Bedingungen für die Existenz von Lösungen der Schätzgleichung (2.2.2) sowie die Asymptotik der so konstruierten Schätzfolgen werden im nächsten Abschnitt untersucht.

Im gesamten Kapitel werden nun folgende Notationen verwendet: Für jedes feste $\theta \in \Theta$ ist die Funktion $\omega \mapsto G_n(\omega, \theta)$ eine \mathbb{R}^d -wertige \mathcal{F}_n -meßbare Zufallsvariable, die wir mit $G_n(\cdot, \theta)$ oder einfach $G_n(\theta)$ bezeichnen. Ebenso ist für festes $\omega \in \Omega$ eine $\mathcal{B}(\Theta)$ -meßbare Funktion $G_n(\omega; \cdot) : \Theta \rightarrow \mathbb{R}^d$ gegeben. Ist diese Funktion in $C^1(\Theta)$, so schreiben wir $\partial_j G_{n,i}(\omega; \theta) := \partial_{\theta_j} G_{n,i}(\omega; \theta)$ für die partiellen Ableitungen von $G_{n,i}(\omega; \cdot)$ nach θ_j an der Stelle $\theta \in \Theta$, $i, j = 1, \dots, d$, sowie $D_\theta G_n(\omega; \theta) := [\partial_j G_{n,i}(\omega; \theta)]_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}$ für die Jacobi-Matrix. Für festes $\theta \in \Theta$ sind die Abbildungen $\partial_j G_{n,i}(\cdot; \theta) : \Omega \rightarrow \mathbb{R}^d$ wiederum \mathcal{F}_n -meßbar, da durch einen Grenzübergang definiert; wir bezeichnen sie kurz mit $\partial_j G_{n,i}(\theta)$. Da für festes $\omega \in \Omega$ nach Voraussetzung $\partial_j G_{n,i}(\omega; \cdot) : \Theta \rightarrow \mathbb{R}^d$ stetig ist, folgt die $\mathcal{F}_n \otimes \mathcal{B}(\Theta)$ -Meßbarkeit von $(\omega; \theta) \mapsto \partial_j G_{n,i}(\omega; \theta)$, siehe Satz A.1. Entsprechendes gilt für höhere Ableitungen, für die wir $\partial_{k_j}^2 G_{n,i}(\omega; \theta) := \partial_{\theta_k} \partial_{\theta_j} G_{n,i}(\omega; \theta)$ etc. schreiben.

Eine Folge $(G_n)_n$ wie in (2.2.1) heißt eine Folge von *Martingal-Schätzfunktionen*, falls für jedes $\theta \in \Theta$ die Folge der Zufallsvariablen $(G_n(\theta))_n$ ein \mathbb{R}^d -wertiges Martingal unter P_{μ_θ} ist, d.h. für alle $n \in \mathbb{N}$ gilt

$$E_{\mu_\theta} [||G_n(\theta)||] < \infty, \quad E_{\mu_\theta} [G_n(\theta) | \mathcal{F}_{n-1}] = G_{n-1}(\theta). \quad (2.2.3)$$

Martingal-Schätzfunktionen wurden eingehend von Bibby und Sørensen in einer mit [BS] beginnenden Serie von Aufsätzen untersucht; ihr großer Vorteil liegt in ihrer relativ einfachen asymptotischen Theorie.

Wir betrachten nun speziell Schätzfunktionen der Bauart

$$G_n(\theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta), \quad (2.2.4)$$

wobei die Funktion $g : E^2 \times \Theta \rightarrow \mathbb{R}^d$ $\mathcal{B}(E^2)$ -meßbar in $(x, y) \in E^2$ und stetig in $\theta \in \Theta$ ist. Dies impliziert die $\mathcal{B}(E^2) \otimes \mathcal{B}(\Theta)$ -Meßbarkeit von g in $(x, y; \theta)$, siehe Satz A.1, so daß G_n tatsächlich eine Schätzfunktion im Sinne von (2.2.1) ist. Für solche G_n ist (2.2.3) offenbar genau dann erfüllt, wenn für $i = 1, \dots, d$ stets $g_i(\cdot, \cdot; \theta) \in L^1(\tilde{\mu}_\theta)$ und

$$\forall n \in \mathbb{N} : E_{\mu_\theta} [g_i(X_{n-1}, X_n; \theta) | X_{n-1}] = 0. \quad (2.2.5)$$

Wegen der Bemerkung nach Satz 2.1.3 ist (2.2.5) wiederum äquivalent zu

$$\int_E Q_\theta(x, dy) g_i(x, y; \theta) = 0 \quad \text{für } \mu_\theta\text{-fast alle } x \in E, \quad (2.2.6)$$

und es gilt

$$\forall \theta \in \Theta, i = 1, \dots, d : \tilde{\mu}_\theta(g_i(\cdot, \cdot; \theta)) = E_{\mu_\theta}[g_i(X_0, X_1)] = 0. \quad (2.2.7)$$

Im folgenden schreiben wir für ein g wie in (2.2.4) zur Vereinfachung der Notation mitunter kurz $\tilde{\mu}_\theta(g_i(\theta))$ anstelle von $\tilde{\mu}_\theta(g_i(\cdot, \cdot; \theta))$.

Beispiele

- Unter der Voraussetzung, daß der Prozeß $(X_n)_n$ unter P_{μ_θ} endliche erste Momente hat, setze

$$m(x; \theta) := E_{\mu_\theta}[X_n | X_{n-1} = x]$$

für eine beliebige Festsetzung der bedingten Erwartung von X_n gegeben $X_{n-1} = x$. Für eine \mathbb{R}^d -wertige Funktion

$$w_m : E \times \Theta \ni (x, \theta) \mapsto w_m(x; \theta) \in \mathbb{R}^d$$

erhält man dann (unter geeigneten Stetigkeits- und Integrierbarkeitsbedingungen an m und w_m) durch die Festsetzung

$$g(x, y; \theta) := (y - m(x; \theta)) \cdot w_m(x; \theta) \quad (2.2.8)$$

Martingal-Schätzfunktionen der Bauart (2.2.4), denn es ist

$$E_{\mu_\theta}[g(X_{n-1}, X_n; \theta) | X_{n-1}] = \underbrace{E_{\mu_\theta}[(X_n - m(X_{n-1}; \theta)) | X_{n-1}]}_{=0} \cdot w_m(X_{n-1}; \theta) = 0 \quad P_{\mu_\theta}\text{-f.s.}$$

Schätzfunktionen dieser Gestalt heißen *linear*. Für $w_m = \nabla_\theta m$ ergibt sich als zugehörige Schätzgleichung die *Least-Squares-Gleichung*, siehe hierzu das nächste Kapitel. Hier und im folgenden schreiben wir ∇_θ für den Gradientenvektor der partiellen Ableitungen nach θ_j , $j = 1, \dots, d$.

Analog sieht man: hat $(X_n)_n$ unter P_{μ_θ} Momente zweiter Ordnung und ist

$$v(x; \theta) := \text{Var}_{\mu_\theta}[X_n | X_{n-1} = x] = E_{\mu_\theta}[(X_n - m(X_{n-1}; \theta))^2 | X_{n-1} = x]$$

die bedingte Varianz von X_n gegeben $X_{n-1} = x$ sowie $w_m, w_v : E \times \Theta \rightarrow \mathbb{R}^d$ geeignet, dann induziert

$$h(x, y; \theta) := (y - m(x; \theta)) \cdot w_m(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta)) \cdot w_v(x; \theta)$$

eine Martingal-Schätzfunktion der Bauart (2.2.4). Solche Schätzfunktionen heißen *quadratisch*. Dabei können die Gewichte w_m und w_v unter geeigneten Bedingungen *optimal* (in einem zu präzisierenden Sinne) gewählt werden, siehe hierzu Abschnitt 5 dieses Kapitels.

- Hat der Prozeß eine stetig differenzierbare, strikt positive Übergangsdichte $p : E^2 \times \Theta \rightarrow \mathbb{R}_{>0}$, d.h. mit einem σ -endlichem Maß λ auf $\mathcal{B}(E)$ gilt

$$\forall \theta \in \Theta, x \in E : Q_\theta(x, dy) = p(x, y; \theta) \lambda(dy),$$

$$\forall (x, y) \in E^2 : \Theta \ni \theta \mapsto p(x, y; \theta) \text{ ist in } C^1(\Theta),$$

so ist die Funktion $\ell : E^2 \times \Theta \rightarrow \mathbb{R}$ mit $\ell(x, y; \theta) := \log p(x, y; \theta)$ wohldefiniert und stetig differenzierbar in θ . Die zu

$$g(x, y; \theta) := \nabla_\theta \ell(x, y; \theta) := [\partial_i \ell(x, y; \theta)]_{i=1, \dots, d}^T \quad (2.2.9)$$

gehörende Schätzgleichung ist die *Likelihood-Gleichung*, siehe Abschnitt 4 unten. Aus der für alle $x \in E$ gültigen Gleichung $\int_E p(x, y; \theta) = 1$ folgt unter geeigneten Voraussetzungen an die Vertauschbarkeit von Integration und Differentiation, daß für alle $x \in E$

$$\int_E \nabla_\theta \ell(x, y; \theta) Q_\theta(x, dy) = \int_E \frac{\nabla_\theta p(x, y; \theta)}{p(x, y; \theta)} \cdot p(x, y; \theta) \lambda(dy) = \nabla_\theta \left[\underbrace{\int_E p(x, y; \theta)}_{\equiv 1} \right] = 0,$$

nach (2.2.6) ist also eine Folge von Martingal-Schätzfunktionen gegeben.

Wir werden im Rest des Kapitels ausschließlich Folgen von Martingal-Schätzfunktionen der Bauart (2.2.4) betrachten, gehen also stets davon aus, daß Bedingung (2.2.5) für die Funktion g erfüllt ist. Für diese Schätzfunktionen gilt nun folgender Zentraler Grenzwertsatz:

Satz 2.2.1 *Sei $(G_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4). Für jedes $\theta \in \Theta$ seien die Funktionen $E^2 \ni (x, y) \mapsto g_i(x, y; \theta)$ in $L^2(\tilde{\mu}_\theta)$, $i = 1 \dots, d$. Dann gilt*

$$\forall \theta \in \Theta : \frac{1}{\sqrt{n}} G_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W(\theta)) \quad \text{unter } P_\theta$$

mit

$$W(\theta) := \tilde{\mu}_\theta (g(\cdot, \cdot; \theta) g(\cdot, \cdot; \theta)^T) = [\tilde{\mu}_\theta (g_i(\cdot, \cdot; \theta) \cdot g_j(\cdot, \cdot; \theta))]_{i, j=1, \dots, d} \in \mathbb{R}^{d \times d}. \quad (2.2.10)$$

Beweis Fixiere ein beliebiges $\theta \in \Theta$. Nach Cramér-Wold reicht es zu zeigen, daß für jedes feste $u \in \mathbb{R}^d$ gilt:

$$\frac{1}{\sqrt{n}} u^T G_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, u^T W(\theta) u) \quad \text{unter } P_\theta.$$

Wegen $g_i(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ für $i = 1, \dots, d$ ist auch die Funktion $h : E^2 \rightarrow \mathbb{R}$,

$$h(x, y) := u^T g(x, y; \theta) = \sum_{i=1}^d u_i g_i(x, y; \theta)$$

in $L^2(\tilde{\mu}_\theta)$. Außerdem gilt nach (2.2.5) für alle $n \in \mathbb{N}$

$$E_{\mu_\theta} [h(X_{n-1}, X_n) | X_{n-1}] = \sum_{i=1}^d u_i \underbrace{E_{\mu_\theta} [g_i(X_{n-1}, X_n; \theta) | X_{n-1}]}_{=0} = 0.$$

Mit dem Zentralen Grenzwertsatz für Martingale 2.1.3 folgt also

$$n^{-1/2} u^T G_n(\theta) = n^{-1/2} \sum_{k=1}^n u^T g(X_{k-1}, X_k; \theta) = n^{-1/2} \sum_{k=1}^n h(X_{k-1}, X_k) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \tilde{\mu}_\theta(h^2)).$$

Dabei ist

$$\tilde{\mu}_\theta(h^2) = \tilde{\mu}_\theta \left(\left[\sum_{i=1}^d u_i g_i(\cdot, \cdot; \theta) \right]^2 \right) = \sum_{i,j=1}^d u_i u_j \cdot \tilde{\mu}_\theta (g_i(\cdot, \cdot; \theta) g_j(\cdot, \cdot; \theta)) = u^T W(\theta) u,$$

womit die Behauptung gezeigt ist. ■

2.3 Existenz und Asymptotik von Lösungen

Betrachte wieder eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4) und die zugehörige Schätzgleichung:

$$G_n(\omega; \theta) = \sum_{k=1}^n g(X_{k-1}(\omega), X_k(\omega); \theta) = 0, \quad (2.3.1)$$

wobei g stetig in θ ist und Bedingung (2.2.5) erfüllt. Wir werden in diesem Abschnitt zeigen, daß unter geeigneten Voraussetzungen, darunter insbesondere Regularitätsbedingungen an g , (stark) konsistente und asymptotisch normale Schätzfolgen existieren, die in einem so gleich zu präzisierenden asymptotischen Sinne die Schätzgleichung (2.3.1) lösen. Die Darstellung in diesem Abschnitt stützt sich wesentlich auf [Sør]. Wir beginnen mit einer Definition:

Definition 2.3.1 Ist $(\hat{\theta}_n)_n$ eine konsistente Schätzfolge mit

$$\forall \theta \in \Theta : P_\theta \left(\left\{ \omega \in \Omega : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0 \right\} \right) \xrightarrow{n \rightarrow \infty} 1, \quad (2.3.2)$$

so nennen wir $(\hat{\theta}_n)_n$ eine konsistente Lösung der Schätzgleichung (2.3.1). Ist $(\hat{\theta}_n)_n$ sogar stark konsistent und gilt

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \left\{ \omega \in \Omega : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0 \right\} \right) = 1, \quad (2.3.3)$$

so nennen wir die Folge eine stark konsistente Lösung von (2.3.1). Eine konsistente Lösung $(\hat{\theta}_n)_n$ heißt eindeutig, falls für jede beliebige konsistente Lösung $(\hat{\vartheta}_n)_n$ gilt

$$\forall \theta \in \Theta : P_\theta \left(\left\{ \omega \in \Omega : \hat{\theta}_n(\omega) = \hat{\vartheta}_n(\omega) \right\} \right) \xrightarrow{n \rightarrow \infty} 1.$$

Entsprechend heißt eine stark konsistente Lösung $(\hat{\theta}_n)_n$ eindeutig, falls für jede stark konsistente Lösung $(\hat{\vartheta}_n)_n$ gilt

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \left\{ \omega \in \Omega : \hat{\theta}_n(\omega) = \hat{\vartheta}_n(\omega) \right\} \right) = 1.$$

Entscheidend für die nun folgende Diskussion ist folgende Begriffsbildung:

Definition 2.3.2 Eine meßbare Funktion $f : E^2 \times \Theta \rightarrow \mathbb{R}$ heißt lokal dominiert integrierbar bezüglich der Familie $(\tilde{\mu}_\theta)_{\theta \in \Theta}$, falls für jedes $\theta \in \Theta$ eine Umgebung $U = U(\theta)$, $\theta \in U \subseteq \Theta$, und eine nichtnegative Funktion $h_\theta : E^2 \rightarrow \mathbb{R}_{\geq 0}$ existiert, so daß gilt:

$$h_\theta \in L^1(\tilde{\mu}_\theta), \quad (2.3.4)$$

$$\text{für } \tilde{\mu}_\theta\text{-fast alle } (x, y) \in E^2 : \sup_{\vartheta \in U} |f(x, y; \vartheta)| \leq h_\theta(x, y). \quad (2.3.5)$$

Bemerkungen Offenbar sind Linearkombinationen lokal dominiert integrierbarer Funktionen wieder lokal dominiert integrierbar. Eine besonders einfache Klasse solcher Funktionen wird gegeben durch den Fall „getrennter Variabler“: ist die Funktion f von der Bauart

$$f(x, y; \theta) = c(\theta) \cdot \tilde{f}(x, y), \quad (2.3.6)$$

wobei $c \in C(\Theta)$ und $\tilde{f} \in L^1(\tilde{\mu}_\theta)$, dann ist f lokal dominiert integrierbar: ist $\theta \in \Theta$ und U eine beliebige kompakte Umgebung von θ mit $U \subseteq \Theta$, so gilt

$$\forall (x, y) \in E^2 : \sup_{\vartheta \in U} |g(x, y; \vartheta)| = \sup_{\vartheta \in U} |c(\vartheta)| \cdot |\tilde{f}(x, y)|$$

mit $C_\theta := \sup_{\vartheta \in U} |c(\vartheta)| < \infty$ aufgrund der Stetigkeit von $c(\cdot)$. Mit $h_\theta := C_\theta \cdot |\tilde{f}|$ sind also (2.3.4) und (2.3.5) erfüllt. Ist

$$f(x, y; \theta) = \sum_{\substack{(\alpha, \beta) \in \mathbb{N}_0^2 \\ \alpha + \beta \leq N}} c_{\alpha, \beta}(\theta) x^\alpha y^\beta$$

ein Polynom vom Grad $\leq N$ in (x, y) mit stetigen Koeffizienten $c_{\alpha, \beta}(\cdot) : \Theta \rightarrow \mathbb{R}$ und hat der Prozeß $(X_n)_n$ unter P_{μ_θ} Momente mindestens der Ordnung $2N$, so ist f lokal dominiert integrierbar: für jedes $(\alpha, \beta) \in \mathbb{N}_0^2$ mit $\alpha + \beta \leq N$ hat $c_{\alpha, \beta}(\theta) x^\alpha y^\beta$ die Gestalt (2.3.6) mit $\tilde{f}(x, y) := x^\alpha y^\beta$ und

$$\tilde{\mu}_\theta \left(\left| \tilde{f}(x, y) \right| \right) = E_{\tilde{\mu}_\theta} \left[\left| X_0^\alpha X_1^\beta \right| \right] \leq (E_{\mu_\theta} [|X_0|^{2\alpha}])^{1/2} \cdot (E_{\mu_\theta} [|X_1|^{2\beta}])^{1/2} \leq E_{\mu_\theta} [|X_0|^{2N}] < \infty.$$

Für den diskretisierten **CIR**-Prozeß ist μ_θ nach Abschnitt 3 stets eine Gamma-Verteilung, und der Prozeß hat somit endliche Momente beliebig hoher Ordnung unter P_{μ_θ} . Daher sind Polynome beliebig hoher Ordnung für den **CIR**-Prozeß stets lokal dominiert integrierbar, sofern die Koeffizienten stetig von θ abhängen. Von dieser Tatsache werden wir im nächsten Kapitel ständig Gebrauch machen. \blacklozenge

Die Bedeutung der lokal dominiert integrierbaren Funktionen für unsere Zwecke liegt in dem folgenden technischen Hilfssatz, von dem wir im weiteren Verlauf der Arbeit mehrfach Gebrauch machen werden:

Lemma 2.3.3 *Sei $g : E^2 \times \Theta \rightarrow \mathbb{R}$ stetig in θ für jedes feste $(x, y) \in E^2$ und lokal dominiert integrierbar bezüglich der Familie $(\tilde{\mu}_\theta)_{\theta \in \Theta}$. Dann gelten für jedes $\theta \in \Theta$ folgende Aussagen:*

1. *Es ist*

$$\lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| = 0 \quad P_\theta\text{-f.s.} \quad (2.3.7)$$

2. *Ist $(\bar{\theta}_n)_{n \in \mathbb{N}_0}$, $\bar{\theta}_n : \Omega \rightarrow \mathbb{R}^d$ eine Folge von Zufallsvariablen mit*

$$\bar{\theta}_n \xrightarrow{n \rightarrow \infty} \theta \quad P_\theta\text{-stochastisch } (P_\theta\text{-f.s.}),$$

so gilt auch

$$\frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \quad P_\theta\text{-stochastisch } (P_\theta\text{-f.s.}). \quad (2.3.8)$$

Bemerkungen Auf der linken Seite von (2.3.7) steht (jedenfalls für genügend kleine δ) eine wohldefinierte Zufallsvariable, denn für beliebiges $(x, y) \in E^2$, $\theta \in \Theta$ und $\delta > 0$ mit $B_\delta(\theta) \subseteq \Theta$ gilt wegen der Stetigkeit von $\vartheta \mapsto g(x, y; \vartheta)$

$$\sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(x, y; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| = \sup_{\vartheta \in B_\delta(\theta) \cap \mathbb{Q}^+} \left| \frac{1}{n} \sum_{k=1}^n g(x, y; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right|.$$

Es gilt tatsächlich folgende etwas stärkere Aussage, die (2.3.7) impliziert:

zu jedem $\varepsilon > 0$ gibt es $\delta = \delta(\varepsilon) > 0$, so daß

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon \quad P_\theta\text{-f.s.} \quad (2.3.9)$$

Setzen wir nämlich $C_{\varepsilon, \delta} := \{ \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon \}$ für $\varepsilon, \delta > 0$ sowie

$$C := \bigcap_{\varepsilon \in \mathbb{Q}^+} \bigcup_{\delta \in \mathbb{Q}^+} C_{\varepsilon, \delta} \in \mathcal{A},$$

so gilt

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \xrightarrow{\delta \downarrow 0} 0 \quad \text{auf } C,$$

und es ist $P_\theta(C) = 1$. ◆

Beweis von Lemma 2.3.3 Fixiere ein beliebiges $\theta \in \Theta$.

- Wir zeigen Aussage (2.3.9), die (2.3.7) impliziert. Sei dazu $\varepsilon > 0$ beliebig. Wähle eine Folge $(\delta_m)_m$ mit $\delta_m \downarrow 0$ für $m \rightarrow \infty$ und $\overline{B_{\delta_1}(\theta)} \subseteq \Theta$. Für jedes feste $(x, y) \in E^2$ und $m \in \mathbb{N}$ ist die Funktion

$$\vartheta \mapsto |g(x, y; \vartheta) - g(x, y; \theta)|$$

stetig auf der kompakten Menge $U_m := \overline{B_{\delta_m}(\theta)} \subseteq \Theta$; daher gibt es ein $\vartheta_m = \vartheta_m(x, y) \in U_m$, so daß

$$|g(x, y; \vartheta_m(x, y)) - g(x, y; \theta)| = \sup_{\vartheta \in U_m} |g(x, y; \vartheta) - g(x, y; \theta)|.$$

Offenbar gilt $\|\vartheta_m(x, y) - \theta\| \leq \delta_m \rightarrow 0$ für $m \rightarrow \infty$. Mit der Stetigkeit von g in ϑ folgt daher

$$\forall (x, y) \in E^2 : \sup_{\vartheta \in U_m} |g(x, y; \vartheta) - g(x, y; \theta)| = |g(x, y; \vartheta_m(x, y)) - g(x, y; \theta)| \xrightarrow{m \rightarrow \infty} 0.$$

Aufgrund der vorausgesetzten lokal dominierten Integrierbarkeit von g gibt es nun eine Umgebung $U \subseteq \Theta$ von θ und eine nichtnegative Funktion h_θ , so daß (2.3.4) und (2.3.5) gelten. Wegen $\delta_m \downarrow 0$ gilt $U_m \subseteq U$ für alle genügend großen m . Für diese m folgt daher

$$\sup_{\vartheta \in U_m} |g(x, y; \vartheta) - g(x, y; \theta)| \leq \sup_{\vartheta \in U} |g(x, y; \vartheta)| + |g(x, y; \theta)| \leq 2h_\theta(x, y)$$

für $\tilde{\mu}_\theta$ -fast alle $(x, y) \in E^2$, und wegen $h_\theta \in L^1(\tilde{\mu}_\theta)$ folgt mit dem Satz über dominierte Konvergenz

$$\tilde{\mu}_\theta \left(\sup_{\vartheta \in U_m} |g(\cdot, \cdot; \vartheta) - g(\cdot, \cdot; \theta)| \right) \xrightarrow{m \rightarrow \infty} 0. \quad (2.3.10)$$

Zu unserem vorgegebenen $\varepsilon > 0$ gibt es folglich ein $m_0 \in \mathbb{N}$, so daß mit $\delta := \delta_{m_0}$ gilt

$$\tilde{\mu}_\theta \left(\sup_{\vartheta \in B_\delta(\theta)} |g(\cdot, \cdot; \vartheta) - g(\cdot, \cdot; \theta)| \right) = \tilde{\mu}_\theta \left(\sup_{\vartheta \in U_{m_0}} |g(\cdot, \cdot; \vartheta) - g(\cdot, \cdot; \theta)| \right) < \varepsilon;$$

insbesondere ist die Funktion $(x, y) \mapsto \sup_{\vartheta \in B_\delta(\theta)} |g(\cdot, \cdot; \vartheta) - g(\cdot, \cdot; \theta)|$ in $L^1(\tilde{\mu}_\theta)$. Nach dem Starken Gesetz der großen Zahlen (2.1.3) folgt daher

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \vartheta)) \right| \\ & \leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n (g(X_{k-1}, X_k; \vartheta) - g(X_{k-1}, X_k; \theta)) + \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \theta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \\ & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sup_{\vartheta \in B_\delta(\theta)} |g(X_{k-1}, X_k; \vartheta) - g(X_{k-1}, X_k; \theta)| \\ & \quad + \limsup_{n \rightarrow \infty} \underbrace{\left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \theta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right|}_{\xrightarrow{n \rightarrow \infty} 0 \text{ } P_\theta\text{-f.s.}} \\ & = \tilde{\mu}_\theta \left(\sup_{\vartheta \in B_\delta(\theta)} |g(\cdot, \cdot; \vartheta) - g(\cdot, \cdot; \theta)| \right) + 0 \\ & < \varepsilon \quad P_\theta\text{-f.s.}, \end{aligned}$$

das ist (2.3.9).

2. Gelte zuerst $\bar{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$ P_θ -fast sicher. Zu beliebigem $\varepsilon > 0$ wähle $\delta = \delta(\varepsilon)$ gemäß (2.3.9). Setze

$$B := \left\{ \lim_{n \rightarrow \infty} \bar{\theta}_n = \theta \right\} \cap \left\{ \lim_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \vartheta)) \right| < \varepsilon \right\}.$$

Dann ist $P_\theta(B) = 1$, und auf der Menge B gilt, da $\bar{\theta}_n \in B_\delta(\theta)$ für alle genügend großen n ,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \\ & \leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon. \end{aligned}$$

Wir haben also gezeigt, daß

$$\forall \varepsilon > 0 : \limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon \quad P_\theta\text{-f.s.} \quad (2.3.11)$$

Schneiden über die abzählbar vielen Mengen von vollem Maß in (2.3.11) zu rationalem $\varepsilon \in \mathbb{Q}^+$ zeigt dann

$$P_\theta \left(\lim_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| = 0 \right) = 1.$$

Gelte nun $\bar{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$ P_θ -stochastisch. Zu beliebigem $\varepsilon > 0$ wähle wieder δ gemäß (2.3.9), so daß

$$P_\theta \left(\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon \right) = 1,$$

also auch

$$P_\theta \left(\sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| < \varepsilon \right) \xrightarrow{n \rightarrow \infty} 1.$$

Auf dem Ereignis $\{\bar{\theta}_n \in B_\delta(\theta)\}$ haben wir die Abschätzung

$$\left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \leq \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right|.$$

Hieraus ergibt sich

$$\begin{aligned} & P_\theta \left(\left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \geq \varepsilon \right) \\ & \leq P_\theta \left(\left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\theta}_n) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \geq \varepsilon, \bar{\theta}_n \in B_\delta(\theta) \right) + P_\theta(\bar{\theta}_n \notin B_\delta(\theta)) \\ & \leq P_\theta \left(\sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)) \right| \geq \varepsilon \right) + P_\theta(\bar{\theta}_n \notin B_\delta(\theta)) \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

und damit die Behauptung.

Wir benötigen noch das folgende, zum Brouwerschen Fixpunktsatz äquivalente Lemma von Aitchison und Silvey ([AS], Lemma 2):

Lemma 2.3.4 (Aitchison & Silvey) Sei $\Theta \subseteq \mathbb{R}^d$ offen, $G : \Theta \rightarrow \mathbb{R}^d$ stetig, $\theta \in \Theta$ fest und $\varepsilon > 0$ so, daß $\overline{B_\varepsilon(\theta)} \subseteq \Theta$. Gilt

$$\forall \vartheta \in \partial B_\varepsilon(\theta) : (\vartheta - \theta)^T G(\vartheta) < 0, \quad (2.3.12)$$

so gibt es einen Punkt $\theta^* \in B_\varepsilon(\theta)$ mit $G(\theta^*) = 0$. Dasselbe gilt, falls $<$ in (2.3.12) durch $>$ ersetzt wird.

Beweis Es gelte (2.3.12). Angenommen, es ist

$$\forall \vartheta \in \overline{B_\varepsilon(\theta)} : G(\vartheta) \neq 0. \quad (2.3.13)$$

Dann wird durch

$$F(\vartheta) := \frac{G(\vartheta)}{\|G(\vartheta)\|} \cdot \varepsilon + \theta$$

eine stetige Funktion $F : \overline{B_\varepsilon(\theta)} \rightarrow \overline{B_\varepsilon(\theta)}$ definiert. Nach dem Brouwerschen Fixpunktsatz hat F einen Fixpunkt, d.h. es gibt ein $\bar{\theta} \in \overline{B_\varepsilon(\theta)}$ mit

$$\bar{\theta} = F(\bar{\theta}) = \frac{G(\bar{\theta})}{\|G(\bar{\theta})\|} \cdot \varepsilon + \theta.$$

Offensichtlich ist $\|\bar{\theta} - \theta\| = \|F(\bar{\theta}) - \theta\| = \varepsilon$, also $\bar{\theta} \in \partial B_\varepsilon(\theta)$. Andererseits gilt aber

$$(\bar{\theta} - \theta)^T G(\bar{\theta}) = (\bar{\theta} - \theta)^T (\bar{\theta} - \theta) \cdot \frac{\|G(\bar{\theta})\|}{\varepsilon} \geq 0$$

im Widerspruch zu (2.3.12). Damit ist die Annahme (2.3.13) ad absurdum geführt, es gibt also ein $\theta^* \in \overline{B_\varepsilon(\theta)}$ mit $G(\theta^*) = 0$. Wäre $\theta^* \in \partial B_\varepsilon(\theta)$, so ergäbe sich aus $(\theta^* - \theta)^T G(\theta^*) = (\theta^* - \theta)^T 0 = 0$ wiederum ein Widerspruch zu (2.3.12). Daher folgt $\theta^* \in B_\varepsilon(\theta)$, und die Behauptung ist gezeigt.

Gilt nun (2.3.12) mit „ $>$ “ statt „ $<$ “, so folgt die Behauptung aus dem schon Bewiesenen angewandt auf die Funktion $\tilde{G} := -G$. ■

Nach diesen Vorbereitungen können nun Regularitätsbedingungen an die Funktion g aus (2.2.4) formuliert werden, die die Durchführung des zu Anfang dieses Abschnitts beschriebenen Programms ermöglichen. Diese Voraussetzungen sind im wesentlichen von zweierlei Typ: zunächst werden Differenzierbarkeitsbedingungen an g und Integrierbarkeitsbedingungen für die entsprechenden partiellen Ableitungen benötigt. Im einzelnen definieren wir:

Bedingung C1 Die Funktion g erfüllt Bedingung **C1**, falls gilt:

- für alle $(x, y) \in E^2$ ist die Funktion $\Theta \ni \theta \mapsto g(x, y; \theta)$ in $C^1(\Theta; \mathbb{R}^d)$;
- für alle $i, j = 1, \dots, d$ ist die Funktion

$$\partial_j g_i : E^2 \times \Theta \rightarrow \mathbb{R}^d$$

lokal dominiert integrierbar.

Bedingung C2 Die Funktion g erfüllt Bedingung **C2**, falls gilt:

- für alle $(x, y) \in E^2$ ist die Funktion $\Theta \ni \theta \mapsto g(x, y; \theta)$ in $C^2(\Theta; \mathbb{R}^d)$;
- für alle $i, j, k = 1, \dots, d$ sind die Funktionen

$$g_k, \partial_j g_k, \partial_{ij}^2 g_k : E^2 \times \Theta \rightarrow \mathbb{R}^d$$

lokal dominiert integrierbar.

Unter jeder der Bedingungen **C1** und **C2** ist dann für jedes $\theta \in \Theta$ die Matrix

$$V(\theta) := \tilde{\mu}_\theta (D_\theta g(\cdot, \cdot; \theta)) = [\tilde{\mu}_\theta (\partial_j g_i(\cdot, \cdot; \theta))]_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d} \quad (2.3.14)$$

definiert. Der zweite Typ von benötigten Voraussetzungen besteht in nun Regularitätsbedingungen an $V(\theta)$:

Bedingung I Die Funktion g erfüllt die Bedingung **I**, falls die Matrix $V(\theta)$ aus (2.3.14) für jedes $\theta \in \Theta$ invertierbar ist.

Bedingung D Die Funktion g erfüllt die Bedingung **D**, falls die Matrix $V(\theta)$ aus (2.3.14) für jedes $\theta \in \Theta$ positiv definit oder negativ definit ist.

Die vorstehenden Bedingungen wurden vorgeschlagen in [Sør]. Wir untersuchen nun zuerst die Frage, wann eine vorgegebene konsistente Lösung der Schätzgleichung (2.3.1) asymptotisch normal ist.

Satz 2.3.5 Sei $(G_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4). Die Funktion g erfülle die Voraussetzungen **C1** und **I**, und für alle $\theta \in \Theta$ und $i = 1, \dots, d$ sei die Funktion $(x, y) \mapsto g_i(x, y; \theta)$ in $L^2(\tilde{\mu}_\theta)$. Ist dann $(\hat{\theta}_n)_n$ eine konsistente Lösung der Schätzgleichung $G_n(\omega; \theta) = 0$, so gilt

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V(\theta)^{-1}W(\theta)V(\theta)^{-T}) \quad \text{unter } P_\theta$$

mit $W(\theta)$ aus (2.2.10) und $V(\theta)$ aus (2.3.14).

Beweis Wir schreiben $\hat{\theta}_{n,i}$ für die Komponenten von $\hat{\theta}_n$. Fixiere ein beliebiges $\theta \in \Theta$.

1. Sei $\delta > 0$ so, daß $B_\delta(\theta) \subseteq \Theta$. Aufgrund der Voraussetzung **C1** ist für jedes $\omega \in \Omega$ die Funktion

$$B_\delta(\theta) \ni \vartheta \mapsto G_n(\omega; \vartheta)$$

stetig differenzierbar. Für $\omega \in B_n := \{\hat{\theta}_n \in B_\delta(\theta)\}$ liegt die Verbindungsstrecke zwischen $\hat{\theta}_n(\omega)$ und θ ganz in Θ . Für diese ω ergibt folglich eine Taylorentwicklung der i -ten Komponente von $G_n(\omega; \cdot)$ um den Punkt θ

$$G_{n,i}(\omega; \hat{\theta}_n(\omega)) = G_{n,i}(\omega; \theta) + \sum_{j=1}^d \partial_j G_{n,i}(\omega; \theta + \lambda_n^{(i)}(\omega)(\hat{\theta}_n(\omega) - \theta)) (\hat{\theta}_{n,j}(\omega) - \theta_j) \quad (2.3.15)$$

mit einer „zufälligen“ Zahl $\lambda_n^{(i)}(\omega) \in [0, 1]$. Für $\omega \in B_n^c$ definieren wir $\lambda_n^{(i)}(\omega) := 0$. Die Abbildung $\lambda_n^{(i)} : \Omega \rightarrow [0, 1]$ kann meßbar, also als eine Zufallsvariable festgelegt werden: dazu betrachte für festes $n \in \mathbb{N}$ und $i \in \{1, \dots, d\}$ die Funktion $h : \Omega \times [0, 1] \rightarrow \mathbb{R}^d$ mit

$$h(\omega; \lambda) := \mathbf{1}_{B_n}(\omega) \cdot \left(G_{n,i}(\omega; \hat{\theta}_n(\omega)) - G_{n,i}(\omega; \theta) - \sum_{j=1}^d \partial_j G_{n,i}(\omega; \theta + \lambda \cdot (\hat{\theta}_n(\omega) - \theta)) \right)$$

sowie die mengenwertige Abbildung $\Lambda : \Omega \rightarrow \mathcal{P}(\mathbb{R})$,

$$\Lambda(\omega) := \begin{cases} \{\lambda \in [0, 1] : h(\omega; \lambda) = 0\} & \text{falls } \omega \in B_n \\ \{0\} & \text{falls } \omega \in B_n^c. \end{cases}$$

Aufgrund des Satzes von Taylor ist $\Lambda(\omega) \subseteq [0, 1]$ für jedes $\omega \in \Omega$ stets nichtleer und offensichtlich auch abgeschlossen und somit kompakt, da die Funktion $\lambda \mapsto h(\omega; \lambda)$ stetig ist. Außerdem haben wir für jede offene Teilmenge $U \subseteq \mathbb{R}$

$$\begin{aligned} \Lambda^-(U) &:= \{\omega \in \Omega : \Lambda(\omega) \cap U \neq \emptyset\} \\ &= \left[B_n \cap \underbrace{\{\omega \in \Omega : \exists \lambda \in U \text{ mit } h(\omega; \lambda) = 0\}}_{\in \mathcal{F}_n \text{ nach Lemma A.1}} \right] \cup \left[B_n^c \cap \underbrace{\{\omega \in \Omega : 0 \in U\}}_{=\emptyset \text{ oder } =\Omega} \right] \in \mathcal{F}_n \end{aligned}$$

nach Lemma A.1, da die Funktion h \mathcal{F}_n -meßbar in ω und stetig in λ ist. Nach dem Measurable Selection Theorem A.2 gibt es daher eine \mathcal{F}_n -meßbare Funktion $\lambda_n^{(i)} : \Omega \rightarrow \mathbb{R}$ mit $\lambda_n^{(i)}(\omega) \in \Lambda(\omega)$ für alle $\omega \in \Omega$, d.h. (2.3.15) gilt.

Wir definieren nun für $n \in \mathbb{N}$ matrixwertige Abbildungen $V^{(n)} : \Omega \rightarrow \mathbb{R}^{d \times d}$ durch

$$V^{(n)}(\omega) := \left[\partial_j G_{n,i}(\omega; \theta + \lambda_n^{(i)}(\omega)(\hat{\theta}_n(\omega) - \theta)) \right]_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}.$$

[Beachte, daß $V^{(n)}(\omega)$ auch für $\omega \in B_n^c$ definiert ist.] Wegen der $\mathcal{F}_n \otimes \mathcal{B}(\Theta)$ -Meßbarkeit der Funktionen $(\omega; \vartheta) \mapsto \partial_j G_{n,i}(\omega; \vartheta)$ sowie der \mathcal{F}_n -Meßbarkeit von $\lambda_n^{(i)}(\cdot)$ und $\hat{\theta}_n(\cdot)$ ist dann auch $V^{(n)}$ eine wohldefinierte Zufallsvariable, und (2.3.15) schreibt sich als Vektorgleichung in der Form

$$G_n(\omega; \hat{\theta}_n(\omega)) = G_n(\omega; \theta) + V^{(n)}(\omega)(\hat{\theta}_n(\omega) - \theta), \quad \omega \in B_n. \quad (2.3.16)$$

2. Wir zeigen nun: es gilt

$$\frac{1}{n}V^{(n)} \xrightarrow{n \rightarrow \infty} V(\theta) \quad P_\theta\text{-stochastisch}; \quad (2.3.17)$$

aufgrund der im ersten Schritt angestellten Vorüberlegungen zur Meßbarkeit ist diese Aussage sinnvoll. Es genügt zu zeigen, daß für alle $i, j = 1, \dots, d$

$$\frac{1}{n}V_{ij}^{(n)} \rightarrow V_{ij}(\theta) \quad P_\theta\text{-stochastisch.}$$

Seien dazu $1 \leq i, j \leq d$ fest. Wegen $\lambda_n^{(i)}(\omega) \in [0, 1]$ und der Konsistenz von $(\hat{\theta}_n)_n$ gilt auch

$$\theta + \lambda_n^{(i)} \cdot (\hat{\theta}_n - \theta) \xrightarrow{n \rightarrow \infty} \theta \quad P_\theta\text{-stochastisch.}$$

Da nach Voraussetzung die Funktion $(x, y; \vartheta) \mapsto \partial_j g_i(x, y; \vartheta)$ lokal dominiert integrierbar ist, folgt also aus Lemma 2.3.3, (2.3.8) auch P_θ -stochastische Konvergenz

$$\frac{1}{n}V_{ij}^{(n)} = \frac{1}{n} \sum_{k=1}^n \partial_j g_i \left(X_{k-1}, X_k; \theta + \lambda_n^{(i)}(\hat{\theta}_n - \theta) \right) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(\partial_j g_i(\cdot, \cdot; \theta)) = V_{ij}(\theta),$$

das ist (2.3.17).

3. Bezeichne \mathcal{G} die Gruppe der invertierbaren $d \times d$ -Matrizen. Nach Voraussetzung **I** ist $V(\theta) \in \mathcal{G}$. Da \mathcal{G} offen in $\mathbb{R}^{d \times d}$ ist, gibt es also ein $\varepsilon > 0$, so daß $\{A \in \mathbb{R}^{d \times d} : |||A - V(\theta)||| < \varepsilon\} \subseteq \mathcal{G}$, wobei $|||\cdot|||$ eine beliebige Matrixnorm auf $\mathbb{R}^{d \times d}$ ist. Mit $C_n := B_n \cap \{\frac{1}{n}V^{(n)} \text{ ist invertierbar}\}$ gilt dann wegen $P_\theta(B_n) \xrightarrow{n \rightarrow \infty} 1$ und der im zweiten Schritt gezeigten stochastischen Konvergenz $\frac{1}{n}V^{(n)} \rightarrow V(\theta)$

$$P_\theta(C_n) \geq P_\theta \left(B_n \cap \left\{ \left\| \frac{1}{n}V^{(n)} - V(\theta) \right\| < \varepsilon \right\} \right) \xrightarrow{n \rightarrow \infty} 1 \quad (2.3.18)$$

und somit P_θ -stochastische, also auch schwache Konvergenz

$$\mathbf{1}_{C_n} \left[\frac{1}{n}V^{(n)} \right] + \mathbf{1}_{C_n^c} E \xrightarrow[n \rightarrow \infty]{\mathcal{L}} V(\theta) \quad \text{unter } P_\theta, \quad (2.3.19)$$

wobei $E \in \mathbb{R}^{d \times d}$ die d -dimensionale Einheitsmatrix bezeichnet. Da die Zufallsvariable auf der linken Seite von (2.3.19) ausschließlich Werte in \mathcal{G} annimmt und die Inversenbildung in dieser Gruppe eine stetige Operation ist, folgt mit dem Satz über die stetige Abbildung

$$\mathbf{1}_{C_n} [n(V^{(n)})^{-1}] + \mathbf{1}_{C_n^c} E \xrightarrow[n \rightarrow \infty]{\mathcal{L}} V(\theta)^{-1} \quad \text{unter } P_\theta,$$

da $V(\theta)^{-1}$ deterministisch und wegen (2.3.18) also auch

$$\mathbf{1}_{C_n} [n(V^{(n)})^{-1}] \xrightarrow[n \rightarrow \infty]{} V(\theta)^{-1} \quad P_\theta\text{-stochastisch.} \quad (2.3.20)$$

4. Wir schreiben (2.3.16) in der Form

$$\frac{1}{\sqrt{n}} G_n(\omega; \hat{\theta}_n(\omega)) - \frac{1}{\sqrt{n}} G_n(\omega; \theta) = \left[\frac{1}{n} V^{(n)}(\omega) \right] \sqrt{n}(\hat{\theta}_n(\omega) - \theta), \quad \omega \in C_n. \quad (2.3.21)$$

Da nach Voraussetzung alle Funktionen $(x, y) \mapsto g_i(x, y; \theta)$ in $L^2(\tilde{\mu}_\theta)$ sind, folgt aus dem Zentralen Grenzwertsatz 2.2.1

$$\frac{1}{\sqrt{n}} G_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W(\theta)) \quad \text{unter } P_\theta$$

mit $W(\theta)$ aus (2.2.10). Da $(\hat{\theta}_n)_n$ eine konsistente Lösung der Schätzgleichung $G_n(\theta) = 0$ ist, gilt (2.3.2) und somit

$$\mathbf{1}_{C_n} \cdot G_n(\hat{\theta}_n) \xrightarrow[n \rightarrow \infty]{} 0 \quad P_\theta\text{-stochastisch.} \quad (2.3.22)$$

Ein Standard-Slutsky-Argument liefert folglich

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta) &= \mathbf{1}_{C_n} \sqrt{n}(\hat{\theta}_n - \theta) + \mathbf{1}_{C_n^c} \sqrt{n}(\hat{\theta}_n - \theta) \\ &= \underbrace{\mathbf{1}_{C_n} [n(V^{(n)})^{-1}]}_{\rightarrow V(\theta)^{-1} P_\theta\text{-stoch.}} \underbrace{\left(\frac{1}{\sqrt{n}} G_n(\hat{\theta}_n) - \frac{1}{\sqrt{n}} G_n(\theta) \right)}_{\xrightarrow{\mathcal{L}} \mathcal{N}(0, W(\theta))} + \underbrace{\mathbf{1}_{C_n^c} \sqrt{n}(\hat{\theta}_n - \theta)}_{\rightarrow 0 P_\theta\text{-stoch.}} \\ &\xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V(\theta)^{-1} W(\theta) V(\theta)^{-T}) \quad \text{unter } P_\theta. \end{aligned}$$

Damit ist der Beweis abgeschlossen. ■

Satz 2.3.5 stellt unter den Voraussetzungen **C1** und **I** sicher, daß eine vorgegebene konsistente Lösung der Schätzgleichung (2.3.1) asymptotisch normal mit Konvergenzgeschwindigkeit \sqrt{n} ist. Diese Voraussetzungen sichern jedoch noch nicht die Existenz von Lösungen der Schätzgleichung. Hierzu muß entweder **C1** zu **C2** oder **I** zu **D** verschärft werden. Wir zeigen nun, daß unter jeder der beiden resultierenden Kombinationen von Bedingungen P_θ -fast sicher „für schließlich alle n “ Lösungen der Schätzgleichung (2.3.1) sogar in beliebig kleinen Umgebungen von $\theta \in \Theta$ existieren, unter jedem der Wahrscheinlichkeitsmaße P_θ .

Satz 2.3.6 Sei $(G_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4), d.h.

$$G_n(\cdot; \theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta).$$

Die Funktion g genüge den Bedingungen **C1** und **D** oder den Bedingungen **C2** und **I**. Dann gilt

$$\forall \theta \in \Theta, \forall \varepsilon > 0 : P_\theta \left(\liminf_{n \rightarrow \infty} \{ \omega \in \Omega : \exists \vartheta \in B_\varepsilon(\theta) \cap \Theta \text{ mit } G_n(\omega; \vartheta) = 0 \} \right) = 1. \quad (2.3.23)$$

Beweis Für jedes $n \in \mathbb{N}$ ist die Menge $\{ \omega \in \Omega : \exists \vartheta \in B_\varepsilon(\theta) \cap \Theta \text{ mit } G_n(\omega; \vartheta) = 0 \}$ ein Ereignis in \mathcal{F}_n , da die Funktion G_n \mathcal{F}_n -meßbar in ω und stetig in ϑ ist (siehe Satz A.1). Daher ist die Aussage (2.3.23) sinnvoll.

Fixiere nun ein beliebiges $\theta \in \Theta$. Sei $\varepsilon > 0$ beliebig. Da Θ offen ist, können wir o.B.d.A. $B_\varepsilon(\theta) \subseteq \Theta$ annehmen.

1. Seien zunächst **C1** und **D** vorausgesetzt. Wir betrachten nun wieder für festes $\omega \in \Omega$ eine Taylorentwicklung der Funktion $B_\varepsilon(\Theta) \ni \vartheta \mapsto G_{n,i}(\omega; \vartheta)$ um den Punkt θ , diesmal mit Integral-Restglied:

$$G_{n,i}(\omega; \vartheta) = G_{n,i}(\omega; \theta) + \sum_{j=1}^d (\vartheta_j - \theta_j) \int_0^1 \partial_j G_{n,i}(\omega; \theta + t \cdot (\vartheta - \theta)) dt \quad (2.3.24)$$

Ähnlich wie im Beweis von Satz 2.3.5 definieren wir für $n \in \mathbb{N}$ matrixwertige Funktionen $R^{(n)} : \Omega \times B_\varepsilon(\theta) \rightarrow \mathbb{R}^{d \times d}$ durch

$$R^{(n)}(\omega; \vartheta) := \int_0^1 D_\theta G_n(\omega; \theta + t \cdot (\vartheta - \theta)) dt,$$

wobei das Integral komponentenweise zu verstehen ist. Damit schreiben sich die d Gleichungen (2.3.24) als Vektorgleichung in der Form

$$\begin{aligned} G_n(\omega; \vartheta) &= G_n(\omega; \theta) + [R^{(n)}(\omega; \vartheta)] (\vartheta - \theta) \\ &= G_n(\omega; \theta) + nV(\theta)(\vartheta - \theta) + n \left(\frac{1}{n} R^{(n)}(\omega; \vartheta) - V(\theta) \right) (\vartheta - \theta). \end{aligned} \quad (2.3.25)$$

Nach Voraussetzung ist $g_i(\cdot, \cdot; \theta) \in L^1(\tilde{\mu}_\theta)$ sowie $\tilde{\mu}_\theta(g_i(\cdot, \cdot; \theta)) = 0$ für alle $i = 1, \dots, d$, da wir es mit einer Folge von Martingal-Schätzfunktionen zu tun haben vgl. (2.2.5), (2.2.7). Also folgt aus dem starken Gesetz der großen Zahlen (2.1.3)

$$\frac{1}{n} G_{n,i}(\theta) = \frac{1}{n} \sum_{k=1}^n g_i(X_{k-1}, X_k; \theta) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(g_i(\cdot, \cdot; \theta)) = 0 \quad P_\theta\text{-f.s.},$$

also auch

$$\frac{1}{n}G_n(\theta) \xrightarrow{n \rightarrow \infty} 0 \quad P_\theta\text{-f.s.} \quad (2.3.26)$$

Wegen Bedingung **D** können wir o.B.d.A. davon ausgehen, daß die Matrix $V(\theta)$ aus (2.3.14) positiv definit ist; der negativ definite Fall läßt sich analog behandeln. Sei $\lambda > 0$ der kleinste Eigenwert von $V(\theta)$. Da nach Voraussetzung **C1** alle Funktionen $\partial_j g_i$ lokal dominiert integrierbar sind, gilt nach Lemma 2.3.3 für $i, j = 1, \dots, d$

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_j G_{n,i}(\vartheta) - V_{ij}(\theta) \right| \\ &= \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n \partial_j g_i(X_{k-1}, X_k; \vartheta) - \tilde{\mu}_\theta(\partial_j g_i(\cdot, \cdot; \theta)) \right| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \end{aligned} \quad (2.3.27)$$

Ist nun $\|\cdot\|$ eine beliebige Norm auf \mathbb{R}^d und bezeichnet $\|\cdot\|$ die durch $\|\cdot\|$ induzierte Matrixnorm auf $\mathbb{R}^{d \times d}$, so folgt aus (2.3.27) wegen der Äquivalenz aller Normen auf $\mathbb{R}^{d \times d}$ auch

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\vartheta) - V(\theta) \right\| \\ & \leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\{ \text{const} \cdot \sum_{i,j=1}^d \left| \frac{1}{n} \partial_j G_{n,i}(\vartheta) - V_{ij}(\theta) \right| \right\} \\ & \leq \text{const} \cdot \sum_{i,j=1}^d \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_j G_{n,i}(\vartheta) - V_{ij}(\theta) \right| \\ & \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \end{aligned} \quad (2.3.28)$$

Für beliebiges $\delta > 0$ und jedes $\bar{\theta} \in \partial B_\delta(\theta)$ gilt offenbar

$$\begin{aligned} \left\| \frac{1}{n} R^{(n)}(\omega; \bar{\theta}) - V(\theta) \right\| &= \left\| \int_0^1 \left(\frac{1}{n} D_\theta G_n(\omega; \theta + t(\bar{\theta} - \theta)) - V(\theta) \right) dt \right\| \\ &\leq \int_0^1 \left\| \frac{1}{n} D_\theta G_n(\omega; \theta + t(\bar{\theta} - \theta)) - V(\theta) \right\| dt \\ &\leq \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\omega; \vartheta) - V(\theta) \right\|. \end{aligned}$$

Setzen wir nun

$$C := \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} G_n(\theta) = 0 \right\} \cap \left\{ \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\vartheta) - V(\theta) \right\| = 0 \right\},$$

so folgt wegen (2.3.26) und (2.3.28), daß $P_\theta(C) = 1$.

Sei nun $\omega \in C$ beliebig. Dann gibt es zu unserem vorgegebenen ε aus (2.3.23) ein $\delta = \delta(\omega)$ mit $0 < \delta < \varepsilon$ sowie $n_0 = n_0(\omega) \in \mathbb{N}$, so daß für alle $n \geq n_0$

$$\|G_n(\omega; \theta)\| < n\delta \cdot \lambda/3, \quad \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\omega; \vartheta) - V(\theta) \right\| < \lambda/3.$$

Sei nun $n \geq n_0$ beliebig. Dann haben wir für jedes feste $\bar{\theta} \in \partial B_\delta(\theta) \subseteq B_\varepsilon(\theta)$ die Abschätzung

$$\begin{aligned} (\bar{\theta} - \theta)^T G_n(\omega; \bar{\theta}) &= (\bar{\theta} - \theta)^T G_n(\omega; \theta) + n(\bar{\theta} - \theta)^T V(\theta)(\bar{\theta} - \theta) \\ &\quad + n(\bar{\theta} - \theta)^T \left(\frac{1}{n} R^{(n)}(\omega; \bar{\theta}) - V(\theta) \right) (\bar{\theta} - \theta) \\ &\geq -\|G_n(\omega; \theta)\| \cdot \underbrace{\|\bar{\theta} - \theta\|}_{=\delta} + n\lambda \|\bar{\theta} - \theta\|^2 \\ &\quad - n \left\| \frac{1}{n} R^{(n)}(\omega; \bar{\theta}) - V(\theta) \right\| \cdot \|\bar{\theta} - \theta\|^2 \\ &\geq n\lambda\delta^2 - \|G_n(\omega; \theta)\| \cdot \delta \\ &\quad - n \cdot \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\omega; \vartheta) - V(\theta) \right\| \cdot \delta^2 \\ &\geq n\lambda\delta^2 - n\delta\lambda/3 \cdot \delta - n\lambda/3 \cdot \delta^2 \\ &= n\delta^2\lambda/3 \\ &> 0, \end{aligned} \tag{2.3.29}$$

und da die Funktion $\vartheta \mapsto G_n(\omega; \vartheta)$ stetig auf Θ ist, folgt aus Lemma 2.3.4 die Existenz eines $\vartheta^* \in B_\delta(\theta) \subseteq B_\varepsilon(\theta)$ mit $G_n(\omega; \vartheta^*) = 0$. Da $\omega \in C$ und $n \geq n_0$ beliebig waren, ist also gezeigt:

$$C \subseteq \liminf_{n \rightarrow \infty} \{ \omega : \exists \vartheta \in B_\varepsilon(\theta) \cap \Theta : G_n(\omega; \vartheta^*) = 0 \},$$

und wegen $P_\theta(C) = 1$ ist damit die Behauptung bewiesen.

2. Seien nun **C2** und **I** vorausgesetzt, d.h. die Matrix $V(\theta)$ ist invertierbar, aber nicht mehr notwendig definit. Die Beweisidee ist die folgende: wegen der Invertierbarkeit von $V(\theta)$ hat die Schätzfunktion

$$\tilde{G}_n(\theta) := V(\theta)^T G_n(\theta) = \sum_{k=1}^n V(\theta)^T g(X_{k-1}, X_k; \theta) =: \sum_{k=1}^n \tilde{g}(X_{k-1}, X_k; \theta)$$

mit $\tilde{g}(x, y; \theta) := V(\theta)^T g(x, y; \theta)$ dieselben Nullstellen wie G_n , und falls die Funktion $A : \Theta \rightarrow \mathbb{R}^{d \times d}$, $\theta \mapsto V(\theta)^T$ stetig differenzierbar ist, folgt

$$\frac{1}{n} D_\theta \tilde{G}_n(\theta) = [D_\theta A(\theta)] \underbrace{\frac{1}{n} G_n(\theta)}_{\rightarrow 0} + V(\theta)^T \underbrace{\frac{1}{n} D_\theta G_n(\theta)}_{\rightarrow V(\theta)} \xrightarrow{n \rightarrow \infty} V(\theta)^T V(\theta) \quad P_\theta\text{-f.s.}$$

[Hierbei ist $D_\theta A(\theta)$ die totale Ableitung von A an der Stelle θ , also ein beschränkter linearer Operator $\Theta \rightarrow \mathbb{R}^{d \times d}$.] Dabei ist $V(\theta)^T V(\theta)$ positiv definit, d.h. \tilde{g} erfüllt die Bedingung **D**. Diese Überlegung legt nahe, daß wir ähnlich wie im ersten Beweisschritt vorgehen können. In der Tat könnten wir den hier betrachteten Fall auf das bereits Bewiesene zurückführen, indem wir zeigen, daß \tilde{g} auch Bedingung **C1** erfüllt.

Es erweist sich aber als vorteilhafter, einen etwas anderen Weg einzuschlagen und die Schätzfunktion G_n statt mit der deterministischen Matrix $V(\theta)^T$ (also dem Limeselement) mit der zufälligen Matrix $\frac{1}{n} [D_\theta G_n(\theta)]^T$ zu multiplizieren. Wir definieren also eine Hilfs-Schätzfunktion

$$H_n(\omega; \theta) := \frac{1}{n} [D_\theta G_n(\omega; \theta)]^T G_n(\omega; \theta).$$

Dann ist

$$\begin{aligned} \partial_j H_{n,i}(\omega; \theta) &= \partial_j \left[\frac{1}{n} \sum_{k=1}^d \partial_i G_{n,k}(\omega; \theta) G_{n,k}(\omega; \theta) \right] \\ &= \frac{1}{n} \sum_{k=1}^d \left(\partial_{ji}^2 G_{n,k}(\omega; \theta) \cdot G_{n,k}(\omega; \theta) + \partial_i G_{n,k}(\omega; \theta) \cdot \partial_j G_{n,k}(\omega; \theta) \right). \end{aligned}$$

Wegen der Voraussetzung **C2** haben wir nach Lemma 2.3.3 für alle $1 \leq i, j, k \leq d$

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} G_{n,k}(\vartheta) \right| \\ &= \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} G_{n,k}(\vartheta) - \underbrace{\tilde{\mu}_\theta(g_k(\theta))}_{=0} \right| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.}, \end{aligned} \tag{2.3.30}$$

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_i G_{n,k}(\vartheta) - \tilde{\mu}_\theta(\partial_i g_k(\theta)) \right| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \tag{2.3.31}$$

und

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_{ji}^2 G_{n,k}(\vartheta) - \tilde{\mu}_\theta(\partial_{ji}^2 g_k(\theta)) \right| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \tag{2.3.32}$$

Sei nun wieder $\|\cdot\|$ eine beliebige Norm auf \mathbb{R}^d und $\|\cdot\|$ die induzierte Matrixnorm. Wie im ersten Beweisschritt ergibt sich dann aus (2.3.30) und (2.3.31) zunächst

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} G_n(\vartheta) \right\| \xrightarrow{\delta \downarrow 0} 0 \quad \text{in } L^\infty(P_\theta), \quad (2.3.33)$$

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\vartheta) - V(\theta) \right\| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \quad (2.3.34)$$

und somit auch

$$\begin{aligned} \left\| \frac{1}{n} H_n(\theta) \right\| &= \left\| \frac{1}{n^2} [D_\theta G_n(\theta)]^T G_n(\theta) \right\| \\ &\leq \underbrace{\left\| \frac{1}{n} D_\theta G_n(\theta) \right\|}_{P_\theta\text{-f.s. beschränkt}} \cdot \underbrace{\left\| \frac{1}{n} G_n(\theta) \right\|}_{\rightarrow 0 P_\theta\text{-f.s.}} \xrightarrow{n \rightarrow \infty} 0 \quad P_\theta\text{-f.s.} \end{aligned} \quad (2.3.35)$$

Wir zeigen nun, daß auch

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta H_n(\vartheta) - V(\theta)^T V(\theta) \right\| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \quad (2.3.36)$$

Dazu beobachten wir, daß

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{\partial_{ji}^2 G_{n,k}(\vartheta)}{n} \cdot \frac{G_{n,k}(\vartheta)}{n} \right| \\ &\leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\{ \left| \frac{1}{n} \partial_{ji}^2 G_{n,k}(\vartheta) - \tilde{\mu}_\theta(\partial_{ji}^2 g_k(\theta)) \right| \cdot \left| \frac{1}{n} G_{n,k}(\cdot; \vartheta) \right| \right. \\ &\quad \left. + \left| \tilde{\mu}_\theta(\partial_{ji}^2 g_k(\theta)) \right| \cdot \left| \frac{1}{n} G_{n,k}(\vartheta) \right| \right\} \\ &\leq \underbrace{\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_{ji}^2 G_{n,k}(\vartheta) - \tilde{\mu}_\theta(\partial_{ji}^2 g_k(\theta)) \right|}_{\xrightarrow{\delta \downarrow 0} 0 \text{ nach (2.3.32)}} \cdot \underbrace{\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} G_{n,k}(\vartheta) \right|}_{\xrightarrow{\delta \downarrow 0} 0 \text{ nach (2.3.30)}} \quad (2.3.37) \\ &\quad + \underbrace{\left| \tilde{\mu}_\theta(\partial_{ji}^2 g_k(\theta)) \right| \cdot \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} G_{n,k}(\vartheta) \right|}_{\xrightarrow{\delta \downarrow 0} 0 \text{ nach (2.3.30)}} \\ &\xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.} \end{aligned}$$

und

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{\partial_i G_{n,k}(\vartheta)}{n} \cdot \frac{\partial_j G_{n,k}(\vartheta)}{n} - V_{ki}(\theta) V_{kj}(\theta) \right| \\
& \leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\{ \left| \left(\frac{1}{n} \partial_i G_{n,k}(\vartheta) - V_{ki}(\theta) \right) \left(\frac{1}{n} \partial_j G_{n,k}(\vartheta) - V_{kj}(\theta) \right) \right| \right. \\
& \quad \left. + \left| V_{kj}(\theta) \left(\frac{1}{n} \partial_i G_{n,k}(\vartheta) - V_{ki}(\theta) \right) \right| + \left| V_{ki}(\theta) \left(\frac{1}{n} \partial_j G_{n,k}(\vartheta) - V_{kj}(\theta) \right) \right| \right\} \\
& \leq \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_i G_{n,k}(\vartheta) - V_{ki}(\theta) \right| \cdot \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_j G_{n,k}(\vartheta) - V_{kj}(\theta) \right| \\
& \quad + |V_{kj}(\theta)| \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_i G_{n,k}(\vartheta) - V_{ki}(\theta) \right| \\
& \quad + |V_{ki}(\theta)| \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_j G_{n,k}(\vartheta) - V_{kj}(\theta) \right| \\
& \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.}
\end{aligned} \tag{2.3.38}$$

Aus (2.3.37) und (2.3.38) erhalten wir für $i, j = 1, \dots, d$

$$\begin{aligned}
& \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \partial_j H_{n,i}(\vartheta) - \underbrace{\sum_{k=1}^d V_{ki}(\theta) \cdot V_{kj}(\theta)}_{=[V(\theta)^T V(\theta)]_{ij}} \right| \\
& = \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \sum_{k=1}^d \left(\frac{\partial_{ji}^2 G_{n,k}(\theta)}{n} \cdot \frac{G_{n,k}(\theta)}{n} \right. \right. \\
& \quad \left. \left. + \frac{\partial_i G_{n,k}(\theta)}{n} \cdot \frac{\partial_j G_{n,k}(\theta)}{n} - V_{ki}(\theta) V_{kj}(\theta) \right) \right| \\
& \leq \sum_{k=1}^d \left(\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{\partial_{ji}^2 G_{n,k}(\vartheta)}{n} \cdot \frac{G_{n,k}(\vartheta)}{n} \right| \right. \\
& \quad \left. + \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{\partial_i G_{n,k}(\vartheta)}{n} \cdot \frac{\partial_j G_{n,k}(\vartheta)}{n} - V_{ki}(\theta) V_{kj}(\theta) \right| \right) \\
& \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.}
\end{aligned}$$

und damit auch (2.3.36). Setze nun

$$C := \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \|H_n(\theta)\| = 0 \right\} \cap \left\{ \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta H_n(\vartheta) - V(\theta)^T V(\theta) \right\| = 0 \right\} \\ \cap \left\{ \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\vartheta) - V(\theta) \right\| = 0 \right\},$$

so daß also $P_\theta(C) = 1$. Wenn wir nun wie in Schritt 1 zeigen können, daß

$$C \subseteq \liminf_{n \rightarrow \infty} \{ \exists \vartheta^* \in B_\varepsilon(\theta) \cap \Theta : G_n(\vartheta^*) = 0 \},$$

dann ist die Behauptung bewiesen. Nach Voraussetzung **I** ist $V(\theta)$ invertierbar und damit die Matrix $V(\theta)^T V(\theta)$ (strikt) positiv definit: sei $\lambda > 0$ ihr kleinster Eigenwert. Bezeichnet außerdem wieder \mathcal{G} die Gruppe der invertierbaren Matrizes in $\mathbb{R}^{d \times d}$, so folgt aus der Offenheit von \mathcal{G} die Existenz eines $\alpha > 0$, so daß für jede Matrix $A \in \mathbb{R}^{d \times d}$ mit $\|A - V(\theta)\| < \alpha$ stets $A \in \mathcal{G}$ ist.

Betrachte nun ein beliebiges $\omega \in C$. Zum vorgegebenen ε aus (2.3.23) wähle zunächst $\delta = \delta(\omega) > 0$, o.B.d.A. $\delta < \varepsilon$, mit

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta H_n(\omega; \vartheta) - V(\theta)^T V(\theta) \right\| < \frac{\lambda}{3}$$

und

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\omega; \vartheta) - V(\theta) \right\| < \alpha,$$

und dann $n_0 = n_0(\omega, \delta) \in \mathbb{N}$, so daß für alle $n \geq n_0$

$$\|H_n(\omega; \theta)\| < n\delta \cdot \lambda/3, \quad \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta H_n(\omega; \vartheta) - V(\theta)^T V(\theta) \right\| < \lambda/3$$

und

$$\sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} D_\theta G_n(\omega; \vartheta) - V(\theta) \right\| < \alpha,$$

Dann ist nach Wahl von α die Matrix $D_\theta G_n(\omega; \vartheta)$ für jedes $n \geq n_0$ und $\vartheta \in B_\delta(\theta)$ invertierbar. Eine völlig analoge Argumentation wie in Schritt 1 (Taylorentwicklung (2.3.24) und Abschätzung (2.3.29) für H_n statt G_n , dann Anwendung von Lemma 2.3.4) liefert für alle $n \geq n_0$ die Existenz eines $\vartheta^* \in B_\delta(\theta) \subseteq B_\varepsilon(\theta)$ mit $H_n(\omega; \vartheta^*) = 0$, was wegen

$$H_n(\omega; \vartheta^*) = \frac{1}{n} [D_\theta G_n(\omega; \vartheta^*)]^T G_n(\omega; \vartheta^*)$$

und der Invertierbarkeit von $D_\theta G_n(\omega; \vartheta^*)$ äquivalent ist zu $G_n(\omega; \vartheta^*) = 0$. Damit ist die Behauptung bewiesen.



Die entscheidende Idee im Beweis von Satz 2.3.23, nämlich die Verwendung von Lemma 2.3.4 in diesem Zusammenhang, stammt aus [Sør]. Satz 2.3.6 zeigt insbesondere, daß für P_θ -fast alle $\omega \in \Omega$ „für schließlich alle n “ mindestens eine Lösung $\vartheta \in \Theta$ der Schätzgleichung $G_n(\omega; \vartheta) = 0$ existiert, unter jedem P_θ : setzen wir

$$N_n(\omega) := \{\vartheta \in \Theta : G_n(\omega; \vartheta) = 0\}, \quad \omega \in \Omega, \quad (2.3.39)$$

so folgt aus (2.3.23) sofort

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \{\omega \in \Omega : N_n(\omega) \neq \emptyset\} \right) = 1. \quad (2.3.40)$$

Wir können nun eine Abbildung $\hat{\theta}_n : \Omega \rightarrow \mathbb{R}^d$ definieren durch

$$\hat{\theta}_n(\omega) := \begin{cases} \vartheta \in N_n(\omega) \text{ beliebig,} & \text{falls } N_n(\omega) \neq \emptyset \\ 0, & \text{falls } N_n(\omega) = \emptyset. \end{cases} \quad (2.3.41)$$

Ist die Abbildung $\hat{\theta}_n$ eindeutig bestimmt, d.h. hat die Gleichung $G_n(\omega; \vartheta) = 0$ für jedes $\omega \in \Omega$ höchstens eine Lösung $\vartheta \in \Theta$, so zeigt eine einfache Überlegung, daß $(\hat{\theta}_n)_n$ eine (eindeutige) stark konsistente Lösung der Schätzgleichung für den Parameter θ ist. Es genügt in der Tat die schwächere Voraussetzung, daß $P_\theta (\liminf_{n \rightarrow \infty} \{\omega \in \Omega : \#N_n(\omega) = 1\}) = 1$, wobei wir mit $\#$ die Kardinalität einer Menge bezeichnet haben. Existieren jedoch mehrere Lösungen der Schätzgleichung, so hat eine *beliebige* Wahl von $\hat{\theta}_n$ wie in (2.3.41) ohne weiteres keinen Grund, konsistent zu sein. Ohne zusätzliche Voraussetzungen erhält man daher lediglich das folgende, essentiell nichtkonstruktive Resultat:

Satz 2.3.7 *Sei $(G_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4). Die Funktion g genüge den Bedingungen **C1** und **D** oder den Bedingungen **C2** und **I**. Sei $\theta \in \Theta$ beliebig, aber fest. Dann gibt eine Folge $(\hat{\theta}_n)_n$ von \mathcal{F}_n -meßbaren Zufallsvariablen $\hat{\theta}_n : \Omega \rightarrow \mathbb{R}^d$, so daß*

$$P_\theta \left(\liminf_{n \rightarrow \infty} \{\omega \in \Omega : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0\} \right) = 1,$$

$$\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta \quad P_\theta\text{-f.s.}$$

Beweis Sei $\theta \in \Theta$ beliebig. Wähle eine Folge von positiven Zahlen $(\varepsilon_k)_k$ mit $\varepsilon_k \downarrow 0$ und $B_{\varepsilon_1}(\theta) \subseteq \Theta$. Setzen wir

$$N_{n,k}(\omega) := \{\vartheta \in B_{\varepsilon_k} : G_n(\omega; \vartheta) = 0\}$$

und

$$A_{n,k} := \{\omega \in \Omega : \exists \vartheta \in B_{\varepsilon_k}(\theta) \text{ mit } G_n(\omega; \vartheta) = 0\} = \{\omega \in \Omega : N_{n,k}(\omega) \neq \emptyset\},$$

so ergibt eine Anwendung von Theorem 2.3.6 mit $\varepsilon = \varepsilon_k$, daß

$$\forall k \in \mathbb{N} : P_\theta \left(\liminf_{n \rightarrow \infty} A_{n,k} \right) = 1,$$

also auch $P_\theta(C) = 1$, wo

$$C := \bigcap_{k \in \mathbb{N}} \liminf_{n \rightarrow \infty} A_{n,k}.$$

Ist nun $\omega \in C$ beliebig, so gibt es zu jedem $k \in \mathbb{N}$ ein $n_k \in \mathbb{N}$, so daß $\omega \in A_{n,k}$, also $N_{n,k}(\omega) \neq \emptyset$ für alle $n \geq n_k$. Für $1 \leq n < n_1$ setze $\hat{\theta}_n(\omega) := 0$, und für $n_k \leq n < n_{k+1}$ mit $k \in \mathbb{N}$ wähle ein Element $\hat{\theta}_n(\omega) \in N_{n,k}(\omega) \neq \emptyset$; aufgrund des Measurable Selection Theorems A.2 kann diese Auswahl so getroffen werden, daß $\hat{\theta}_n$ \mathcal{F}_n -meßbar ist, vgl. hierzu den Beweis von Satz 2.3.8 unten. Für $n_k \leq n < n_{k+1}$ ist dann nach Definition $\hat{\theta}_n(\omega) \in B_{\varepsilon_k}(\theta) \subseteq \Theta$ und $G_n(\omega; \hat{\theta}_n(\omega)) = 0$, insbesondere gilt also

$$\forall n \geq n_1 : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0.$$

Man überzeugt man sich leicht davon, daß $\hat{\theta}_n(\omega) \xrightarrow{n \rightarrow \infty} \theta$. Da $\omega \in C$ beliebig war und $P_\theta(C) = 1$, ist damit die Behauptung gezeigt.

Bemerkung Für den Spezialfall der Least-Squares-Gleichung, d.h. g wie in (2.2.8), wurde Satz 2.3.7 unter stärkeren Voraussetzungen in [KN], Theorem 3.1, bewiesen. Ein Beweis mit stochastischer statt fast sicherer Konvergenz findet sich für den Spezialfall der Likelihood-Gleichung (d.h. g wie in (2.2.9)) in [Bil1], Theorem 2.1, ebenfalls unter stärkeren Voraussetzungen, sowie für allgemeines g (unter minimal stärkeren Voraussetzungen) in [Sør], Theorem 3.6. \blacklozenge

Wesentlich ist, daß die Definition der Folge $(\hat{\theta}_n)_n$ in Satz 2.3.7 über $N_{n,k}$ von $\theta \in \Theta$ abhängt, also speziell auf die Wahl von θ zugeschnitten ist. Man hat also strenggenommen gar keine Schätzfolge für den unbekannt Parameter konstruiert. Außerdem hat eine zu $\theta \in \Theta$ gemäß Satz 2.3.7 gewählte Folge $\hat{\theta}_n$ hat a priori keinen Grund, auch unter $P_{\tilde{\theta}}$, $\tilde{\theta} \neq \theta$ fast sicher gegen $\tilde{\theta}$ zu konvergieren, da die Bedingungen **C1** und **C2** „lokal in θ “ sind. Insbesondere liefert Satz 2.3.7 nicht die Existenz einer stark konsistenten Schätzfolge. (Für eine Diskussion dieses Problems im Rahmen von i.i.d.-Variablen und der Likelihood-Gleichung siehe [Leh], S. 420 f.)

Das Problem läßt sich jedoch umgehen, wenn man die zusätzliche Annahme macht, daß sich der Parameter $\theta \in \Theta$ überhaupt (stark) konsistent schätzen läßt. Ist $(\hat{\tau}_n)_n$ eine (stark) konsistente Schätzfolge, so kann man eine (stark) konsistente Lösung der Schätzgleichung

$G_n(\vartheta) = 0$ durch „Projektion von $\hat{\tau}_n(\omega)$ in die Lösungsmenge $N_n(\omega)$ “ gewinnen, d.h. man definiert in symbolischer Schreibweise

$$\hat{\theta}_n(\omega) := \mathbf{1}_{N_n(\omega) \neq \emptyset} \operatorname{arginf}_{\vartheta \in N_n(\omega)} \|\hat{\tau}_n(\omega) - \vartheta\|.$$

Dieser Ausweg wurde z.B. vorgeschlagen von [Leh], S. 421 und [Pfa], S. 202. Die Folge $(\hat{\theta}_n)_n$ ist dann sogar asymptotisch normal mit Konvergenzrate \sqrt{n} , ohne daß man eine entsprechende Annahme an die Folge der Hilfsschätzer $(\hat{\tau}_n)_n$ machen muß. Wir erhalten als Hauptresultat dieses Abschnitts:

Theorem 2.3.8 *Sei $(G_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4), d.h.*

$$G_n(\theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta).$$

*Die Funktion g genüge den Bedingungen **C1** und **D** oder den Bedingungen **C2** und **I**, und für alle $\theta \in \Theta$ und $i = 1, \dots, d$ seien die Funktionen $E^2 \ni (x, y) \mapsto g(x, y; \theta)$ in $L^2(\tilde{\mu}_\theta)$. Ferner sei der Parameter $\theta \in \Theta$ (stark) konsistent schätzbar. Dann existiert eine (stark) konsistente Lösung $(\hat{\theta}_n)_n$ der Schätzgleichung (2.3.1), und es gilt*

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V(\theta)^{-1}W(\theta)V(\theta)^{-T}), \quad (2.3.42)$$

wo $V(\theta)$ und $W(\theta)$ durch (2.3.14) und (2.2.10) gegeben sind.

Beweis

1. Sei zunächst $(\hat{\tau}_n)_n$ eine beliebige Schätzfolge für den Parameter θ , also $\hat{\tau}_n : \Omega \rightarrow \mathbb{R}^d$ \mathcal{F}_n -meßbar für alle $n \in \mathbb{N}$. Mit N_n aus (2.3.39) definiere die zufällige Menge

$$D_n(\omega) := \{\vartheta \in N_n(\omega) : \|\hat{\tau}_n(\omega) - \vartheta\| = \operatorname{dist}(\hat{\tau}_n(\omega), N_n(\omega))\} \subseteq N_n(\omega).$$

Es ist $D_n(\omega)$ stets kompakt, und wegen der Stetigkeit von $G_n(\omega; \cdot) : \Theta \rightarrow \mathbb{R}^d$ ist $N_n(\omega) = G_n(\cdot; \omega)^{-1}(\{0\})$ stets abgeschlossen. Ist $N_n(\omega) \neq \emptyset$, so folgt also auch $D_n(\omega) \neq \emptyset$ [denn für jede nichtleere abgeschlossene Teilmenge $C \subseteq \mathbb{R}^d$ und festes $y \in \mathbb{R}^d$ gibt es ein $x \in C$ mit $\|y - x\| = \operatorname{dist}(y, C)$]. Wir definieren nun eine mengenwertige Funktion $F_n : \Omega \rightarrow \mathcal{P}(\mathbb{R}^d)$ durch

$$F_n(\omega) := \begin{cases} D_n(\omega), & \text{falls } N_n(\omega) \neq \emptyset \\ \{0\} & \text{falls } N_n(\omega) = \emptyset. \end{cases}$$

Dann ist $F_n(\omega)$ für jedes $\omega \in \Omega$ kompakt und nichtleer. Wir zeigen nun: für jede offene Menge $U \subseteq \mathbb{R}^d$ gilt

$$\{\omega \in \Omega : F_n(\omega) \cap U \neq \emptyset\} \in \mathcal{F}_n. \quad (2.3.43)$$

Dazu betrachte zunächst die Funktion

$$\Omega \ni \omega \mapsto h_n(\omega) := \text{dist}(\hat{\tau}(\omega), N_n(\omega))$$

(mit der Konvention $\text{dist}(y, \emptyset) = \infty$, so daß also h_n Werte in $[0, \infty]$ annimmt) und zeige: h_n ist \mathcal{F}_n -meßbar. Dazu reicht zu zeigen: für jedes $0 \leq b < \infty$ ist $\{h_n \leq b\} \in \mathcal{F}_n$. Wir haben unter Ausnutzung der Abgeschlossenheit von $N_n(\omega)$

$$\begin{aligned} \{\omega : h_n(\omega) \leq b\} &= \{\omega : \text{dist}(\hat{\tau}_n(\omega), N_n(\omega)) \leq b\} \\ &= \{\omega : N_n(\omega) \neq \emptyset, \min_{\vartheta \in N_n(\omega)} \|\hat{\tau}_n(\omega) - \vartheta\| \leq b\} \\ &= \{\omega : \exists \vartheta \in N_n(\omega) \text{ mit } \|\hat{\tau}_n(\omega) - \vartheta\| \leq b\} \in \mathcal{F}_n \end{aligned}$$

nach Satz A.1, da die Funktion $(\omega; \vartheta) \mapsto \hat{\tau}_n(\omega) - \vartheta$ \mathcal{F}_n -meßbar in ω und stetig in ϑ ist. Völlig analog hat man für beliebiges kompaktes $K \subseteq U$ die Meßbarkeit von $\omega \mapsto \text{dist}(\hat{\tau}(\omega), N_n(\omega) \cap K)$. Außerdem ist $\{\omega : N_n(\omega) \cap K \neq \emptyset\} = \{\omega : \exists \vartheta \in \Theta \cap K \text{ mit } G_n(\omega; \vartheta) = 0\} \in \mathcal{F}_n$ wiederum nach Satz A.1, da G_n \mathcal{F}_n -meßbar in ω und stetig in ϑ ist; genauso ist $\{\omega : N_n(\omega) \neq \emptyset\} \in \mathcal{F}_n$. Für jede kompakte Menge $K \subseteq U$ gilt somit

$$\begin{aligned} &\{\omega : D_n(\omega) \cap K \neq \emptyset\} \\ &= \{\omega : \exists \vartheta \in N_n(\omega) \cap K \text{ mit } \|\hat{\tau}_n(\omega) - \vartheta\| = \text{dist}(\hat{\tau}_n(\omega), N_n(\omega))\} \\ &= \underbrace{\{\omega : N_n(\omega) \cap K \neq \emptyset\}}_{\in \mathcal{F}_n} \cap \underbrace{\{\omega : \text{dist}(\hat{\tau}_n(\omega), N_n(\omega) \cap K) = \text{dist}(\hat{\tau}_n(\omega), N_n(\omega))\}}_{\in \mathcal{F}_n} \in \mathcal{F}_n. \end{aligned}$$

Mit einer kompakten Ausschöpfung $(K_m)_{m \in \mathbb{N}}$ von U erhalten wir somit

$$\{\omega : D_n(\omega) \cap U \neq \emptyset\} = \bigcup_{m \in \mathbb{N}} \{\omega : D_n(\omega) \cap K_m \neq \emptyset\} \in \mathcal{F}_n,$$

und es folgt

$$\begin{aligned} &\{\omega \in \Omega : F_n(\omega) \cap U \neq \emptyset\} \in \mathcal{F}_n \\ &= \underbrace{\{\omega : D_n(\omega) \cap U \neq \emptyset\}}_{\in \mathcal{F}_n} \cup \underbrace{\{\omega : N_n(\omega) = \emptyset, 0 \in U\}}_{= \emptyset \text{ oder } = \{\omega : N_n(\omega) \neq \emptyset\} \in \mathcal{F}_n} \in \mathcal{F}_n, \end{aligned}$$

das ist (2.3.43).

Nach dem Measurable Selection Theorem A.2 gibt es nun eine \mathcal{F}_n -meßbare Zufallsvariable $\hat{\theta}_n : \Omega \rightarrow \mathbb{R}^d$ mit $\hat{\theta}_n(\omega) \in F(\omega)$ für alle $\omega \in \Omega$. Dies bedeutet nach Definition von F_n , daß für jedes $\omega \in \Omega$ mit $N_n(\omega) \neq \emptyset$ gilt:

$$\hat{\theta}_n(\omega) \in N_n(\omega) \subseteq \Theta, \quad \|\hat{\tau}_n(\omega) - \hat{\theta}_n(\omega)\| = \text{dist}(\hat{\tau}_n(\omega), N_n(\omega)). \quad (2.3.44)$$

Da die Voraussetzungen von Satz 2.3.6 erfüllt sind, gilt (2.3.40), und es folgt

$$\begin{aligned} & \lim_{n \rightarrow \infty} P_\theta \left(\{\omega : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0\} \right) \\ & \geq P_\theta \left(\liminf_{n \rightarrow \infty} \{\omega : \hat{\theta}_n(\omega) \in \Theta, G_n(\omega; \hat{\theta}_n(\omega)) = 0\} \right) \\ & \geq P_\theta \left(\liminf_{n \rightarrow \infty} \{\omega : N_n(\omega) \neq \emptyset\} \right) = 1. \end{aligned} \quad (2.3.45)$$

2. Sei nun $(\hat{\tau}_n)_n$ als stark konsistent vorausgesetzt. Fixiere ein beliebiges $\theta \in \Theta$. Ähnlich wie im Beweis von Satz 2.3.7 setzen wir $N_{n,\varepsilon}(\omega) := B_\varepsilon(\theta) \cap N_n(\omega)$ für $\varepsilon > 0$ und definieren die Menge

$$C := \bigcap_{\varepsilon \in \mathbb{Q}^+} \liminf_{n \rightarrow \infty} \{\omega : N_{n,\varepsilon}(\omega) \neq \emptyset\} \cap \left\{ \lim_{n \rightarrow \infty} \hat{\tau}_n = \theta \right\} \in \mathcal{A}.$$

Aus Satz 2.3.6 sowie der vorausgesetzten starken Konsistenz von $(\hat{\tau}_n)_n$ folgt dann $P_\theta(C) = 1$. Sei nun $\omega \in C$ und $\varepsilon > 0$ beliebig. Dann gibt es $n_0 \in \mathbb{N}$, so daß für alle $n \geq n_0$ gilt:

$$N_n(\omega) \supseteq N_{n,\varepsilon/2} \neq \emptyset, \quad \|\hat{\tau}_n(\omega) - \theta\| < \varepsilon/2.$$

Nach (2.3.44) gilt folglich für $n \geq n_0$

$$\|\hat{\tau}_n(\omega) - \hat{\theta}_n(\omega)\| = \text{dist}(\hat{\tau}_n(\omega), N_n(\omega)) \leq \underbrace{\|\hat{\tau}_n(\omega) - \theta\|}_{< \varepsilon/2} + \underbrace{\text{dist}(\theta, N_n(\omega))}_{< \varepsilon/2, \text{ da } N_n(\omega) \cap B_{\varepsilon/2}(\theta) \neq \emptyset} < \varepsilon. \quad (2.3.46)$$

Gezeigt ist also $\|\hat{\tau}_n - \hat{\theta}_n\| \xrightarrow{n \rightarrow \infty} 0$ P_θ -fast sicher und damit auch $\|\hat{\theta}_n - \theta\| \xrightarrow{n \rightarrow \infty} 0$ P_θ -fast sicher. Zusammen mit (2.3.45) zeigt dies, daß durch $(\hat{\theta}_n)_n$ eine stark konsistente Lösung der Schätzgleichung gegeben ist.

3. Nun sei nur noch Konsistenz von $(\hat{\tau}_n)_n$ vorausgesetzt. Fixiere ein beliebiges $\theta \in \Theta$ und $\delta > 0$. Wir müssen zeigen, daß

$$P_\theta \left(\|\hat{\theta}_n - \theta\| \leq \delta \right) \xrightarrow{n \rightarrow \infty} 1. \quad (2.3.47)$$

Setze nun

$$A_n := \{\omega : N_{n,\delta/2}(\omega) \neq \emptyset\}, \quad B_n := \{\|\hat{\tau}_n - \theta\| \leq \delta/2\}.$$

Nach Satz 2.3.6 ist

$$\lim_{n \rightarrow \infty} P_\theta(A_n) \geq P_\theta \left(\liminf_{n \rightarrow \infty} \{\omega : N_{n,\delta/2} \neq \emptyset\} \right) = 1$$

und wegen der vorausgesetzten Konsistenz von $(\hat{\tau}_n)_n$ damit auch

$$P_\theta(A_n \cap B_n) \xrightarrow{n \rightarrow \infty} 1.$$

Genau wie in (2.3.46) ergibt sich für $\omega \in A_n \cap B_n$ aufgrund der Definition von $\hat{\theta}_n$, daß $\|\hat{\tau}_n(\omega) - \hat{\theta}_n(\omega)\| \leq \delta$, und es folgt

$$P_\theta \left(\|\hat{\tau}_n - \hat{\theta}_n\| \leq \delta \right) \geq P_\theta(A_n \cap B_n) \xrightarrow{n \rightarrow \infty} 1.$$

Hieraus folgt sofort (2.3.47) und damit $\hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta$ P_θ -stochastisch. Zusammen mit (2.3.45) heißt das: $(\hat{\theta}_n)_n$ ist eine konsistente Lösung der Schätzgleichung.

4. Die asymptotische Normalität (2.3.42) der Schätzfolge $(\hat{\theta}_n)_n$ ergibt sich nun sofort aus Satz 2.2.1. Damit ist der Beweis abgeschlossen. ■

2.4 Likelihood-Schätzung und LAN

Im letzten Abschnitt dieses Kapitels sollen die bisher gewonnen Resultate verwendet werden, um unter einfachen Bedingungen die Existenz einer (stark) konsistenten eindeutigen Lösung der Likelihood-Gleichung sowie lokal-asymptotische Normalität (im folgenden kurz: **LAN**) für die Folge von Experimenten

$$\mathcal{E}_n := (\Omega, \mathcal{F}_n, \{P_{n,\theta} : \theta \in \Theta\})$$

zu zeigen, wobei

$$P_{n,\theta} := P_\theta|_{\mathcal{F}_n}.$$

Im ganzen Abschnitt sollen folgende Voraussetzungen gelten: es gebe ein σ -endliches Maß λ auf $(E, \mathcal{B}(E))$, so daß $Q_\theta(x, dy) \sim \lambda(dy)$ für alle $x \in E$ und $\theta \in \Theta$. Dies ist äquivalent dazu, daß die Übergangswahrscheinlichkeit Q_θ durch eine Übergangsdichte gegeben ist, die strikt positiv festgelegt werden kann, d.h. es gibt eine Funktion $p : E^2 \times \Theta \rightarrow \mathbb{R}_{>0}$ mit

$$\forall \theta \in \Theta, \forall x \in E : Q_\theta(x, dy) = p(x, y; \theta) \lambda(dy). \quad (2.4.1)$$

Ist die Startverteilung ν nicht bekannt, sondern ebenfalls Bestandteil der Parametrisierung, $\nu \in \{\nu_\theta : \theta \in \Theta\}$, so soll auch $\nu_\theta \sim \lambda$ für alle $\theta \in \Theta$ gelten, was äquivalent ist zur Existenz einer Funktion $f : E \times \Theta \rightarrow \mathbb{R}_{>0}$ mit

$$\forall \theta \in \Theta : \nu_\theta(dy) = f(y; \theta) \lambda(dy). \quad (2.4.2)$$

Im folgenden gehen wir von dieser Situation aus; der Fall, daß ν bekannt ist, läßt sich analog (einfacher) behandeln. Für eine Menge $A = \{X_0 \in B_0, \dots, X_n \in B_n\} \in \mathcal{F}_n$ mit $B_0, \dots, B_n \in \mathcal{B}(E)$ setzen wir mit $\lambda^{n+1} := \otimes_{i=0}^n \lambda$

$$\Lambda_n(A) := \lambda^{n+1}(B_0 \times \dots \times B_n);$$

dadurch wird eindeutig ein σ -endliches Maß Λ_n auf \mathcal{F}_n definiert, für welches offenbar

$$\mathcal{L}(X_0, \dots, X_n | \Lambda_n) = \lambda^{n+1}$$

gilt und bezüglich dessen wir $\{P_{n,\theta} : \theta \in \Theta\}$ als dominierte Familie auffassen können: man rechnet sofort nach, daß aufgrund der Markov-Eigenschaft für eine Menge $A = \{X_0 \in B_0, \dots, X_n \in B_n\} \in \mathcal{F}_n$ mit $B_0, \dots, B_n \in \mathcal{B}(E)$

$$P_{n,\theta}(A) = \int_A f(X_0; \theta) \prod_{i=1}^n p(X_{i-1}, X_i; \theta) d\Lambda_n \quad (2.4.3)$$

gilt. Da die Mengen dieser Gestalt einen durchschnittsstabilen Erzeuger von \mathcal{F}_n bilden, gilt (2.4.3) sogar für alle $A \in \mathcal{F}_n$, d.h. $f(X_0; \theta) \prod_{i=1}^n p(X_{i-1}, X_i; \theta)$ ist eine strikt positive Festlegung der Dichte von $P_{n,\theta}$ bezüglich Λ_n :

$$\forall n \in \mathbb{N}, \forall \theta \in \Theta : \frac{dP_{n,\theta}}{d\Lambda_n} = f(X_0; \theta) \prod_{i=1}^n p(X_{i-1}, X_i; \theta) > 0. \quad (2.4.4)$$

Damit sind für jedes $n \in \mathbb{N}$ die Wahrscheinlichkeitsmaße $\{P_{n,\theta} : \theta \in \Theta\}$ äquivalent zu Λ_n und daher auch untereinander, und für die Dichtequotienten (Likelihood-Ratios) ergibt sich

$$\forall \vartheta, \theta \in \Theta : \frac{dP_{n,\vartheta}}{dP_{n,\theta}} = \frac{f(X_0; \vartheta)}{f(X_0; \theta)} \prod_{i=1}^n \frac{p(X_{i-1}, X_i; \vartheta)}{p(X_{i-1}, X_i; \theta)}.$$

Die log-Likelihood-Funktion ist gegeben durch

$$\log \frac{dP_{n,\theta}}{d\Lambda_n} = \log f(X_0; \theta) \prod_{i=1}^n p(X_{i-1}, X_i; \theta) = \log f(X_0; \theta) + \sum_{i=1}^n \log p(X_{i-1}, X_i; \theta). \quad (2.4.5)$$

Da der Term $\log f(X_0; \theta)$, wie wir sehen werden, asymptotisch vernachlässigbar ist, arbeiten wir im folgenden anstelle von (2.4.5) mit der modifizierten log-Likelihood-Funktion

$$L_n(\theta) := \sum_{i=1}^n \log p(X_{i-1}, X_i; \theta).$$

zu arbeiten. (Ist die Startverteilung bekannt, so fällt L_n mit der exakten log-Likelihood-Funktion zusammen.) Ist L_n differenzierbar in θ , so ist der *Score-Vektor* der partiellen Ableitungen von L_n nach θ_i gegeben durch

$$\nabla_{\theta} L_n(\theta) = \sum_{i=1}^n \nabla_{\theta} \log p(X_{i-1}, X_i; \theta),$$

und die Gleichung

$$\nabla_{\theta} L_n(\cdot; \theta) = 0 \quad (2.4.6)$$

heißt *Likelihood-Gleichung*. Diese Gleichung ist offenbar von der Bauart (2.2.4). Unter geeigneten Bedingungen an die Übergangsdichte p können wir die in Abschnitt 2.3 entwickelten Resultate verwenden, um die Existenz von (stark) konsistenten und asymptotisch normalen Lösungen der Likelihood-Gleichung (2.4.6) zu zeigen. Praktisch dieselben Bedingungen stellen sicher, daß die Folge der Experimente $(\mathcal{E}_n)_n = (\Omega, \mathcal{F}_n, \{P_{n,\theta} : \theta \in \Theta\})$ an jeder Stelle $\theta \in \Theta$ lokal-asymptotisch normal ist. Wir arbeiten mit der folgenden Definition von **LAN** (siehe z.B. [Pfa], Condition 8.1.1):

Definition 2.4.1 (LAN-Bedingung) Die Folge von Experimenten $(\mathcal{E}_n)_n$ heißt lokal-asymptotisch normal an der Stelle $\theta \in \Theta$, falls es eine (strikt) positiv definite Matrix $K(\theta) \in \mathbb{R}^{d \times d}$ sowie Zufallsvariablen $S_n = S_n(\theta) : \Omega \rightarrow \mathbb{R}^d$ mit

$$S_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, K(\theta)) \quad \text{unter } P_\theta \quad (2.4.7)$$

und eine Folge $\delta_n = \delta_n(\theta) \rightarrow 0$ für $n \rightarrow \infty$ gibt, so daß für jede beschränkte Folge $(u_n)_n \subseteq \mathbb{R}^d$ gilt:

$$\log \frac{dP_{n,\theta+\delta_n(\theta)u_n}}{dP_{n,\theta}} = u_n^T S_n - \frac{1}{2} u_n^T K(\theta) u_n + o_{P_\theta}(1), \quad n \rightarrow \infty. \quad (2.4.8)$$

Bemerkungen Die Gültigkeit der **LAN**-Bedingung (2.4.1) impliziert eine Reihe von Folgerungen, für die wir z.B. auf [IH], Kapitel 9 ff. verweisen. Insbesondere stellt sich $K(\theta)^{-1}$ als untere Schranke für die asymptotische Varianz *regulärer* Schätzfolgen heraus; zu diesem Begriff siehe [IH], Definition 9.1. Unter moderaten Zusatzvoraussetzungen kann außerdem jede konsistente Schätzfolge, die die richtige Konvergenzgeschwindigkeit besitzt, zu einer regulären und effizienten Schätzfolge umgebaut werden; für eine einfache Version dieses auf Le Cam zurückgehenden Verfahrens siehe z.B. [Hö], Kapitel 7 B. \blacklozenge

Wir kommen nun zum Hauptresultat dieses Abschnitts. Zusätzlich zu den Bedingungen, die im folgenden Theorem formuliert werden, sollen natürlich (2.4.1) und (2.4.2) sowie die generellen Voraussetzungen aus Abschnitt 2.1 erfüllt sein, d.h. insbesondere: $(X_n)_n$ ist unter jedem P_θ ein ergodischer Markov-Prozeß mit einem eindeutigen invarianten Wahrscheinlichkeitsmaß μ_θ .

Theorem 2.4.2 Für die Übergangsdichte p und die Dichte f der Startverteilung seien folgende Voraussetzungen erfüllt:

- Für jedes $x \in E$ ist die Funktion $\Theta \ni \theta \mapsto f(x; \theta)$ stetig.
- Für alle $(x, y) \in E^2$ ist die Funktion $\Theta \ni \theta \mapsto p(x, y; \theta)$ in $C^2(\Theta)$, und somit ist auch für alle $(x, y) \in E^2$ die Funktion

$$\Theta \ni \theta \mapsto \ell(x, y; \theta) := \log p(x, y; \theta)$$

wohldefiniert und in $C^2(\Theta)$.

- Für alle $\theta \in \Theta$ und $i = 1, \dots, d$ ist die Funktion $E^2 \ni (x, y) \mapsto \partial_i \ell(x, y; \theta)$ in $L^2(\tilde{\mu}_\theta)$.
- Für alle $i, j = 1, \dots, d$ ist die Funktion $\partial_{ij}^2 \ell$ lokal dominiert integrierbar.
- Es gilt

$$\forall \theta \in \Theta, x \in E, i, j = 1, \dots, d : \int_E \partial_i p(x, y; \theta) \lambda(dy) = 0 = \int_E \partial_{ij}^2 p(x, y; \theta) \lambda(dy). \quad (2.4.9)$$

- Für alle $\theta \in \Theta$ ist die Fisher-Information

$$K(\theta) := [\tilde{\mu}_\theta(\partial_i \ell(\cdot, \cdot; \theta) \cdot \partial_j \ell(\cdot, \cdot; \theta))]_{i,j=1,\dots,d} \in \mathbb{R}^{d \times d}$$

invertierbar.

Dann gilt **LAN** an jeder Stelle $\theta \in \Theta$ mit $K(\theta)$ in (2.4.7) und normierender Folge $\delta_n(\theta) = n^{-1/2}$.

Ist ferner der Parameter $\theta \in \Theta$ stark konsistent (bzw. konsistent) schätzbar, so gibt es eine eindeutige stark konsistente (bzw. eindeutige konsistente) Lösung $(\hat{\theta}_n)_n$ der Likelihood-Gleichung (2.4.6) mit

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \{ \omega : \hat{\theta}_n(\omega) \text{ ist lokale Maximalstelle von } L_n(\omega; \cdot) \} \right) = 1, \quad (2.4.10)$$

bzw.

$$\forall \theta \in \Theta : P_\theta \left(\{ \omega : \hat{\theta}_n(\omega) \text{ ist lokale Maximalstelle von } L_n(\omega; \cdot) \} \right) \xrightarrow{n \rightarrow \infty} 1. \quad (2.4.11)$$

In beiden Fällen gilt dabei

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, K(\theta)^{-1}) \quad \text{unter } P_\theta. \quad (2.4.12)$$

Bemerkung Bedingung (2.4.9) folgt sofort aus der Gleichung $\int_E p(x, y; \theta) = 1$, falls für jedes $x \in E$ die Funktion

$$\Theta \ni \theta \mapsto \int_E p(x, y; \theta) \lambda(dy) \quad (2.4.13)$$

zweimal partiell unter dem Integralzeichen differenziert werden darf. Die Fisher-Information $K(\theta)$ ist gerade die zu $g = \nabla_\theta \ell$ gehörige Matrix $W(\theta)$ aus (2.2.10). Da $K(\theta)$ als Kovarianzmatrix stets positiv semidefinit ist, folgt aus der der Nichtsingularitätsannahme, daß $K(\theta)$ sogar (strikt) positiv definit ist. \blacklozenge

Beweis Fixiere ein beliebiges $\theta \in \Theta$. Im folgenden bezeichne $\nabla \nabla_\theta^T L_n(\omega; \theta)$ die Hesse-Matrix der zweiten partiellen Ableitungen von $L_n(\omega; \cdot)$ an der Stelle θ .

1. Sei $(u_n)_n \subseteq \mathbb{R}^d$ beschränkt. Eine Taylorentwicklung von $L_n(\omega; \cdot)$ um θ ergibt für die log-Likelihood-Ratios (für genügend große n , so daß $\theta + n^{-1/2}u_n \in \Theta$)

$$\begin{aligned}
& \log \frac{dP_{n, \theta + n^{-1/2}u_n}}{dP_{n, \theta}}(\omega) \\
&= f(X_0(\omega); \theta + n^{-1/2}u_n) - f(X_0(\omega); \theta) + L_n(\omega; \theta + n^{-1/2}u_n) - L_n(\omega; \theta) \\
&= f(X_0(\omega); \theta + n^{-1/2}u_n) - f(X_0(\omega); \theta) \\
&\quad + n^{-1/2}u_n^T \nabla_{\theta} L_n(\omega; \theta) + \frac{1}{2}n^{-1}u_n^T [\nabla \nabla_{\theta}^T L_n(\omega; \theta + \lambda_n(\omega)n^{-1/2}u_n)] u_n \\
&= +u_n^T S_n(\omega) - \frac{1}{2}u_n^T K(\theta)u_n \\
&\quad + f(X_0(\omega); \theta + n^{-1/2}u_n) - f(X_0(\omega); \theta) + \frac{1}{2}u_n^T (K(\theta) + K_n(\omega))u_n,
\end{aligned} \tag{2.4.14}$$

mit

$$S_n(\omega) := n^{-1/2} \nabla_{\theta} L_n(\omega; \theta) = n^{-1/2} \sum_{k=1}^n \nabla_{\theta} \ell(X_{k-1}(\omega), X_k(\omega); \theta)$$

und

$$\begin{aligned}
K_n(\omega) &:= n^{-1} \nabla \nabla_{\theta}^T L_n(\omega; \theta + \lambda_n(\omega)n^{-1/2}u_n) \\
&= n^{-1} \sum_{k=1}^n \nabla \nabla_{\theta}^T \ell(X_{k-1}(\omega), X_k(\omega); \theta + \lambda_n(\omega)n^{-1/2}u_n).
\end{aligned}$$

Dabei ist $\lambda_n : \Omega \rightarrow [0, 1]$ wie in (2.3.15) eine Abbildung, die meßbar, also als Zufallsvariable festgelegt werden kann (analoge Argumentation wie im Beweis von Satz 2.3.5).

2. Es ist $(L_n)_n$ eine Folge von Martingal-Schätzfunktionen: nach Voraussetzung ist für jedes $i = 1, \dots, d$ und $\theta \in \Theta$ die Funktion $\partial_i \ell(\cdot, \cdot; \theta) : E^2 \rightarrow \mathbb{R}$ in $L^2(\tilde{\mu}_{\theta}) \subseteq L^1(\tilde{\mu}_{\theta})$, und nach (2.4.9) gilt

$$\forall \theta \in \Theta, x \in E : \int_E \partial_i \ell(x, y; \theta) Q_{\theta}(x, dy) = \int_E \frac{\partial_i p(x, y; \theta)}{p(x, y; \theta)} p(x, y; \theta) \lambda(dy) = 0,$$

das ist (2.2.6). Aus dem Zentralen Grenzwertsatz (2.2.1) folgt nun sofort

$$S_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, K(\theta)) \quad \text{unter } P_{\theta},$$

das ist (2.4.7).

Betrachte nun die zu $\nabla_{\theta} \ell$ gehörige Matrix $V(\theta)$ aus (2.3.14), also $V_{ij}(\theta) = \tilde{\mu}_{\theta}(\partial_j \partial_i \ell(\cdot, \cdot; \theta))$, und zeige: es ist $V(\theta) = -K(\theta)$. Aus (2.4.9) ergibt sich nämlich

$$\tilde{\mu}_{\theta} \left(\frac{\partial_{ji}^2 p(\cdot, \cdot; \theta)}{p(\cdot, \cdot; \theta)} \right) = \int_E \tilde{\mu}_{\theta}(dx) \underbrace{\int_E \frac{\partial_{ji}^2 p(x, y; \theta)}{p(x, y; \theta)} p(x, y; \theta) \lambda(dy)}_{=0} = 0$$

und somit

$$\begin{aligned}
V_{ij}(\theta) &= \tilde{\mu}_\theta (\partial_{ji}^2 \ell(\cdot, \cdot; \theta)) \\
&= \tilde{\mu}_\theta \left(\frac{\partial_{ji}^2 p(\cdot, \cdot; \theta) \cdot p(\cdot, \cdot; \theta) - \partial_i p(\cdot, \cdot; \theta) \cdot \partial_j (\cdot, \cdot; \theta)}{p(\cdot, \cdot; \theta)^2} \right) \\
&= \underbrace{\tilde{\mu}_\theta \left(\frac{\partial_{ji}^2 p(\cdot, \cdot; \theta)}{p(\cdot, \cdot; \theta)} \right)}_{=0} - \tilde{\mu}_\theta \left(\frac{\partial_i p(\cdot, \cdot; \theta)}{p(\cdot, \cdot; \theta)} \cdot \frac{\partial_j p(\cdot, \cdot; \theta)}{p(\cdot, \cdot; \theta)} \right) \\
&= -\tilde{\mu}_\theta (\partial_i \ell(\cdot, \cdot; \theta) \cdot \partial_j \ell(\cdot, \cdot; \theta)) \\
&= -K_{ij}(\theta).
\end{aligned}$$

Da nun alle Funktionen $\partial_{ij}^2 \ell$ lokal dominiert integrierbar sind und $\theta + \lambda_n \cdot n^{-1/2} u_n \xrightarrow{n \rightarrow \infty} \theta$, folgt aus Lemma 2.3.8, daß

$$K_{ij}^{(n)} = \frac{1}{n} \sum_{k=1}^n \partial_{ij}^2 \ell(X_{k-1}, X_k; \theta + \lambda_n \cdot n^{-1/2} u_n) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta (\partial_{ij}^2 \ell(\cdot, \cdot; \theta)) = V_{ij}(\theta) \quad P_\theta\text{-f.s.},$$

also $K^{(n)} \xrightarrow{n \rightarrow \infty} V(\theta) = -K(\theta)$ P_θ -f.s. Zusammen mit der Stetigkeit von f in ϑ folgt insbesondere

$$f(X_0; \theta + n^{-1/2} u_n) - f(X_0; \theta) + \frac{1}{2} u_n^T (K(\theta) + K^{(n)}) u_n = o_{P_\theta}.$$

Wegen der Entwicklung (2.4.14) ist damit die **LAN**-Bedingung (2.4.1) nachgewiesen.

3. Es ist schon gezeigt, daß $(L_n)_n$ eine Folge von Martingal-Schätzfunktionen der Bauart (2.2.4) mit $g = \nabla_\theta \ell$ ist. Da die Matrix $V(\theta) = -K(\theta)$ negativ definit ist, folgt zusammen mit dem schon Bewiesenen, daß $\nabla_\theta \ell$ die Bedingungen **C1** und **D** aus Abschnitt 2.3 erfüllt. Daher können wir Theorem 2.3.8 anwenden, um zu schließen, daß es eine konsistente (stark konsistente) Lösung $(\hat{\theta}_n)_n$ der Likelihood-Gleichung (2.4.6) gibt, falls der Parameter $\theta \in \Theta$ überhaupt konsistent (stark konsistent) schätzbar ist.

Wegen der negativen Definitheit von $V(\theta)$ gibt es $\alpha > 0$ so, daß für $A \in \mathbb{R}^{d \times d}$ mit $\|A - V(\theta)\| < \alpha$ stets auch A negativ definit ist. Da die Funktionen $\partial_{ij}^2 \ell$ nach Voraussetzung lokal dominiert integrierbar sind, folgt nach Lemma (2.3.3)

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left| \frac{1}{n} \sum_{k=1}^n \partial_{ji}^2 \ell(X_{k-1}, X_k; \vartheta) - V_{ij}(\theta) \right| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.},$$

also auch

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} \nabla \nabla_\theta^T L_n(\cdot; \vartheta) - V(\theta) \right\| \xrightarrow{\delta \downarrow 0} 0 \quad P_\theta\text{-f.s.}$$

Sei nun $(\hat{\vartheta}_n)_n$ eine beliebige stark konsistente Lösung der Likelihood-Gleichung. Definiere

$$C := \left\{ \hat{\theta}_n \xrightarrow{n \rightarrow \infty} \theta \right\} \cap \left\{ \hat{\vartheta}_n \xrightarrow{n \rightarrow \infty} \theta \right\} \cap \left\{ \lim_{\delta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} \nabla \nabla_\theta^T L_n(\vartheta) - V(\theta) \right\| = 0 \right\} \\ \cap \liminf_{n \rightarrow \infty} \{ \hat{\theta}_n \in \Theta, \nabla_\theta L_n(\hat{\theta}_n) = 0 \} \cap \liminf_{n \rightarrow \infty} \{ \hat{\vartheta}_n \in \Theta, \nabla_\theta L_n(\hat{\vartheta}_n) = 0 \},$$

so daß also $P_\theta(C) = 1$. Für $\omega \in C$ gibt es dann $\delta = \delta(\omega) > 0$ und $n_0 = n_0(\omega, \delta) \in \mathbb{N}$, so daß für $n \geq n_0$ stets folgendes gilt:

$$\hat{\theta}_n(\omega), \hat{\vartheta}_n(\omega) \in B_\delta(\theta) \cap \Theta, \quad \nabla_\theta L_n(\omega; \hat{\theta}_n(\omega)) = 0 = \nabla_\theta L_n(\omega; \hat{\vartheta}_n(\omega))$$

sowie

$$\sup_{\vartheta \in B_\delta(\theta)} \left\| \frac{1}{n} \nabla \nabla_\theta^T L_n(\omega; \vartheta) - V(\theta) \right\| < \alpha,$$

also auch

$$\nabla \nabla_\theta^T L_n(\omega; \vartheta) \text{ ist negativ definit für alle } \vartheta \in B_\delta(\theta)$$

nach Wahl von α . Somit kann also die Gleichung $\nabla_\theta L_n(\omega; \vartheta) = 0$ für $n \geq n_0$ nur höchstens eine Lösung $\vartheta \in B_\delta(\theta)$ haben, und es folgt

$$\forall n \geq n_0 : \hat{\theta}_n(\omega) = \hat{\vartheta}_n(\omega).$$

Da wegen $\hat{\theta}_n(\omega) \in B_\delta(\theta)$ insbesondere auch $\nabla \nabla_\theta^T L(\omega; \hat{\theta}_n(\omega))$ negativ definit ist, ist $\hat{\theta}_n(\omega)$ eine lokale Maximalstelle von $\vartheta \mapsto L_n(\omega; \vartheta)$. Insgesamt haben wir also gezeigt, daß

$$C \subseteq \liminf_{n \rightarrow \infty} \{ \hat{\theta}_n = \hat{\vartheta}_n \} \cap \liminf_{n \rightarrow \infty} \{ \hat{\theta}_n \text{ ist lokale Maximalstelle von } L_n \}.$$

Wegen $P_\theta(C) = 1$ sind damit sowohl die Eindeutigkeit als auch (2.4.10) für eine stark konsistente Lösung bewiesen. Im Fall, daß $(\hat{\theta}_n)_n$ lediglich eine konsistente Lösung der Likelihood-Gleichung liefert, erhält man mit analogen Argumenten Eindeutigkeit und Aussage (2.4.11). Schließlich gilt nach Theorem 2.3.5

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V(\theta)^{-1} W(\theta) V(\theta)^{-T}) = \mathcal{N}(0, K(\theta)^{-1})$$

wegen $K(\theta)^T = K(\theta) = W(\theta) = -V(\theta)$.

■

2.5 Bemerkungen zu optimalen Schätzfunktionen

Im letzten Abschnitt dieses Kapitels beschäftigen wir uns kurz mit optimalen Schätzfunktionen. Verschiedene Optimalitätsbegriffe für Schätzfunktionen (auch für Prozesse in stetiger Zeit) wurden insbesondere von Godambe und Heyde formuliert und untersucht; siehe z.B. [GH] für eine allgemeine Diskussion. Wir beschränken uns hier auf einige elementare Beobachtungen über Optimalität in einem Sinne, der dem asymptotischen Optimalitätsbegriff in [GH] entspricht.

Es bezeichne \mathcal{G} das System aller Funktionen $g : E^2 \times \Omega \rightarrow \mathbb{R}^d$, die die Bedingungen **C1** und **I** aus Abschnitt 2.3 sowie für alle $\theta \in \Theta$

$$g(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta),$$

$$\int_E Q_\theta(x, dy) g(x, y; \theta) = 0 \quad \text{für } \mu_\theta\text{-fast alle } x \in E \quad (2.5.1)$$

erfüllen. Für jedes $g \in \mathcal{G}$ ist dann also $(G_n)_n$ mit

$$G_n(\theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta)$$

nach Abschnitt 2.2 eine Folge von Martingal-Schätzfunktionen, und nach Theorem 2.3.5 gilt für jede konsistente Lösung $(\hat{\theta}_n)_n$ der zugehörigen Martingal-Schätzgleichung $G_n(\theta) = 0$, daß

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, V_g(\theta)^{-1} W_g(\theta) V_g(\theta)^{-T}).$$

Hierbei schreiben wir $V_g(\theta)$ und $W_g(\theta)$ für die Matrizes aus (2.3.14) und (2.2.10), um zu betonen, daß die Matrizes von g abhängen. Es ist also

$$V_g(\theta) = \tilde{\mu}_\theta(D_\theta g(\cdot, \cdot; \theta)), \quad W_g(\theta) = \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)g(\cdot, \cdot; \theta)^T).$$

Betrachte nun eine Teilmenge $\tilde{\mathcal{G}} \subseteq \mathcal{G}$; dies können zum Beispiel alle g einer bestimmten Bauart sein, etwa die in den Beispielen in Abschnitt 2.2 betrachteten linearen oder quadratischen Schätzfunktionen. Dann liegt die Frage nahe, ob die Kovarianzmatrix $V_g(\theta)^{-1} W_g(\theta) V_g(\theta)^{-T}$ in $g \in \tilde{\mathcal{G}}$ minimiert werden kann, d.h. ob es ein $g_* \in \tilde{\mathcal{G}}$ gibt mit

$$\forall g \in \tilde{\mathcal{G}}, \theta \in \Theta : V_g(\theta)^{-1} W_g(\theta) V_g(\theta)^{-T} \geq V_{g_*}(\theta)^{-1} W_{g_*}(\theta) V_{g_*}(\theta)^{-T}. \quad (2.5.2)$$

Nenne jedes $g_* \in \tilde{\mathcal{G}}$ mit (2.5.2) ein optimales Element in $\tilde{\mathcal{G}}$. Ein solches ist keineswegs eindeutig bestimmt: ist g_* optimal in $\tilde{\mathcal{G}}$ und $A \in \mathbb{R}^{d \times d}$ invertierbar, so überzeugt man sich leicht davon, daß auch Ag_* optimal in $\tilde{\mathcal{G}}$ ist.

Existiert ein optimales $g_* \in \tilde{\mathcal{G}}$, so ist also durch $V_{g_*}(\theta)^{-1}W_{g_*}(\theta)V_{g_*}(\theta)^{-T}$ eine untere Schranke für die asymptotische Varianz einer jeden konsistenten Lösung der Martingal-Schätzgleichung

$$G_n(\theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta) = 0, \quad g \in \tilde{\mathcal{G}} \text{ beliebig,}$$

gegeben. Hilfreich für die Bestimmung optimaler Elemente ist das folgende, an die Cramér-Rao-Schranke erinnernde Lemma:

Lemma 2.5.1 *Existiert $g_* \in \tilde{\mathcal{G}}$ mit*

$$\forall g \in \tilde{\mathcal{G}}, \forall \theta \in \Theta : \tilde{\mu}_\theta(g(\cdot, \cdot; \theta)g_*(\cdot, \cdot; \theta)^T) = -V_g(\theta), \quad (2.5.3)$$

so ist g_ optimal in $\tilde{\mathcal{G}}$.*

Beweis Die Funktion $g_* \in \tilde{\mathcal{G}}$ erfülle (2.5.3). Dann gilt insbesondere

$$\forall \theta \in \Theta : W_{g_*}(\theta) = \tilde{\mu}_\theta(g_*(\theta)g_*(\theta)^T) = -V_{g_*}(\theta) \quad (2.5.4)$$

und $V_{g_*}(\theta)$ ist symmetrisch, also vereinfacht sich die rechte Seite von (2.5.2) zu $-V_{g_*}(\theta)^{-1}W_{g_*}(\theta)^{-1}$. Durch Multiplikation mit $V_g(\theta)$ von links und $V_g(\theta)^T$ von rechts sieht man dann, daß (2.5.2) äquivalent ist zu

$$\forall g \in \tilde{\mathcal{G}}, \theta \in \Theta : W_g(\theta) \geq -V_g(\theta)V_{g_*}(\theta)^{-1}V_g(\theta)^T. \quad (2.5.5)$$

Sei nun $g \in \tilde{\mathcal{G}}$ beliebig. Setze

$$h(x, y; \theta) := g(x, y; \theta) - V_g(\theta)V_{g_*}(\theta)^{-1}g_*(x, y; \theta).$$

Dann gilt für jedes $\theta \in \Theta$

$$\begin{aligned} 0 &\leq \tilde{\mu}_\theta(h(\theta)h(\theta)^T) \\ &= \tilde{\mu}_\theta(g(\theta)g(\theta)^T) - \tilde{\mu}_\theta\left(g(\theta)\left(V_g(\theta)V_{g_*}(\theta)^{-1}g_*(\theta)\right)^T\right) - \tilde{\mu}_\theta\left(V_g(\theta)V_{g_*}(\theta)^{-1}g_*(\theta)g(\theta)^T\right) \\ &\quad + \tilde{\mu}_\theta\left(V_g(\theta)V_{g_*}(\theta)^{-1}g_*(\theta)\left(V_g(\theta)V_{g_*}(\theta)^{-1}g_*(\theta)\right)^T\right) \\ &= \tilde{\mu}_\theta(g(\theta)g(\theta)^T) - \underbrace{\tilde{\mu}_\theta(g(\theta)g_*(\theta)^T)}_{=-V_g(\theta)} V_{g_*}(\theta)^{-1}V_g(\theta)^T - V_g(\theta)V_{g_*}(\theta)^{-1} \underbrace{\tilde{\mu}_\theta(g_*(\theta)g(\theta)^T)}_{=-V_g(\theta)^T} \\ &\quad + V_g(\theta)V_{g_*}(\theta)^{-1} \underbrace{\tilde{\mu}_\theta(g_*(\theta)g_*(\theta)^T)}_{=-V_{g_*}(\theta)} V_{g_*}(\theta)^{-1}V_g^T(\theta) \\ &= W_g(\theta) + V_g(\theta)V_{g_*}(\theta)^{-1}V_g(\theta)^T + V_g(\theta)V_{g_*}(\theta)^{-1}V_g(\theta)^T - V_g(\theta)V_{g_*}(\theta)^{-1}V_{g_*}(\theta)V_{g_*}(\theta)^{-1}V_g(\theta)^T \\ &= W_g(\theta) + V_g(\theta)V_{g_*}(\theta)^{-1}V_g(\theta)^T, \end{aligned}$$

das ist (2.5.5).

Bemerkungen

1. Existiert ein Element $g_* \in \tilde{\mathcal{G}}$ mit (2.5.3), so ist $V_{g_*}(\theta) = -\tilde{\mu}_\theta(g_*(\theta)g_*(\theta)^T)$ negativ semidefinit und nach Voraussetzung **I** invertierbar, also auch strikt negativ definit. Damit erfüllt g_* die Bedingungen **C1** und **D** aus Abschnitt 2.3. Unter der Voraussetzung, daß der Parameter $\theta \in \Theta$ überhaupt konsistent schätzbar ist, gibt es also nach Theorem 2.3.8 eine konsistente Lösung $(\hat{\theta}_n^*)_n$ der Schätzgleichung

$$G_n^*(\theta) := \sum_{k=1}^n g_*(X_{k-1}, X_k; \theta) = 0,$$

für die

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n^* - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W_{g_*}(\theta)^{-1}) \quad \text{unter } P_\theta$$

gilt; die untere Schranke $-V_{g_*}(\theta)^{-1} = W_{g_*}(\theta)^{-1}$ wird dann also erreicht. Entsprechende Schätzfolgen werden auch als *Quasi-Likelihood-Schätzfolgen* bezeichnet.

2. Anders als durch die Optimalitätsbedingung (2.5.2) ist das Element g_* durch die Bedingung (2.5.3) in Lemma 2.5.1 im wesentlichen eindeutig bestimmt: seien g_* , \tilde{g}_* mit (2.5.3) gegeben. Schreiben wir $\text{tr}(A)$ für die Spur einer Matrix $A \in \mathbb{R}^{d \times d}$ und verwenden wir die euklidische Norm auf \mathbb{R}^d ,

$$\|x\| := \sqrt{x^T x} = \sqrt{\text{tr}(xx^T)}, \quad x \in \mathbb{R}^d,$$

so ergibt sich aus (2.5.3)

$$\begin{aligned} & \forall \theta \in \Theta : \tilde{\mu}_\theta (\|g_*(\theta) - \tilde{g}_*(\theta)\|^2) \\ &= \tilde{\mu}_\theta (\text{tr} [(g_*(\theta) - \tilde{g}_*(\theta))(g_*(\theta) - \tilde{g}_*(\theta))^T]) \\ &= \text{tr} (\tilde{\mu}_\theta (g_*(\theta)g_*(\theta)^T) - \tilde{\mu}_\theta (\tilde{g}_*(\theta)g_*(\theta)^T) - \tilde{\mu}_\theta (g_*(\theta)\tilde{g}_*(\theta)^T) + \tilde{\mu}_\theta (\tilde{g}_*(\theta)\tilde{g}_*(\theta)^T)) \\ &= \text{tr} (-V_{g_*}(\theta) + V_{\tilde{g}_*}(\theta) + V_{g_*}(\theta) - V_{\tilde{g}_*}(\theta)) \\ &= 0, \end{aligned}$$

d.h. für alle $\theta \in \Theta$ gilt $g_*(\cdot, \cdot; \theta) = \tilde{g}_*(\cdot, \cdot; \theta)$ $\tilde{\mu}_\theta$ -fast sicher.

3. Der Prozeß $(X_n)_n$ erfülle zusätzlich die Regularitätsvoraussetzungen aus Abschnitt 2.4, d.h. es sei

$$Q_\theta(x, dy) = p(x, y; \theta) \lambda(dy)$$

mit einer strikt positiven und in θ stetig differenzierbaren Übergangsdichte p . Definiere wieder $\ell(x, y; \theta) = \log p(x, y; \theta)$ und setze voraus, daß $\nabla_\theta \ell \in \mathcal{G}$. Für beliebiges $g \in \mathcal{G}$ ist dann

$$D_\theta(p \cdot g) = g \cdot (\nabla_\theta p)^T + p \cdot D_\theta g.$$

Unter der zusätzlichen Voraussetzung, daß für jedes $g \in \mathcal{G}$ die Funktion

$$\theta \mapsto \int_E g(x, y; \theta) p(x, y; \theta) \lambda(dy)$$

komponentenweise partiell unter dem Integralzeichen differenziert werden darf, gilt dann wegen (2.5.1) für μ_θ -fast alle $x \in E$

$$\begin{aligned} 0 &= D_\theta \left[\underbrace{\int_E g(x, y; \theta) \cdot p(x, y; \theta) \lambda(dy)}_{=0 \text{ für } \mu_\theta\text{-fast alle } x} \right] \\ &= \int_E D_\theta [g(x, y; \theta) \cdot p(x, y; \theta)] \lambda(dy) \\ &= \int_E g(x, y; \theta) [\nabla_\theta \ell(x, y; \theta)]^T p(x, y; \theta) \lambda(dy) + \int_E p(x, y; \theta) D_\theta g(x, y; \theta) \lambda(dy), \end{aligned}$$

also auch

$$\begin{aligned} \tilde{\mu}_\theta (g(\theta) [\nabla_\theta \ell(\theta)]^T) &= \int_E \mu_\theta(dx) \int_E g(x, y; \theta) [\nabla_\theta \ell(x, y; \theta)]^T p(x, y; \theta) \lambda(dy) \\ &= - \int_E \mu_\theta(dx) \int_E p(x, y; \theta) D_\theta g(x, y; \theta) \lambda(dy) \\ &= -\tilde{\mu}_\theta (D_\theta g(\theta)) \\ &= -V_g(\theta). \end{aligned}$$

Also ist $\nabla_\theta \ell$ nach Lemma 2.5.1 ein optimales Element in \mathcal{G} , und die zugehörige untere Schranke $W_{\nabla_\theta \ell}(\theta)^{-1}$ ist gerade die Inverse der Fisher-Information. ◆

Im Rest dieses Abschnitts soll gezeigt werden, wie Lemma 2.5.1 zur expliziten Berechnung optimaler Elemente verwendet werden kann. Wir betrachten dabei die Teilmenge

$$\tilde{\mathcal{G}} = \mathcal{G}_2 := \left\{ g \in \mathcal{G} : g(x, y; \theta) = w_m(x; \theta)(y - m(x; \theta)) + w_v(\theta) \left((y - m(x; \theta))^2 - v(x; \theta) \right) \right\},$$

wobei

$$m(x; \theta) := E_{\mu_\theta}[X_n | X_{n-1} = x], \quad v(x; \theta) := \text{Var}_{\mu_\theta}[X_n | X_{n-1} = x]$$

Festlegungen der bedingten Erwartung bzw. Varianz von X_n gegeben $X_{n-1} = x$ sind, siehe das Beispiel in Abschnitt 2.2. Die durch $g \in \mathcal{G}_2$ induzierten Schätzfunktionen G_n sind gerade die dort angesprochenen quadratischen Schätzfunktionen. Gesucht sind die optimalen Gewichte w_m^* und w_v^* . Wir nehmen an, daß der Prozeß $(X_n)_n$ unter P_{μ_θ} endliche vierte Momente besitzt. Dann sind auch die dritten und vierten zentrierten bedingten Momente

$$\mu_3(x; \theta) := E_{\mu_\theta} [(X_n - m(X_{n-1}; \theta))^3 | X_{n-1} = x],$$

$$\mu_4(x; \theta) := E_{\mu_\theta} [(X_n - m(X_{n-1}; \theta))^4 | X_{n-1} = x]$$

wohldefiniert.

Wir wollen nun ein optimales Element g_* durch den Ansatz

$$\tilde{\mu}_\theta (g(\theta)g_*(\theta)^T) \stackrel{!}{=} -V_g(\theta) \quad (2.5.6)$$

bestimmen. Für beliebiges $g \in \mathcal{G}_2$ und $\theta \in \Theta$ ist

$$\begin{aligned} D_\theta g(x, y; \theta) &= D_\theta [(y - m(x; \theta)) \cdot w_m(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta)) \cdot w_v(x; \theta)] \\ &= (y - m(x; \theta)) D_\theta w_m(x; \theta) - w_m(x; \theta) [\nabla_\theta m(x; \theta)]^T + ((y - m(x; \theta))^2 - v(x; \theta)) D_\theta w_v(x; \theta) \\ &\quad - 2(y - m(x; \theta)) w_v(x; \theta) [\nabla_\theta m(x; \theta)]^T - w_v(x; \theta) [\nabla_\theta v(x; \theta)]^T; \end{aligned}$$

also berechnet sich die Matrix $V_g(\theta)$ zu

$$\begin{aligned} V_g(\theta) &= \tilde{\mu}_\theta (D_\theta g(\theta)) \\ &= \underbrace{E_{\mu_\theta} [(X_1 - m(X_0; \theta)) \cdot D_\theta w_m(X_0; \theta)]}_{=0} \\ &\quad + \underbrace{E_{\mu_\theta} [((X_1 - m(X_0; \theta))^2 - v(X_0; \theta)) \cdot D_\theta w_v(X_0; \theta)]}_{=0} \\ &\quad - 2 \underbrace{E_{\mu_\theta} [(X_1 - m(X_0; \theta)) \cdot w_v(X_0; \theta) \nabla_\theta m(X_0; \theta)^T]}_{=0} \\ &\quad - E_{\mu_\theta} [w_m(X_0; \theta) [\nabla_\theta m(X_0; \theta)]^T + w_v(X_0; \theta) [\nabla_\theta v(X_0; \theta)]^T] \\ &= -\mu_\theta (w_m(\cdot; \theta) [\nabla_\theta m(\cdot; \theta)]^T + w_v(\cdot; \theta) [\nabla_\theta v(\cdot; \theta)]^T). \end{aligned} \quad (2.5.7)$$

Für das gesuchte Element $g_* \in \mathcal{G}_2$ mit

$$g_*(x, y; \theta) = (y - m(x; \theta)) w_m^*(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta)) w_v^*(x; \theta) \quad (2.5.8)$$

ergibt sich

$$\begin{aligned} g(x, y; \theta) g_*(x, y; \theta)^T &= (y - m(x; \theta))^2 w_m(x; \theta) w_m^*(x; \theta)^T + ((y - m(x; \theta))^2 - v(x; \theta))^2 w_v(x; \theta) w_v^*(x; \theta)^T \\ &\quad + (y - m(x; \theta)) ((y - m(x; \theta))^2 - v(x; \theta)) [w_m(x; \theta) w_v^*(x; \theta)^T + w_v(x; \theta) w_m^*(x; \theta)^T], \end{aligned}$$

und somit

$$\begin{aligned}
& \tilde{\mu}_\theta(g(\theta)g_*(\theta)^T) \\
&= E_{\mu_\theta} [(X_1 - m(X_0; \theta))^2 w_m(X_0; \theta) w_m^*(X_0; \theta)^T] \\
&\quad + E_{\mu_\theta} \left[\left((X_1 - m(X_0; \theta))^2 - v(X_0; \theta) \right)^2 w_v(X_0; \theta) w_v^*(X_0; \theta)^T \right] \\
&\quad + E_{\mu_\theta} \left[(X_1 - m(X_0; \theta))^3 (w_m(X_0; \theta) w_v^*(X_0; \theta)^T + w_v(X_0; \theta) w_m^*(X_0; \theta)^T) \right] \\
&\quad - \underbrace{E_{\mu_\theta} \left[(X_1 - m(X_0; \theta)) v(X_0; \theta) (w_m(X_0; \theta) w_v^*(X_0; \theta)^T + w_v(X_0; \theta) w_m^*(X_0; \theta)^T) \right]}_{=0} \\
&= E_{\mu_\theta} [v(X_0; \theta) w_m(X_0; \theta) w_m^*(X_0; \theta)^T] \\
&\quad + E_{\mu_\theta} [(\mu_4(X_0; \theta) - v(X_0; \theta)^2) w_v(X_0; \theta) w_v^*(X_0; \theta)^T] \\
&\quad + E_{\mu_\theta} [\mu_3(X_0; \theta) (w_m(X_0; \theta) w_v^*(X_0; \theta)^T + w_v(X_0; \theta) w_m^*(X_0; \theta)^T)] \\
&= \mu_\theta \left(v(\cdot; \theta) w_m(\cdot; \theta) w_m^*(\cdot; \theta)^T + (\mu_4(\cdot; \theta) - v(\cdot; \theta)^2) w_v(\cdot; \theta) w_v^*(\cdot; \theta)^T \right. \\
&\quad \left. + \mu_3(\cdot; \theta) (w_m(\cdot; \theta) w_v^*(\cdot; \theta)^T + w_v(\cdot; \theta) w_m^*(\cdot; \theta)^T) \right) \\
&= \mu_\theta \left(w_m [v(\cdot; \theta) w_m^*(\cdot; \theta) + \mu_3(\cdot; \theta) w_v^*(\cdot; \theta)]^T \right. \\
&\quad \left. + w_v(\cdot; \theta) [(\mu_4(\cdot; \theta) - v(\cdot; \theta)^2) w_v^*(\cdot; \theta) + \mu_3(\cdot; \theta) w_m^*(\cdot; \theta)]^T \right). \tag{2.5.9}
\end{aligned}$$

Der Ansatz (2.5.6) liefert dann zusammen mit (2.5.7) und (2.5.9)

$$v(x; \theta) w_m^*(x; \theta) + \mu_3(x; \theta) w_v^*(x; \theta) \stackrel{!}{=} \nabla_\theta m(x; \theta), \tag{2.5.10}$$

$$(\mu_4(x; \theta) - v(x; \theta)^2) w_v^*(x; \theta) + \mu_3(x; \theta) w_m^*(x; \theta) \stackrel{!}{=} \nabla_\theta v(x; \theta). \tag{2.5.11}$$

Auflösen nach w_m^* und w_v^* liefert dann die optimalen Gewichte

$$w_m^*(x; \theta) = \frac{(\mu_4(x; \theta) - v(x; \theta)^2) \nabla_\theta m(x; \theta) - \mu_3(x; \theta) \nabla_\theta v(x; \theta)}{v(x; \theta) (\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2} \tag{2.5.12}$$

$$w_v^*(x; \theta) = \frac{v(x; \theta) \nabla_\theta v(x; \theta) - \mu_3(x; \theta) \nabla_\theta m(x; \theta)}{v(x; \theta) (\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2}. \tag{2.5.13}$$

Wir haben mittels der obigen Rechnungen also gezeigt:

Satz 2.5.2 *Die Ausdrücke (2.5.12) und (2.5.13) seien wohldefiniert. Definiere g_* durch (2.5.8). Gilt dann $g_* \in \mathcal{G}$, so erfüllt g_* die Voraussetzungen von Lemma 2.5.1 und ist somit insbesondere ein optimales Element in \mathcal{G}_2 .*

Bemerkungen Die Ausdrücke (2.5.12) und (2.5.13) für die optimalen Gewichte w_m^* und w_v^* sind eine d -dimensionale Version von (1.9) - (1.11) in [Wef].

Man beachte, daß der Nenner in (2.5.12) und (2.5.13) stets positiv ist, falls das Wahrscheinlichkeitsmaß $Q_\theta(x, dy)$ keine Punktmassen hat: nach der Cauchy-Schwarz-Ungleichung gilt

$$\begin{aligned} \mu_3(x; \theta)^2 &= \left(E_\theta^x [(X_1 - m(X_0; \theta))^3] \right)^2 \\ &= \left(E_\theta^x [(X_1 - m(x; \theta))^3] - \underbrace{E_\theta^x [v(x; \theta)(X_1 - m(x; \theta))]}_{=0} \right)^2 \\ &= \left(E_\theta^x [((X_1 - m(x; \theta))^2 - v(x; \theta)) \cdot (X_1 - m(x; \theta))] \right)^2 \\ &\leq E_\theta^x [((X_1 - m(x; \theta))^2 - v(x; \theta))^2] \cdot E_\theta^x [(X_1 - m(x; \theta))^2] \\ &= (\mu_4(x; \theta) - v(x; \theta)^2) \cdot v(x; \theta) \end{aligned}$$

mit Gleichheit genau dann, wenn es eine Konstante $K = K(x, \theta) \in \mathbb{R}$ gibt mit

$$|(X_1 - m(x; \theta))^2 - v(x; \theta)| = K \cdot |X_1 - m(x; \theta)| \quad P_\theta^x\text{-f.s.}, \quad (2.5.14)$$

siehe etwa [LL], Theorem 2.3. Gilt aber (2.5.14), so erfüllt X_1 P_θ^x -fast sicher eine von zwei quadratischen Gleichungen, und $\mathcal{L}(X_1 | P_\theta^x) = Q_\theta(x, \cdot)$ ist somit auf endlich viele Punkte konzentriert. \blacklozenge

Kapitel 3

Parameterschätzung im diskretisierten CIR-Modell

3.1 Vorbereitungen

In diesem Kapitel kehren wir zum diskretisierten **CIR**-Modell zurück. Mit den Bezeichnungen aus Kapitel 2 ist dann also

$$E := \mathbb{R}_{\geq 0}, \quad \Theta := \mathbb{R}_{> 0} \times \mathbb{R}_{< 0} \times \mathbb{R}_{> 0}.$$

Für die Elemente des Parameterraums schreiben wir statt $\theta = (\theta_1, \theta_2, \theta_3)$ auch weiterhin meistens (a, b, σ^2) . Partielle Ableitungen nach den Variablen a, b, σ^2 bezeichnen wir mit $\partial_a, \partial_b, \partial_{\sigma^2}$ oder auch $\partial_i, i = 1, \dots, 3$; es ist also $\partial_1 \equiv \partial_a, \partial_2 \equiv \partial_b, \partial_3 \equiv \partial_{\sigma^2}$. Mit λ bezeichnen wir das Lebesguemaß auf $E = \mathbb{R}_{\geq 0}$ sowie mit $\lambda^2 := \lambda \otimes \lambda$ das Lebesguemaß auf E^2 . Integrale bezüglich λ werden mit $\int_E f(y) dy := \int_E f(y) \lambda(dy)$ bezeichnet.

Betrachte nun wieder den diskretisierten **CIR**-Prozeß $(X_n)_{n \in \mathbb{N}_0}$ aus Abschnitt 1.2, aufgefaßt als kanonischer Prozeß auf dem kanonischen Pfadraum. Nach Abschnitt 1.2 ist dieser für jedes $\theta = (a, b, \sigma^2) \in \Theta$ ein homogener Markov-Prozeß mit Übergangswahrscheinlichkeit

$$Q_\theta(x, dy) = p(x, y; \theta) dy, \quad x \in E,$$

wobei p die Übergangsdichte aus (1.2.3) ist. Wegen $p(x, y; \theta) > 0$ für $y > 0$ gilt insbesondere

$$\forall \theta \in \Theta, x \in E : Q_\theta(x, dy) \sim \lambda. \tag{3.1.1}$$

Nach Abschnitt 1.2 existiert zu jedem $Q_\theta, \theta \in \Theta$, ein eindeutiges invariantes Wahrscheinlichkeitsmaß, nämlich die Gamma-Verteilung

$$\mu_\theta(dx) := \Gamma(\alpha(\theta), r(\theta))(dx) = \gamma(x; \theta) dx,$$

wobei nach (1.1.9)

$$\alpha(\theta) = q(\theta) + 1 = \frac{2a}{\sigma^2}, \quad r(\theta) = \frac{-2b}{\sigma^2}. \quad (3.1.2)$$

Die Verteilung hat also die Lebesgue-Dichte

$$\gamma(x; \theta) = \frac{r(\theta)^{\alpha(\theta)}}{\Gamma(\alpha(\theta))} x^{\alpha(\theta)-1} e^{-r(\theta)x}, \quad x \in E.$$

Wegen $\gamma(x; \theta) > 0$ für $x > 0$ ist $\mu_\theta \sim \lambda$ für alle $\theta \in \Theta$. Mit (3.1.1) folgt daher insbesondere

$$\forall \theta \in \Theta, x \in E : Q_\theta(x, dy) \ll \mu_\theta(dy).$$

Nach Satz 2.1.2 ist folglich die Ergodizitätsannahme 2.1.1 erfüllt. Damit erfüllt der Prozeß $(X_n)_n$ alle Voraussetzungen aus Abschnitt 2.1. Wir halten außerdem fest, daß das Maß $\tilde{\mu}_\theta$ aus (2.1.2) durch eine λ^2 -Dichte gegeben ist,

$$\tilde{\mu}_\theta(d(x, y)) = \gamma(x; \theta)p(x, y; \theta) dx dy.$$

Wegen $\gamma(x; \theta)p(x, y; \theta) > 0$ für $x, y > 0$ gilt dabei auch

$$\tilde{\mu}_\theta \sim \lambda^2, \quad \theta \in \Theta. \quad (3.1.3)$$

Schließlich erhalten wir aus Satz 1.1.4, daß die Momente beliebig hoher Ordnung des Prozesses endlich sind. Insbesondere gilt

$$E_\theta^x [X_1^n] = \sum_{k=0}^n \binom{n}{k} (xe^{b\Delta})^{n-k} \left(\frac{e^{b\Delta} - 1}{2b} \right)^k \prod_{j=1}^k (2a + \sigma^2(n - k - 1 + j)), \quad n \in \mathbb{N}_0. \quad (3.1.4)$$

Wir werden im folgenden die im zweiten Kapitel entwickelte Theorie auf das **CIR**-Modell anwenden und stark konsistente sowie asymptotisch normale Schätzfolgen für den Parameter θ angeben sowie lokal-asymptotische Normalität beweisen.

3.2 Schätzung mittels quadratischer Schätzfunktionen

In diesem Abschnitt betrachten wir Schätzfolgen für den Parameter θ im **CIR**-Modell, die durch stark konsistente Lösungen von Martingal-Schätzgleichungen der Form

$$G_n(\theta) = \sum_{k=1}^n g(X_{k-1}, X_k; \theta) = 0 \quad (3.2.1)$$

definiert sind. Dabei soll $(G_n)_n$ eine Folge von *quadratischen* Martingal-Schätzfunktionen sein, d.h. in der Terminologie von Abschnitt 2.5 $g \in \mathcal{G}_2$ sein. Es sei kurz daran erinnert, daß dies bedeutet: g ist von der Gestalt

$$g(x, y; \theta) = (y - m(x; \theta))w_m(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta))w_v(x; \theta)$$

und erfüllt die Bedingungen **C1** und **I** aus Abschnitt 2.3 sowie $g(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ für jedes $\theta \in \Theta$. Dabei bezeichnet $m(x; \theta)$ eine Festlegung der bedingten Erwartung $E_{\mu_\theta} [X_n | X_{n-1} = x]$ und $v(x; \theta)$ eine Festlegung der bedingten Varianz $E_{\mu_\theta} [(X_n - m(X_{n-1}; \theta))^2 | X_{n-1} = x]$. Diese existieren, da der Prozeß unter P_{μ_θ} endliche Momente beliebig hoher Ordnung hat. Nach der Markov-Eigenschaft gilt

$$E_{\mu_\theta} [X_n | X_{n-1}] = E_\theta^{X_{n-1}} [X_1] \quad P_{\mu_\theta}\text{-f.s.},$$

also können wir mit der Festlegung

$$m(x; \theta) = E_\theta^x [X_1] \quad (3.2.2)$$

arbeiten. Mit Gleichung (3.1.4) berechnet sich $m(x; \theta)$ für $x \in E, \theta \in \Theta$ zu

$$m(x; \theta) = xe^{b\Delta} + \frac{e^{b\Delta} - 1}{2b} \cdot 2a =: \gamma_0(\theta) + \gamma_1(\theta)x, \quad (3.2.3)$$

wobei

$$\gamma_0(\theta) := \frac{a}{b} (e^{b\Delta} - 1) > 0, \quad \gamma_1(\theta) := e^{b\Delta} > 0. \quad (3.2.4)$$

Ganz analog arbeiten wir mit der Festlegung

$$v(x; \theta) = E_\theta^x [X_1^2] - m(x; \theta)^2 \quad (3.2.5)$$

für die bedingte Varianz; wieder mit (3.1.4) folgt für $x \in E, \theta \in \Theta$

$$\begin{aligned} v(x; \theta) &= E_\theta^x [X_1^2] - m(x; \theta)^2 \\ &= x^2 e^{2b\Delta} + 2xe^{b\Delta} \frac{e^{b\Delta} - 1}{2b} (2a + \sigma^2) + \left(\frac{e^{b\Delta} - 1}{2b} \right)^2 2a(2a + \sigma^2) - \left(xe^{b\Delta} + \frac{a}{b} (e^{b\Delta} - 1) \right)^2 \\ &= xe^{b\Delta} \frac{e^{b\Delta} - 1}{b} \sigma^2 + \frac{(e^{b\Delta} - 1)^2}{2b^2} a \sigma^2 \\ &=: \sigma^2 (\eta_0(\theta) + \eta_1(\theta)x), \end{aligned}$$

wobei

$$\eta_0(\theta) := \frac{a}{2b^2} (e^{b\Delta} - 1)^2 > 0, \quad \eta_1(\theta) := \frac{1}{b} e^{b\Delta} (e^{b\Delta} - 1) > 0. \quad (3.2.6)$$

Offenbar gilt $\gamma_0, \gamma_1, \eta_0, \eta_1 \in C^\infty(\Theta)$. Wir beobachten dabei, daß $\gamma_0, \gamma_1, \eta_0, \eta_1$ und somit auch m ausschließlich von den Variablen a und b und nicht von σ^2 abhängen. Wir können sie daher je nach Zweckmäßigkeit als Funktionen von (a, b) allein oder als in σ^2 konstante Funktionen von $\theta = (a, b, \sigma^2)$ auffassen.

Wir benötigen noch wie in Abschnitt 2.5 die zentrierten bedingten Momente dritter und vierter Ordnung μ_3 und μ_4 . Analog zur obigen Herleitung von m und v berechnet man unter Verwendung von (3.1.4) Festlegungen von μ_3 und μ_4 zu

$$\mu_3(x; \theta) = \mu_{30}(\theta) + \mu_{31}(\theta)x; \quad (3.2.7)$$

$$\mu_4(x; \theta) = \mu_{40}(\theta) + \mu_{41}(\theta)x + \mu_{42}(\theta)x^2 \quad (3.2.8)$$

wobei

$$\begin{aligned} \mu_{30}(\theta) &= \frac{a\sigma^4}{2b^3}(e^{b\Delta} - 1)^3 > 0, & \mu_{40}(\theta) &= \frac{3a\sigma^4(a + \sigma^2)}{4b^4}(e^{b\Delta} - 1)^4 > 0 \\ \mu_{31}(\theta) &= \frac{3\sigma^4}{2b^2}e^{b\Delta}(e^{b\Delta} - 1)^2 > 0, & \mu_{41}(\theta) &= \frac{3\sigma^4(a + \sigma^2)}{b^3}e^{b\Delta}(e^{b\Delta} - 1)^3 > 0, \\ \mu_{42}(\theta) &= \frac{3\sigma^4}{b^2}e^{2b\Delta}(e^{b\Delta} - 1)^2 > 0. \end{aligned} \quad (3.2.9)$$

Schließlich sei kurz bemerkt, daß wegen $X_n \geq 0$ für alle $n \in \mathbb{N}_0$ für den **CIR**-Prozeß sämtliche bedingten Momente nicht nur unter P_{μ_θ} , sondern sogar unter jedem P_θ^ν mit beliebiger Startverteilung ν wohldefiniert und durch die obigen Festlegungen gegeben sind.

Wir werden im folgenden zunächst den *Conditional Least Squares*-Ansatz (im folgenden kurz: **CLS**-Ansatz) verwenden, um eine stark konsistente und asymptotisch normale Schätzfolge für den Parameter θ zu konstruieren. Diese Schätzfolge hat eine explizite Darstellung und löst eine quadratische Martingal-Schätzgleichung der Bauart (3.2.1). Damit steht dann eine Folge von Hilfsschätzern zur Verfügung, die die Anwendung der Resultate aus Kapitel 2, insbesondere von Theorem 2.3.8, auf das **CIR**-Modell ermöglicht. Die Darstellung in diesem Unterabschnitt folgt der Arbeit [OR].

3.2.1 Conditional Least Squares-Schätzung

Der **CLS**-Ansatz wurde insbesondere von Klimko und Nelson eingehend untersucht; siehe [KN]. Ein kurzer Überblick findet sich in [HH]. Man geht dabei aus von einem (nicht notwendig Markovschen) zeitdiskreten stochastischen Prozeß $(\tilde{X}_n)_{n \in \mathbb{N}_0}$ auf einem Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}_\theta)$, dessen Verteilung von einem Parameter $\theta \in \tilde{\Theta}$, $\tilde{\Theta} \subseteq \mathbb{R}^d$ offen, abhängt. Mit $\tilde{\mathcal{F}}_n := \sigma(\tilde{X}_k : 0 \leq k \leq n)$ sei für alle $n \in \mathbb{N}$ und $\theta \in \tilde{\Theta}$ die bedingte Erwartung $E_\theta(\tilde{X}_n | \tilde{\mathcal{F}}_{n-1})$ definiert. Für eine endliche Folge von Beobachtungen $(\tilde{X}_0(\omega), \dots, \tilde{X}_n(\omega))$ betrachtet man nun die Quadratsumme der Fehler der „besten Prognose“

$$Q_n(\omega; \theta) := \sum_{k=1}^n \left[\tilde{X}_k(\omega) - E_\theta(\tilde{X}_k | \tilde{\mathcal{F}}_{k-1})(\omega) \right]^2 \quad (3.2.10)$$

in Abhängigkeit von $\theta \in \tilde{\Theta}$. Beim **CLS**-Verfahren sucht man für festes $\omega \in \tilde{\Omega}$ nach (relativen oder sogar globalen) Minima der Funktion $\theta \mapsto Q_n(\omega; \theta)$. Ist diese Funktion differenzierbar in θ , so gilt für jede Minimalstelle θ^* von $Q_n(\omega; \cdot)$ notwendig

$$\nabla_\theta Q_n(\omega; \theta^*) = 0. \quad (3.2.11)$$

Gleichung (3.2.11) heißt *Least-Squares-Equation*. Unter geeigneten Bedingungen erhält man somit eine Schätzfolge für den unbekannt Parameter θ , die (asymptotisch) die Least-Squares-Gleichung (3.2.11) löst.

Schätzung des Driftkoeffizienten

Betrachte nun wieder den **CIR**-Prozeß. Nach den Überlegungen aus Abschnitt (3.1) sind die bedingten Erwartungen $E_\theta(X_n | \mathcal{F}_{n-1})$ stets definiert und gegeben durch

$$E_\theta [X_n | \mathcal{F}_{n-1}] = m(X_{n-1}; \theta) = \gamma_0(\theta) + \gamma_1(\theta)X_{n-1} \quad P_\theta\text{-f.s.}$$

mit m , γ_0 und γ_1 aus (3.2.2) und (3.2.4). Für die Funktion Q_n aus (3.2.10) ergibt sich

$$Q_n(\theta) = \sum_{k=1}^n [X_k - m(X_{k-1}; \theta)]^2 = \sum_{k=1}^n [X_k - \gamma_0(\theta) - \gamma_1(\theta)X_{k-1}]^2,$$

und die Least-Squares-Gleichung (3.2.11) schreibt sich als

$$0 = \nabla_\theta Q_n(\theta) = -2 \sum_{k=1}^n (X_k - m(X_{k-1}; \theta)) \cdot \nabla_\theta m(X_{k-1}; \theta). \quad (3.2.12)$$

Die Least-Squares-Gleichung ist also eine lineare Martingal-Schätzgleichung der Bauart (3.2.1) und läßt sich somit als Spezialfall der allgemeinen Theorie aus Kapitel 2 auffassen.

Da wie schon bemerkt γ_0 und γ_1 ausschließlich von a und b abhängen, ist m und damit auch Q_n konstant entlang der θ_3 -Koordinatenrichtung. Durch Lösen der Gleichung (3.2.12) können wir also sinnvollerweise nur den Driftkoeffizienten (a, b) schätzen; für die Schätzung von σ^2 wird ein anderer Ansatz benötigt. Dies geschieht im nächsten Unterabschnitt. Aus dem genannten Grund zerlegen wir den Parameterraum in

$$\Theta =: \Phi \times \mathbb{R}_{>0}$$

mit $\Phi := \mathbb{R}_{>0} \times \mathbb{R}_{<0}$ und schreiben

$$\theta =: (\phi, \sigma^2) \quad \text{mit} \quad \phi = (\phi_1, \phi_2) := (a, b).$$

Wir fassen auch Q_n je nach Zweckmäßigkeit als Funktion von ϕ oder als in σ^2 konstante Funktion von $\theta = (\phi, \sigma^2)$ auf. Aus der Least-Squares-Gleichung (3.2.12) erhalten wir dann nach Multiplikation mit $-1/2$ das Gleichungssystem

$$\begin{aligned} 0 &= -\frac{1}{2} \partial_a Q_n(\phi) = \sum_{k=1}^n (X_k - m(X_{k-1}; \phi)) \cdot \partial_a m(X_{k-1}; \phi) \\ 0 &= -\frac{1}{2} \partial_b Q_n(\phi) = \sum_{k=1}^n (X_k - m(X_{k-1}; \phi)) \cdot \partial_b m(X_{k-1}; \phi), \end{aligned} \quad (3.2.13)$$

welches in ϕ zu lösen ist. Zur Abkürzung führen wir folgende Schreibweisen ein:

$$\begin{aligned}\bar{X}_n &:= \frac{1}{n} \sum_{k=1}^n X_k, & \bar{X}'_n &:= \frac{1}{n} \sum_{k=1}^n X_{k-1}, \\ \Sigma_n^{(1)} &:= \sum_{k=1}^n (X_k - \bar{X}_n) (X_{k-1} - \bar{X}'_n) = \sum_{k=1}^n X_k X_{k-1} - n \bar{X}_n \bar{X}'_n, \\ \Sigma_n^{(2)} &:= \sum_{k=1}^n (X_{k-1} - \bar{X}'_n)^2 = \sum_{k=1}^n X_{k-1}^2 - n (\bar{X}'_n)^2.\end{aligned}\tag{3.2.14}$$

Mit diesen Konventionen ergibt sich

$$\begin{aligned}0 &\stackrel{!}{=} \sum_{k=1}^n (X_k - m(X_{k-1}; \phi)) \cdot \partial_a m(X_{k-1}; \phi) \\ &= \sum_{k=1}^n (X_k - \gamma_0(\phi) - \gamma_1(\phi) X_{k-1}) (\partial_a \gamma_0(\phi) + \underbrace{\partial_a \gamma_1(\phi)}_{=0} X_{k-1}) \\ &= \partial_a \gamma_0(\phi) \sum_{k=1}^n (X_k - \gamma_0(\phi) - \gamma_1(\phi) X_{k-1}) \\ &= n \partial_a \gamma_0(\phi) (\bar{X}_n - \gamma_0(\phi) - \gamma_1(\phi) \bar{X}'_n)\end{aligned}\tag{3.2.15}$$

und

$$\begin{aligned}0 &\stackrel{!}{=} \sum_{k=1}^n (X_k - m(X_{k-1}; \phi)) \cdot \partial_b m(X_{k-1}; \phi) \\ &= \sum_{k=1}^n (X_k - \gamma_0(\phi) - \gamma_1(\phi) X_{k-1}) (\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi) X_{k-1}) \\ &= n \partial_b \gamma_0(\phi) (\bar{X}_n - \gamma_0(\phi) - \gamma_1(\phi) \bar{X}'_n) \\ &\quad + \partial_b \gamma_1(\phi) \left(\sum_{k=1}^n X_k X_{k-1} - n \gamma_0(\phi) \bar{X}'_n - \gamma_1(\phi) \sum_{k=1}^n X_{k-1}^2 \right).\end{aligned}\tag{3.2.16}$$

Nun ist wegen $\partial_a \gamma_0(\phi) = \frac{1}{b}(e^{b\Delta} - 1) > 0$ die rechte Seite von (3.2.15) offensichtlich genau dann gleich 0, wenn der Ausdruck in der Klammer gleich 0 ist, also wenn

$$0 = (\bar{X}_n - \gamma_0(\phi) - \gamma_1(\phi) \bar{X}'_n) = \bar{X}_n - \frac{a}{b}(e^{b\Delta} - 1) - e^{b\Delta} \bar{X}'_n,$$

das ist gleichwertig zu

$$a = \frac{\bar{X}_n - e^{b\Delta} \bar{X}'_n}{e^{b\Delta} - 1} b.$$

Setzen wir dies in (3.2.16) ein, so ergibt sich wegen $\partial_b \gamma_1(\phi) = \Delta e^{b\Delta} > 0$

$$\begin{aligned}
0 &= \sum_{k=1}^n X_k X_{k-1} - n \bar{X}'_n \gamma_0(\phi) - \gamma_1(\phi) \sum_{k=1}^n X_{k-1}^2 \\
&= \sum_{k=1}^n X_k X_{k-1} - n \bar{X}'_n (\bar{X}_n - \gamma_1(\phi) \bar{X}'_n) - \gamma_1(\phi) \sum_{k=1}^n X_{k-1}^2 \\
&= \sum_{k=1}^n X_k X_{k-1} - n \bar{X}_n \bar{X}'_n - \gamma_1(\phi) \left(\sum_{k=1}^n X_{k-1}^2 - n \bar{X}_n'^2 \right) \\
&= \Sigma_n^{(1)} - e^{b\Delta} \Sigma_n^{(2)},
\end{aligned}$$

was unter der Voraussetzung $0 < \Sigma_n^{(1)} < \Sigma_n^{(2)}$ offenbar äquivalent ist zu

$$b = \frac{1}{\Delta} \log \frac{\Sigma_n^{(1)}}{\Sigma_n^{(2)}} < 0.$$

Wir haben somit gezeigt: gilt $0 < \Sigma_n^{(1)} < \Sigma_n^{(2)}$, so hat das Gleichungssystem (3.2.13) eine eindeutige Lösung (\hat{a}_n, \hat{b}_n) in \mathbb{R}^2 , nämlich

$$\hat{b}_n := \frac{1}{\Delta} \log \left(\frac{\Sigma_n^{(1)}}{\Sigma_n^{(2)}} \right) < 0,$$

$$\hat{a}_n := \frac{\bar{X}_n - e^{\hat{b}_n \Delta} \bar{X}'_n}{e^{\hat{b}_n \Delta} - 1} \hat{b}_n = \frac{\bar{X}_n \Sigma_n^{(2)} - \bar{X}'_n \Sigma_n^{(1)}}{\Sigma_n^{(1)} - \Sigma_n^{(2)}} \cdot \frac{1}{\Delta} \log \frac{\Sigma_n^{(1)}}{\Sigma_n^{(2)}} \geq 0;$$

dabei gilt $(\hat{a}_n, \hat{b}_n) \in \Phi = \mathbb{R}_{>0} \times \mathbb{R}_{<0}$, falls zusätzlich $\hat{a}_n > 0$. Folgerichtig definieren wir

$$B_n := \{0 < \Sigma_n^{(1)} < \Sigma_n^{(2)}\} \quad (3.2.17)$$

und einen Schätzer $\hat{\phi}_n$ für ϕ durch

$$\hat{\phi}_n := \mathbf{1}_{B_n} \cdot \begin{pmatrix} \hat{a}_n \\ \hat{b}_n \end{pmatrix} = \mathbf{1}_{B_n} \cdot \frac{1}{\Delta} \log \left(\frac{\Sigma_n^{(1)}}{\Sigma_n^{(2)}} \right) \begin{pmatrix} \frac{\bar{X}_n \Sigma_n^{(2)} - \bar{X}'_n \Sigma_n^{(1)}}{\Sigma_n^{(1)} - \Sigma_n^{(2)}} \\ 1 \end{pmatrix} \quad (3.2.18)$$

Man beachte, daß wegen $\Phi = \mathbb{R}_{>0} \times \mathbb{R}_{<0}$ stets $\{\hat{\phi}_n \in \Phi\} \subseteq B_n$ gilt. Als Ergebnis der bisherigen Diskussion sei festgehalten:

Feststellung 3.2.1 *Sei $\hat{\phi}_n$ definiert wie in (3.2.18). Dann ist $\hat{\phi}_n$ auf dem Ereignis B_n die eindeutige Lösung des Gleichungssystems (3.2.13).*

Wir zeigen nun, daß $(\hat{\phi}_n)_n$ stark konsistent für den Parameter ϕ ist:

Lemma 3.2.2 Für alle $\theta = (\phi, \sigma^2) \in \Theta$ gilt

$$\hat{\phi}_n \xrightarrow{n \rightarrow \infty} \phi \quad P_\theta\text{-f.s.} \quad (3.2.19)$$

und

$$P_\theta \left(\liminf_{n \rightarrow \infty} \left\{ \hat{\phi}_n \in \Phi, \nabla_\phi Q_n(\hat{\phi}_n) = 0 \text{ und } \hat{\phi}_n \text{ ist lokale Minimalstelle von } \Phi \ni \phi \mapsto Q_n(\phi) \right\} \right) = 1. \quad (3.2.20)$$

Beweis Fixiere ein beliebiges $\theta = (\phi, \sigma^2) = (a, b, \sigma^2) \in \Theta$.

1. Zeige zuerst (3.2.19). Da der Prozeß $(X_n)_n$ unter P_{μ_θ} Momente jeder Ordnung hat, folgt mit dem Starken Gesetz der Großen Zahlen 2.1.3

$$\begin{aligned} \bar{X}_n &\xrightarrow{n \rightarrow \infty} E_{\mu_\theta}[X_0], & \bar{X}'_n &\xrightarrow{n \rightarrow \infty} E_{\mu_\theta}[X_0] & P_\theta\text{-f.s.}, \\ \frac{1}{n} \sum_{k=1}^n X_{k-1}^2 &\xrightarrow{n \rightarrow \infty} E_{\mu_\theta}[X_0^2] & P_\theta\text{-f.s.} \end{aligned}$$

sowie

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n X_k X_{k-1} &\xrightarrow{n \rightarrow \infty} E_{\mu_\theta}[X_0 \cdot X_1] = E_{\mu_\theta}[X_0 \cdot E_{\mu_\theta}[X_1 | X_0]] \\ &= E_{\mu_\theta}[X_0(\gamma_0(\phi) + \gamma_1(\phi)X_0)] \\ &= \gamma_0(\phi)E_{\mu_\theta}[X_0] + \gamma_1(\phi)E_{\mu_\theta}[X_0^2] & P_\theta\text{-f.s.} \end{aligned}$$

Daraus folgt (vgl. (1.1.13))

$$\frac{1}{n} \Sigma_n^{(2)} = \frac{1}{n} \sum_{k=1}^n X_{k-1}^2 - (\bar{X}'_n)^2 \xrightarrow{n \rightarrow \infty} E_{\mu_\theta}[X_0^2] - E_{\mu_\theta}[X_0]^2 = \text{Var}_{\mu_\theta}[X_0] = \frac{a\sigma^2}{2b^2}$$

und

$$\begin{aligned} \frac{1}{n} \Sigma_n^{(1)} &= \frac{1}{n} \sum_{k=1}^n X_k X_{k-1} - \bar{X}_n \bar{X}'_n \xrightarrow{n \rightarrow \infty} \gamma_0(\phi)E_{\mu_\theta}[X_0] + \gamma_1(\phi)E_{\mu_\theta}[X_0^2] - (E_{\mu_\theta}[X_0])^2 \\ &= \frac{a}{b}(e^{b\Delta} - 1) \left(-\frac{a}{b}\right) + e^{b\Delta} \frac{a(\sigma^2 + 2a)}{2b^2} - \left(-\frac{a}{b}\right)^2 \\ &= e^{b\Delta} \frac{a\sigma^2}{2b^2} & P_\theta\text{-f.s.} \end{aligned}$$

Insbesondere folgt auch

$$\frac{1}{n} (\Sigma_n^{(2)} - \Sigma_n^{(1)}) \xrightarrow{n \rightarrow \infty} \frac{a\sigma^2}{2b^2} (1 - e^{b\Delta}) > 0 \quad P_\theta\text{-f.s.}$$

Mit $B_n = \left\{0 < \Sigma_n^{(1)} < \Sigma_n^{(2)}\right\}$ aus (3.2.17) gilt folglich

$$P_\theta \left(\liminf_{n \rightarrow \infty} B_n \right) = 1 \quad (3.2.21)$$

sowie

$$\hat{b}_n = \mathbf{1}_{B_n} \cdot \frac{1}{\Delta} \log \left(\frac{\frac{1}{n} \Sigma_n^{(1)}}{\frac{1}{n} \Sigma_n^{(2)}} \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\Delta} \log \left(\frac{e^{b\Delta} \cdot a\sigma^2 / 2b^2}{a\sigma^2 / 2b^2} \right) = b \quad P_\theta\text{-f.s.}$$

Hieraus ergibt sich dann sofort

$$\hat{a}_n = \frac{\bar{X}_n - e^{\hat{b}_n \Delta} \bar{X}'_n}{e^{\hat{b}_n \Delta} - 1} \hat{b}_n \xrightarrow{n \rightarrow \infty} \frac{-\frac{a}{b} - e^{b\Delta} \left(-\frac{a}{b}\right)}{e^{b\Delta} - 1} b = a \quad P_\theta\text{-f.s.},$$

das ist (3.2.19).

2. Wegen Feststellung 3.2.1 und (3.2.21) gilt insbesondere auch

$$P_\theta \left(\liminf_{n \rightarrow \infty} \{ \hat{\phi}_n \in \Phi, \nabla_\phi Q_n(\hat{\phi}_n) = 0 \} \right) = 1. \quad (3.2.22)$$

Definiere nun

$$f(x, y; \phi) := (y - m(x; \phi))^2 = (y - \gamma_0(\phi) - \gamma_1(\phi)x)^2,$$

so daß also

$$Q_n(\phi) = \sum_{k=1}^n f(X_{k-1}, X_k; \phi).$$

Die zweiten partiellen Ableitungen von f nach a und b sind gegeben durch

$$\partial_{aa}^2 f(x, y; \phi) = 2 (\partial_a \gamma_0(\phi))^2,$$

$$\partial_{ab}^2 f(x, y; \phi) = 2 [\partial_a \gamma_0(\phi) (\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi)x) - \partial_a^2 \gamma_0(\phi) (y - m(x; \phi))],$$

$$\partial_{bb}^2 f(x, y; \phi) = 2 [(\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi)x)^2 - (\partial_{bb}^2 \gamma_0(\phi) + \partial_b^2 \gamma_1(\phi)x) (y - m(x; \phi))].$$

Hieraus ersieht man, daß $\partial_{ij}^2 f$ für $i, j = a, b$ Polynome in (x, y) sind, deren Koeffizienten wegen $\gamma_0, \gamma_1 \in C^\infty(\Phi)$ stetig von ϕ abhängen. Da der Prozeß $(X_n)_n$ unter P_{μ_θ} Momente beliebig hoher Ordnung hat, sind folglich alle zweiten partiellen Ableitungen $\partial_{ij}^2 f$ lokal dominiert integrierbar, vgl. die Bemerkungen nach Definition 2.3.2. Aus Lemma 2.3.3 folgt daher

$$\frac{1}{n} \partial_{ij}^2 Q_n(\hat{\phi}_n) = \frac{1}{n} \sum_{k=1}^n \partial_{ij}^2 f(X_{k-1}, X_k; \hat{\phi}_n) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta (\partial_{ij}^2 g(\cdot, \cdot; \phi)) \quad P_\theta\text{-f.s.},$$

für $i, j = a, b$, also auch

$$\frac{1}{n} \nabla \nabla_{\phi}^T Q_n(\hat{\phi}_n) \xrightarrow{n \rightarrow \infty} [\tilde{\mu}_{\theta} (\partial_{ij}^2 f(\cdot, \cdot; \phi))]_{i,j=a,b} =: V(\phi), \quad (3.2.23)$$

wobei wir mit $\nabla \nabla_{\phi}^T Q_n(\phi)$ die Hesse-Matrix von Q_n an der Stelle ϕ bezeichnet haben. Für die Matrix $V(\phi)$ berechnet man

$$V_{11}(\phi) = 2(\partial_a \gamma_0(\phi))^2,$$

$$\begin{aligned} V_{21}(\phi) = V_{12}(\phi) &= 2\partial_a \gamma_0(\phi) (\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi) E_{\mu_{\theta}}[X_0]) - 2\partial_a^2 \gamma_0(\phi) \underbrace{E_{\mu_{\theta}}[X_1 - m(X_0; \phi)]}_{=0 \text{ nach Def. von } m} \\ &= 2\partial_a \gamma_0(\phi) (\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi) E_{\mu_{\theta}}[X_0]), \end{aligned}$$

$$\begin{aligned} V_{22}(\phi) &= 2E_{\mu_{\theta}} [(\partial_b \gamma_0(\phi) + \partial_b \gamma_1(\phi) X_0)^2] - 2 \underbrace{E_{\mu_{\theta}} [(\partial_b^2 \gamma_0(\phi) + \partial_b^2 \gamma_1(\phi) X_0) (X_1 - m(X_0; \phi))]}_{=0 \text{ nach Def. von } m} \\ &= 2(\partial_b \gamma_0(\phi))^2 + 4\partial_b \gamma_0(\phi) \partial_b \gamma_1(\phi) E_{\mu_{\theta}}[X_0] + 2(\partial_b \gamma_1(\phi))^2 E_{\mu_{\theta}}[X_0^2]. \end{aligned}$$

Folglich ist $V_{11}(\phi) > 0$ sowie

$$\begin{aligned} \det V(\phi) &= 4(\partial_a \gamma_0(\phi))^2 (\partial_b \gamma_1(\phi))^2 (E_{\mu_{\theta}}[X_0^2] - E_{\mu_{\theta}}[X_0]^2) \\ &= 4(\partial_a \gamma_0(\phi))^2 (\partial_b \gamma_1(\phi))^2 \cdot \text{Var}_{\mu_{\theta}}[X_0] > 0; \end{aligned}$$

die Matrix $V(\phi)$ ist somit positiv definit. Wegen (3.2.23) ist daher auch

$$P_{\theta} \left(\liminf_{n \rightarrow \infty} \left\{ \nabla \nabla_{\phi}^T Q_n(\hat{\phi}_n) \text{ ist positiv definit} \right\} \right) = 1.$$

Setzen wir nun

$$C_n := \left\{ \hat{\phi}_n \in \Phi, \nabla_{\phi} Q_n(\hat{\phi}_n) = 0 \right\} \cap \left\{ \nabla \nabla_{\phi}^T Q_n(\hat{\phi}_n) \text{ ist positiv definit} \right\}, \quad (3.2.24)$$

so ist $\hat{\phi}_n$ auf dem Ereignis C_n eine lokale Minimalstelle der Funktion $\Phi \ni \phi \mapsto Q_n(\phi)$, und nach dem gerade Bewiesenen sowie (3.2.22) ist

$$P_{\theta} \left(\liminf_{n \rightarrow \infty} C_n \right) = 1,$$

das ist (3.2.20)

Bemerkung Das eben bewiesene Resultat liefert nicht, daß $\hat{\phi}_n$ asymptotisch eine *globale* Minimalstelle von $\Phi \ni \phi \mapsto Q_n(\phi)$, also $Q_n(\phi) \geq Q_n(\hat{\phi}_n)$ für alle $\phi \in \Phi$ ist. Für das Ereignis C_n aus (3.2.24) gilt jedoch $C_n \subseteq B_n$, und auf B_n ist $\hat{\phi}_n$ nach Feststellung 3.2.1 der einzige kritische Punkt von $\phi \mapsto Q_n(\phi)$. Damit ist $\hat{\phi}_n$ zumindest in dem Sinne „optimal“, daß es asymptotisch außer $\hat{\phi}_n$ keine weiteren Minimalstellen von $\Phi \ni \phi \mapsto Q_n(\phi)$ gibt. ♦

Schätzung des Diffusionskoeffizienten

Wir wenden uns nun der Schätzung des Diffusionskoeffizienten σ^2 zu. Dafür benötigen wir die bedingte Varianz. Nach den Überlegungen aus Abschnitt 3.1 ist diese gegeben durch

$$\text{Var}_\theta[X_n | \mathcal{F}_{n-1}] = v(X_{n-1}; \theta) = \sigma^2(\eta_0(\phi) + \eta_1(\phi)X_{n-1}) \quad (3.2.25)$$

mit v , η_0 und η_1 aus (3.2.5) und (3.2.6), wobei wir wieder η_0 und η_1 als Funktionen von $\phi = (a, b)$ auffassen.

Ist nun $(\bar{\phi}_n)_n$ eine beliebige Schätzfolge für $\phi \in \Phi$, so bieten sich verschiedene Möglichkeiten zur Konstruktion eines Schätzers für σ^2 an. Für festes $\phi \in \Phi$ können wir etwa $(X_k - m(X_{k-1}, \phi))^2$ mit seinem Kompensator unter P_θ gleichsetzen,

$$(X_k - m(X_{k-1}, \phi))^2 \stackrel{!}{=} E_\theta [(X_k - m(X_{k-1}, \phi))^2 | \mathcal{F}_{k-1}] = \sigma^2(\eta_0(\phi) + \eta_1(\phi)X_{k-1}),$$

und nach σ^2 auflösen. Das führt zu

$$\sigma^2 = \frac{(X_k - m(X_{k-1}, \phi))^2}{\eta_0(\phi) + \eta_1(\phi)X_{k-1}}.$$

Daher bietet sich

$$\boxed{\bar{\rho}_n^2 := \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{1}{n} \sum_{k=1}^n \frac{(X_k - m(X_{k-1}, \bar{\phi}_n))^2}{\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1}}} \quad (3.2.26)$$

als Schätzer für σ^2 an. Dieser Schätzer ist auf dem Ereignis $\{\bar{\phi}_n \in \Phi\}$ die eindeutige Lösung in σ^2 der Gleichung

$$\frac{1}{n} \sum_{k=1}^n \frac{(X_k - m(X_{k-1}, \bar{\phi}_n))^2 - \sigma^2 (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})}{\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1}} = 0.$$

Man beachte, daß wegen $\eta_0(\phi) > 0$, $\eta_1(\phi) > 0$ für alle $\phi \in \Phi$ und $X_k \geq 0$ für alle $k \in \mathbb{N}_0$ stets $\eta_0(\phi) + \eta_1(\phi)X_{k-1} > 0$ und somit die rechte Seite von (3.2.26) wohldefiniert ist.

Ein anderer Ansatz besteht darin, Gleichung (3.2.25) für festes $\phi \in \Phi$ als lineare Regressionsgleichung im Parameter σ^2 aufzufassen,

$$(X_k - m(X_{k-1}, \phi))^2 = \sigma^2(\eta_0(\phi) + \eta_1(\phi)X_{k-1}) + \epsilon_k, \quad (3.2.27)$$

und die Methode der kleinsten Quadrate zu verwenden. Dann müssen wir also für festes $\phi \in \Phi$ die Funktion

$$\sigma^2 \mapsto \sum_{k=1}^n [(X_k - m(X_{k-1}, \phi))^2 - \sigma^2(\eta_0(\phi) + \eta_1(\phi)X_{k-1})]^2$$

minimieren. Differentiation nach σ^2 ergibt

$$0 \stackrel{!}{=} -2 \sum_{k=1}^n [(X_k - m(X_{k-1}; \phi))^2 - \sigma^2(\eta_0(\phi) + \eta_1(\phi)X_{k-1})] \cdot (\eta_0(\phi) + \eta_1(\phi)X_{k-1}).$$

Diese Gleichung hat die eindeutige Lösung

$$\sigma_n^2(\phi) := \frac{\sum_{k=1}^n (\eta_0(\phi) + \eta_1(\phi)X_{k-1})(X_k - m(X_{k-1}; \phi))^2}{\sum_{k=1}^n (\eta_0(\phi) + \eta_1(\phi)X_{k-1})^2}.$$

Zu jeder Schätzfolge $(\bar{\phi}_n)_n$ für ϕ erhalten wir dann eine Schätzfolge $(\bar{\sigma}_n^2)_n$ für σ^2 durch die Festsetzung

$$\boxed{\bar{\sigma}_n^2 := \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \sigma_n^2(\bar{\phi}_n) = \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{\sum_{k=1}^n (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})(X_k - m(X_{k-1}; \bar{\phi}_n))^2}{\sum_{k=1}^n (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})^2}} \quad (3.2.28)$$

Wir halten fest:

Feststellung 3.2.3 *Auf dem Ereignis $\{\bar{\phi}_n \in \Phi\}$ ist $\bar{\sigma}_n^2$ die eindeutige Lösung in σ^2 der Gleichung*

$$\sum_{k=1}^n [(X_k - m(X_{k-1}; \bar{\phi}_n))^2 - \sigma^2(\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})] \cdot (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1}) = 0$$

und $\bar{\rho}_n^2$ die eindeutige Lösung in σ^2 der Gleichung

$$\frac{1}{n} \sum_{k=1}^n \frac{(X_k - m(X_{k-1}; \bar{\phi}_n))^2 - \sigma^2(\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})}{\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1}} = 0.$$

Ist nun die Schätzfolge $(\bar{\phi}_n)_n$ (stark) konsistent, so überträgt sich diese Eigenschaft auf $(\bar{\rho}_n^2)_n$ und $(\bar{\sigma}_n^2)_n$:

Lemma 3.2.4 *Sei $(\bar{\phi}_n)_n$ eine Schätzfolge für den Parameter $\phi \in \Phi$. Dann gilt: Ist $(\bar{\phi}_n)_n$ konsistent (stark konsistent), d.h.*

$$\forall \theta = (\phi, \sigma^2) \in \Theta : \bar{\phi}_n \xrightarrow{n \rightarrow \infty} \phi \quad P_\theta\text{-stochastisch (} P_\theta\text{-f.s.)},$$

so sind die durch (3.2.26) und (3.2.28) definierten Schätzfolgen $(\bar{\rho}_n^2)_n$ und $(\bar{\sigma}_n^2)_n$ konsistent (stark konsistent) für den Parameter σ^2 , d.h.

$$\forall \theta = (\phi, \sigma^2) \in \Theta : \bar{\sigma}_n^2 \xrightarrow{n \rightarrow \infty} \sigma^2 \quad \text{und} \quad \bar{\rho}_n^2 \xrightarrow{n \rightarrow \infty} \sigma^2 \quad P_\theta\text{-stochastisch (} P_\theta\text{-f.s.)}.$$

Beweis Fixiere $\theta = (\phi, \sigma^2) \in \Theta$.

1. Betrachte zunächst $(\bar{\sigma}_n^2)_n$ und setze starke Konsistenz voraus. Definiere

$$\begin{aligned} f(x, y; \phi) &:= (\eta_0(\phi) + \eta_1(\phi)x)(y - m(x; \phi))^2 \\ &= (\eta_0(\phi) + \eta_1(\phi)x)(y - \gamma_0(\phi) - \gamma_1(\phi)x)^2 \geq 0 \end{aligned}$$

und

$$g(x, y; \phi) := (\eta_0(\phi) + \eta_1(\phi)x)^2 > 0, \quad x, y \in \mathbb{R}_{\geq 0}, \phi \in \Phi.$$

Dann sind für festes $\phi \in \Phi$ offenbar $(x, y) \mapsto f(x, y; \phi)$ und $(x, y) \mapsto g(x, y; \phi)$ Polynome in (x, y) , und wegen der Stetigkeit von η_0, η_1 und m in ϕ hängen die Koeffizienten stetig von ϕ ab. Damit ist f lokal dominiert integrierbar, vgl. die Bemerkungen nach Definition 2.3.2, und aus Lemma 2.3.3 folgt

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n f(X_{k-1}, X_k; \bar{\phi}_n) &\xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(f(\cdot, \cdot; \phi)) = E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0)(X_1 - m(X_0; \phi))^2] \\ &= E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0) \cdot v(X_0; \phi)] \\ &= \sigma^2 E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0)^2] \quad P_\theta\text{-f.s.} \end{aligned}$$

sowie

$$\frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\phi}_n) \xrightarrow{n \rightarrow \infty} \tilde{\mu}_\theta(g(\cdot, \cdot; \phi)) = E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0)^2] \quad P_\theta\text{-f.s.}$$

Da wegen der vorausgesetzten starken Konsistenz zudem $\mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \xrightarrow{n \rightarrow \infty} 1$ P_θ -fast sicher, erhalten wir

$$\begin{aligned} \bar{\sigma}_n^2 &= \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{\frac{1}{n} \sum_{k=1}^n (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})(X_k - m(X_{k-1}; \bar{\phi}_n))^2}{\frac{1}{n} \sum_{k=1}^n (\eta_0(\bar{\phi}_n) + \eta_1(\bar{\phi}_n)X_{k-1})^2} \\ &= \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{\frac{1}{n} \sum_{k=1}^n f(X_{k-1}, X_k; \bar{\phi}_n)}{\frac{1}{n} \sum_{k=1}^n g(X_{k-1}, X_k; \bar{\phi}_n)} \xrightarrow{n \rightarrow \infty} \frac{\sigma^2 E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0)^2]}{E_{\mu_\theta} [(\eta_0(\phi) + \eta_1(\phi)X_0)^2]} = \sigma^2 \quad P_\theta\text{-f.s.} \end{aligned}$$

Ist $(\bar{\sigma}_n^2)_n$ nur konsistent, so schließt man analog unter Verwendung von Lemma 2.3.3.

2. Betrachte nun $(\bar{\rho}_n^2)_n$. Dieser Schätzer hat die Darstellung

$$\bar{\rho}_n^2 = \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{1}{n} \sum_{k=1}^n h(X_{k-1}, X_k; \bar{\phi}_n)$$

mit

$$h(x, y; \phi) := \frac{(y - m(x; \phi))^2}{\eta_0(\phi) + \eta_1(\phi)x} \geq 0, \quad x, y \in \mathbb{R}_{\geq 0}, \phi \in \Phi.$$

Offenbar gilt wegen $\eta_0(\phi), \eta_1(\phi) > 0$ die Abschätzung

$$h(x, y; \phi) \leq \frac{(y - m(x; \phi))^2}{\eta_0(\phi)}$$

und hier ist die Funktion auf der rechten Seite wieder ein Polynom in (x, y) mit in ϕ stetigen Koeffizienten, also lokal dominiert integrierbar. Wiederum mit Lemma 2.3.3 folgt

$$\begin{aligned} \bar{\rho}_n^2 &= \mathbf{1}_{\{\bar{\phi}_n \in \Phi\}} \cdot \frac{1}{n} \sum_{k=1}^n h(X_{k-1}, X_k; \bar{\phi}_n) \xrightarrow{n \rightarrow \infty} E_{\mu_\theta} [h(X_0, X_1; \phi)] \\ &= E_{\mu_\theta} \left[\frac{(X_1 - m(X_0; \phi))^2}{\eta_0(\phi) + \eta_1(\phi)X_0} \right] \\ &= E_{\mu_\theta} \left[\frac{v(X_0; \phi)}{\eta_0(\phi) + \eta_1(\phi)X_0} \right] = \sigma^2 \end{aligned}$$

mit P_θ -fast sicher bzw. stochastischer Konvergenz, je nachdem ob $(\bar{\phi}_n)_n$ stark konsistent bzw. nur konsistent ist. Damit ist die Behauptung gezeigt. ■

Definition und Asymptotik des Conditional Least Squares-Schätzers

Mit diesen Ergebnissen können jetzt zwei explizite Schätzfolgen für den Parameter $\theta = (\phi, \sigma^2) \in \Theta$ angegeben werden. Sei $\hat{\phi}_n$ wie in (3.2.18) und seien $\hat{\rho}_n^2, \hat{\sigma}_n^2$ die gemäß (3.2.26), (3.2.28) zugehörigen Schätzer für σ^2 . Wir definieren

$$\hat{\theta}_n := \mathbf{1}_{\{\hat{\phi}_n \in \Phi\}} \cdot (\hat{\phi}_n, \hat{\sigma}_n^2), \quad \tilde{\theta}_n := \mathbf{1}_{\{\hat{\phi}_n \in \Phi\}} \cdot (\hat{\phi}_n, \hat{\rho}_n^2) \quad (3.2.29)$$

Wir fassen nun m, η_0 und η_1 wieder als in σ^2 konstante Funktionen von $\theta = (\phi, \sigma^2)$ auf. Feststellung 3.2.1 und Feststellung 3.2.3 zeigen dann zusammen:

Feststellung 3.2.5 *Auf dem Ereignis $\{\hat{\phi}_n \in \Phi\} \subseteq B_n$ ist $\hat{\theta}_n$ die eindeutige Lösung der quadratischen Schätzgleichung*

$$0 = G_n(\theta) := \sum_{k=1}^n (X_k - m(X_{k-1}; \theta)) w_m(X_{k-1}; \theta) + ((X_k - m(X_{k-1}; \theta))^2 - v(X_{k-1}; \theta)) w_v(X_{k-1}; \theta), \quad (3.2.30)$$

wobei

$$w_m(x; \theta) := \begin{pmatrix} \partial_a m(x; \theta) \\ \partial_b m(x; \theta) \\ 0 \end{pmatrix} = \begin{pmatrix} \partial_a \gamma_0(\theta) \\ \partial_b \gamma_0(\theta) + \partial_b \gamma_1(\theta)x \\ 0 \end{pmatrix}, \quad (3.2.31)$$

$$w_v(x; \theta) := \begin{pmatrix} 0 \\ 0 \\ \eta_0(\theta) + \eta_1(\theta)x \end{pmatrix}. \quad (3.2.32)$$

Ebenso ist $\tilde{\theta}_n$ auf $\{\hat{\phi}_n \in \Phi\}$ die eindeutige Lösung von

$$0 = \tilde{G}_n(\theta) := \sum_{k=1}^n (X_k - m(X_{k-1}; \theta)) \tilde{w}_m(X_{k-1}; \theta) + ((X_k - m(X_{k-1}; \theta))^2 - v(X_{k-1}; \theta)) \tilde{w}_v(X_{k-1}; \theta) \quad (3.2.33)$$

mit

$$\tilde{w}_m = w_m, \quad \tilde{w}_v(x; \theta) := \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\eta_0(\theta) + \eta_1(\theta)x} \end{pmatrix}. \quad (3.2.34)$$

Beweis Wir zeigen die Behauptung für $(\hat{\theta}_n)_n$. Nach Feststellung 3.2.1 und 3.2.3 ist klar, daß $\hat{\theta}_n$ auf $\{\hat{\phi}_n \in \Phi\}$ die Gleichung (3.2.30) löst. Zu zeigen bleibt die Eindeutigkeit. Sei $\omega \in \{\hat{\phi}_n \in \Phi\} \subseteq B_n$ beliebig und $\theta = (\phi, \sigma^2) \in \Theta$ mit $G_n(\omega; \theta) = 0$. Aufgrund der speziellen Gestalt von w_m und w_v erhalten wir dann zunächst

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \sum_{k=1}^n (X_k(\omega) - m(X_{k-1}(\omega); \phi)) \begin{pmatrix} \partial_a m(X_{k-1}(\omega); \phi) \\ \partial_b m(X_{k-1}(\omega); \phi) \end{pmatrix},$$

d.h. $\theta = (\phi, \sigma^2)$ löst das Gleichungssystem (3.2.13). Wegen $\omega \in B_n$ und der Eindeutigkeit in Feststellung 3.2.1 folgt, daß $\phi = \hat{\phi}_n(\omega)$. Aus der Gestalt von w_v erhält man dann

$$0 = \sum_{k=1}^n \left[\left(X_k(\omega) - m(X_{k-1}(\omega); \hat{\phi}_n(\omega)) \right)^2 - \sigma^2 \left(\eta_0(\hat{\phi}_n(\omega)) + \eta_1(\hat{\phi}_n(\omega)) X_{k-1}(\omega) \right) \right] \times \\ \times \left(\eta_0(\hat{\phi}_n(\omega)) + \eta_1(\hat{\phi}_n(\omega)) X_{k-1}(\omega) \right),$$

d.h. σ^2 löst die erste Gleichung in Feststellung 3.2.3. Wegen $\hat{\phi}_n(\omega) \in \Phi$ und der Eindeutigkeitsaussage in Feststellung 3.2.3 folgt dann $\sigma^2 = \hat{\sigma}_n^2(\omega)$, also insgesamt $\theta = (\hat{\phi}_n(\omega), \hat{\sigma}_n^2(\omega)) = \hat{\theta}_n(\omega)$. Das zeigt die behauptete Eindeutigkeit. \blacksquare

Wegen Feststellung 3.2.1 und $\partial_{\sigma^2} m(x; \theta) \equiv 0$ ist außerdem klar, daß $\hat{\theta}_n$ und $\tilde{\theta}_n$ auf dem Ereignis B_n die Least-Squares-Gleichung (3.2.12) lösen, d.h. auf B_n gilt $\nabla_{\theta} Q_n(\hat{\theta}_n) = 0 = \nabla_{\theta} Q_n(\tilde{\theta}_n)$. Wir fassen die bisherigen Erkenntnisse über die beiden Schätzfolgen $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ in den folgenden Theoremen zusammen:

Theorem 3.2.6 Die Schätzfolge $(\hat{\theta}_n)_n$ ist die eindeutige stark konsistente Lösung der quadratischen Schätzgleichung (3.2.30). Außerdem ist $(\hat{\theta}_n)_n$ eine stark konsistente Lösung der Least-Squares-Gleichung (3.2.12) mit

$$\forall \theta \in \Theta : P_{\theta} \left(\liminf_{n \rightarrow \infty} \left\{ \hat{\theta}_n \text{ ist lokale Minimalstelle von } \Theta \ni \theta \mapsto Q_n(\theta) \right\} \right) = 1. \quad (3.2.35)$$

Beweis Die starke Konsistenz von $(\hat{\theta}_n)_n$ ergibt sich aus den Lemmata 3.2.2 und 3.2.4. Insbesondere gilt also auch

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \{\hat{\theta}_n \in \Theta\} \right) = 1. \quad (3.2.36)$$

Wegen $\{\hat{\phi}_n \in \Phi\} \subseteq \{\hat{\theta}_n \in \Theta\}$ und Feststellung 3.2.5 gilt $G_n(\hat{\theta}_n) = 0$ auf dem Ereignis $\{\hat{\theta}_n \in \Theta\}$. Zusammen mit (3.2.36) zeigt dies, daß $(\hat{\theta}_n)_n$ eine stark konsistente Lösung der Schätzgleichung (3.2.30) im Sinne von Definition 2.3.1 ist. Die Eindeutigkeit ergibt sich sofort aus der Eindeutigkeitsaussage in Feststellung 3.2.5.

Oben wurde schon bemerkt, daß $\nabla_\theta Q_n(\hat{\theta}_n) = 0$ auf dem Ereignis B_n gilt. Da nach (3.2.21) $P_\theta (\liminf_{n \rightarrow \infty} B_n) = 1$ für alle $\theta \in \Theta$, ist $(\hat{\theta}_n)_n$ also auch eine stark konsistente Lösung von (3.2.12). Schließlich ist die Funktion Q_n konstant in $\theta_3 = \sigma^2$, d.h. $Q_n(\theta) = Q_n(\phi)$ für $\theta = (\phi, \sigma^2)$. Ist $\hat{\phi}_n$ lokale Minimalstelle von $\Phi \ni \phi \mapsto Q_n(\phi)$, so ist also $\hat{\theta}_n = (\hat{\phi}_n, \hat{\sigma}_n^2)$ lokale Minimalstelle von $\Theta \ni \theta \mapsto Q_n(\theta)$. Wegen (3.2.20) in Lemma 3.2.2 folgt daher (3.2.35). ■

Völlig analog beweist man das folgende Resultat über die Schätzfolge $(\tilde{\theta}_n)_n$:

Theorem 3.2.7 *Die Schätzfolge $(\tilde{\theta}_n)_n$ ist die eindeutige stark konsistente Lösung der quadratischen Schätzgleichung (3.2.33) sowie eine stark konsistente Lösung der Least-Squares-Gleichung (3.2.12), und es gilt*

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \left\{ \tilde{\theta}_n \text{ ist lokale Minimalstelle von } \Theta \ni \theta \mapsto Q_n(\theta) \right\} \right) = 1. \quad (3.2.37)$$

Das vorstehende Resultat rechtfertigt es, die Folgen $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ als *Conditional-Least-Squares*-Schätzfolgen, im folgenden kurz: **CLS**-Schätzfolgen für den Parameter $\theta \in \Theta$ zu bezeichnen. Für die asymptotische Verteilung der Schätzfehler kommt es jedoch ausschließlich darauf an, daß $\hat{\theta}_n$ und $\tilde{\theta}_n$ die quadratischen Martingal-Schätzgleichungen (3.2.30) bzw. (3.2.33) lösen. Diesbezüglich gilt folgendes Resultat:

Theorem 3.2.8 *Die Schätzfolgen $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ sind asymptotisch normal mit*

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, V(\theta)^{-1} W(\theta) V(\theta)^{-T} \right) \quad (3.2.38)$$

und

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left(0, \tilde{V}(\theta)^{-1} \tilde{W}(\theta) \tilde{V}(\theta)^{-T} \right). \quad (3.2.39)$$

Dabei sind

$$V(\theta) = -\mu_\theta \left(w_m(\cdot; \theta) [\nabla_\theta m(\cdot; \theta)]^T + w_v(\cdot; \theta) [\nabla_\theta v(\cdot; \theta)]^T \right), \quad (3.2.40)$$

$$\tilde{V}(\theta) = -\mu_\theta \left(w_m(\cdot; \theta) [\nabla_\theta m(\cdot; \theta)]^T + \tilde{w}_v(\cdot; \theta) [\nabla_\theta v(\cdot; \theta)]^T \right) \quad (3.2.41)$$

sowie

$$W(\theta) = \mu_\theta \left(v(\cdot; \theta) w_m(\cdot; \theta) w_m(\cdot; \theta)^T + (\mu_4(\cdot; \theta) - v(\cdot; \theta)^2) w_v(\cdot; \theta) w_v(\cdot; \theta)^T \right. \\ \left. + \mu_3(\cdot; \theta) (w_m(\cdot; \theta) w_v(\cdot; \theta)^T + w_v(\cdot; \theta) w_m(\cdot; \theta)^T) \right), \quad (3.2.42)$$

$$\tilde{W}(\theta) = \mu_\theta \left(v(\cdot; \theta) w_m(\cdot; \theta) w_m(\cdot; \theta)^T + (\mu_4(\cdot; \theta) - v(\cdot; \theta)^2) \tilde{w}_v(\cdot; \theta) \tilde{w}_v(\cdot; \theta)^T \right. \\ \left. + \mu_3(\cdot; \theta) (w_m(\cdot; \theta) \tilde{w}_v(\cdot; \theta)^T + \tilde{w}_v(\cdot; \theta) w_m(\cdot; \theta)^T) \right). \quad (3.2.43)$$

Beweis Wir weisen die Voraussetzungen von Theorem 2.3.5 nach.

1. Betrachte zunächst $(\hat{\theta}_n)_n$. Mit w_m und w_v aus (3.2.31) und (3.2.32) definiere

$$g(x, y; \theta) := w_m(x; \theta)(y - m(x; \theta)) + w_v(x; \theta) \left((y - m(x; \theta))^2 - v(x; \theta) \right),$$

so daß sich die Schätzgleichung (3.2.30) in der Form

$$\sum_{k=1}^n g(X_k, X_{k-1}; \theta) = 0$$

schreibt. Offenbar ist g komponentenweise ein Polynom vom Grad ≤ 3 in den Variablen x und y mit von θ abhängigen C^∞ -Koeffizienten; gleiches gilt somit auch für sämtliche partiellen Ableitungen $\partial_j g_i$, $i, j = 1, \dots, 3$. Da der **CIR**-Prozeß unter P_{μ_θ} Momente beliebig hoher Ordnung hat, sind diese also lokal dominiert integrierbar bezüglich $(\tilde{\mu}_\theta)_{\theta \in \Theta}$, und für $i = 1, \dots, 3$ und $\theta \in \Theta$ ist $g_i(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$. Insbesondere erfüllt g die Bedingung **C1** aus Kapitel 2. Die Behauptung folgt somit aus Theorem 2.3.5, falls g auch Bedingung **I** erfüllt, also die Matrix $V(\theta) = \tilde{\mu}_\theta(D_\theta g(\theta))$ für jedes $\theta \in \Theta$ invertierbar ist. In Abschnitt 2.5, Gleichung (2.5.7) wurde bereits nachgerechnet, daß $V(\theta)$ für quadratische Schätzfunktionen gerade durch den Ausdruck (3.2.40) gegeben ist. Die Invertierbarkeit von $V(\theta)$ ergibt sich aus der Bemerkung im Anschluß an diesen Beweis. Eine Rechnung wie in (2.5.9) zeigt, daß die Matrix $W(\theta) = \tilde{\mu}_\theta(g(\theta)g(\theta)^T)$ für quadratische Schätzfunktionen gerade die in (3.2.42) angegebene Gestalt hat. Damit folgt die Behauptung aus Theorem 2.3.5. Im Anschluß an diesen Beweis werden explizite Formeln für die Einträge von $V(\theta)$ und $W(\theta)$ angegeben.

2. Betrachte nun $(\tilde{\theta}_n)_n$. Die zugehörige Schätzgleichung (3.2.33) läßt sich schreiben als

$$\sum_{k=1}^n \tilde{g}(X_{k-1}, X_k; \theta) = 0$$

mit

$$\tilde{g}(x, y; \theta) := w_m(x; \theta)(y - m(x; \theta)) + \tilde{w}_v(x; \theta) \left((y - m(x; \theta))^2 - v(x; \theta) \right),$$

wobei $\tilde{w}_v(x; \theta) = 1/(\eta_0(\theta) + \eta_1(\theta)x)$ nach (3.2.34). Für $i = 1, 2$ ist $\tilde{g}_i(\cdot, \cdot; \theta) = g_i(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ für alle $\theta \in \Theta$ und $\partial_j \tilde{g}_i = \partial_j g_i$ lokal dominiert integrierbar für $j = 1, \dots, 3$ nach dem ersten Beweisschritt; zu betrachten bleibt nur die dritte Komponente

$$\tilde{g}_3(x, y; \theta) = \frac{1}{\eta_0(\theta) + \eta_1(\theta)x} \left((y - m(x; \theta))^2 - v(x; \theta) \right).$$

Offenbar ist für $j = 1, 2, 3$

$$\partial_j \tilde{g}_3(x, y; \theta) = \frac{\tilde{q}_j(x, y; \theta)}{(\eta_0(\theta) + \eta_1(\theta)x)^2},$$

wobei $\tilde{q}_j : E^2 \times \Theta \rightarrow \mathbb{R}$ ein Polynom in (x, y) mit stetig von θ abhängigen Koeffizienten ist. Da wegen $x \geq 0$ und $\eta_0(\theta), \eta_1(\theta) > 0$ stets $(\eta_0(\theta) + \eta_1(\theta)x)^2 \geq (\eta_0(\theta))^2 > 0$ abgeschätzt werden kann, ergibt sich

$$|\partial_j \tilde{g}_3(x, y; \theta)| \leq \frac{|\tilde{q}_j(x, y; \theta)|}{(\eta_0(\theta))^2},$$

also ist $\partial_j \tilde{g}_3$ lokal dominiert integrierbar. Außerdem ist

$$\begin{aligned} \tilde{\mu}_\theta (|\tilde{g}_3(\cdot, \cdot; \theta)|^2) &= E_{\mu_\theta} \left[\frac{\left((X_1 - m(X_0; \theta))^2 - v(X_0; \theta) \right)^2}{(\eta_0(\theta) + \eta_1(\theta)X_0)^2} \right] \\ &\leq \frac{1}{(\eta_0(\theta))^2} E_{\mu_\theta} \left[\left((X_1 - m(X_0; \theta))^2 - v(X_0; \theta) \right)^2 \right] < \infty, \end{aligned}$$

das zeigt $\tilde{g}_3(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$. Nach (2.5.7) und einer Rechnung wie in (2.5.9) sind wiederum $\tilde{V}(\theta) = \tilde{\mu}_\theta (D_\theta \tilde{g}(\theta))$ und $\tilde{W}(\theta) = \tilde{\mu}_\theta (\tilde{g}(\theta)\tilde{g}(\theta)^T)$ durch (3.2.41) und (3.2.43) gegeben, und $\tilde{V}(\theta)$ ist invertierbar, siehe die anschließende Bemerkung. Die Behauptung folgt damit wiederum aus Theorem 2.3.5. ■

Bemerkung Ein Vergleich der Kovarianzmatrizes in (3.2.38) und (3.2.39) scheint nur auf numerische Weise möglich zu sein. Für die Schätzfolge $(\hat{\theta}_n)_n$ können die Matrizes $V(\theta)$ und $W(\theta)$ explizit durch $\theta = (a, b, \sigma^2)$ ausgedrückt werden. Es ist $w_m(x; \theta) = \nabla_\theta m(x; \theta)$, also

$$V(\theta) = -\mu_\theta \left(\nabla_\theta m(\cdot; \theta) [\nabla_\theta m(\cdot; \theta)]^T + w_v(\cdot; \theta) [\nabla_\theta v(\cdot; \theta)]^T \right).$$

Wegen $\partial_{\sigma^2} m(x; \theta) = 0$ ist

$$\nabla_{\theta} m(x; \theta) [\nabla_{\theta} m(x; \theta)]^T = \begin{pmatrix} (\partial_a \gamma_0(\theta))^2 & \partial_a \gamma_0(\theta) (\partial_b \gamma_0(\theta) + \partial_b \gamma_1(\theta)x) & 0 \\ \partial_a \gamma_0(\theta) (\partial_b \gamma_0(\theta) + \partial_b \gamma_1(\theta)x) & (\partial_b \gamma_0(\theta) + \partial_b \gamma_1(\theta)x)^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

sowie aufgrund der Definition von w_v

$$\begin{aligned} & w_v(x; \theta) [\nabla_{\theta} v(x; \theta)]^T \\ &= \begin{pmatrix} 0 \\ 0 \\ \eta_0(\phi) + \eta_1(\phi)x \end{pmatrix} (\partial_a v(x; \theta), \partial_b v(x; \theta), \partial_{\sigma^2} v(x; \theta)) \\ &= (\eta_0(\phi) + \eta_1(\phi)x) \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \sigma^2 \partial_a \eta_0(\phi) & \sigma^2 (\partial_b \eta_0(\phi) + \partial_b \eta_1(\phi)x) & \eta_0(\phi) + \eta_1(\phi)x \end{pmatrix}. \end{aligned}$$

Im folgenden schreiben wir $\nu_k := \nu_k(\theta) := E_{\mu_{\theta}}[X_0^k]$ für das k -te Moment von X_0 unter $P_{\mu_{\theta}}$ und lassen der besseren Übersichtlichkeit halber das Argument θ der Funktionen $\nu_k, \gamma_0, \gamma_1, \eta_0, \eta_1$ sowie V, \tilde{V} und W weg. Dann berechnet man

$$\begin{aligned} V_{11} &= -(\partial_a \gamma_0)^2, \\ V_{12} &= V_{21}(\theta) = -\partial_a \gamma_0 (\partial_b \gamma_0 + \partial_b \gamma_1 \nu_1), \\ V_{13} &= V_{23}(\theta) = 0, \\ V_{22} &= -((\partial_b \gamma_0)^2 + 2\partial_b \gamma_0 \partial_b \gamma_1 \nu_1 + (\partial_b \gamma_1)^2 \nu_2), \\ V_{31} &= -\sigma^2 \partial_a \eta_0 (\eta_0 + \eta_1 \nu_1), \\ V_{32} &= -\sigma^2 (\eta_0 \partial_b \eta_0 + (\eta_0 \partial_b \eta_1 + \eta_1 \partial_b \eta_0) \nu_1 + \eta_1 \partial_b \eta_1 \nu_2), \\ V_{33} &= -(\eta_0^2 + 2\eta_0 \eta_1 \nu_1 + \eta_1^2 \nu_2). \end{aligned}$$

Insbesondere folgt wegen der Blockgestalt von V

$$\begin{aligned} \det V &= V_{33} \cdot (V_{11} V_{22} - V_{12}^2) \\ &= -(\eta_0^2 + 2\eta_0 \eta_1 \nu_1 + \eta_1^2 \nu_2) \cdot (\partial_a \gamma_0)^2 (\partial_b \gamma_1)^2 \underbrace{(\nu_2 - \nu_1^2)}_{= \text{Var}_{\mu_{\theta}}(X_0)} < 0, \end{aligned}$$

d.h. V ist invertierbar. Nun zur Berechnung von $W(\theta)$. Nach (3.2.42) ist

$$\begin{aligned} W(\theta) &= \mu_{\theta} \left(v(\cdot; \theta) w_m(\cdot; \theta) w_m(\cdot; \theta)^T + (\mu_4(\cdot; \theta) - v(\cdot; \theta)^2) w_v(\cdot; \theta) w_v(\cdot; \theta)^T \right. \\ &\quad \left. + \mu_3(\cdot; \theta) (w_m(\cdot; \theta) w_v(\cdot; \theta)^T + w_v(\cdot; \theta) w_m(\cdot; \theta)^T) \right). \end{aligned}$$

Dabei ist $w_m(x; \theta)w_m(x; \theta)^T = \nabla_\theta m(x; \theta)[\nabla_\theta m(x; \theta)]^T$ wie oben,

$$w_m(x; \theta)w_v(x; \theta)^T = \begin{pmatrix} 0 & 0 & \partial_a \gamma_0(\theta)(\eta_0(\theta) + \eta_1(\theta)x) \\ 0 & 0 & (\partial_b \gamma_0(\theta) + \partial_b \gamma_1(\theta)x)(\eta_0(\theta) + \eta_1(\theta)x) \\ 0 & 0 & 0 \end{pmatrix}$$

sowie

$$w_v(x; \theta)w_v(x; \theta)^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & (\eta_0(\theta) + \eta_1(\theta)x)^2 \end{pmatrix}.$$

Im folgenden lassen wir das Argument θ wieder weg. Dann berechnet sich W zu

$$\begin{aligned} W_{11} &= E_{\mu_\theta} [\sigma^2(\eta_0 + \eta_1 X_0)(\partial_a \gamma_0)^2] = \sigma^2(\partial_a \gamma_0)^2(\eta_0 + \eta_1 \nu_1), \\ W_{12} &= E_{\mu_\theta} [\sigma^2(\eta_0 + \eta_1 X_0) \cdot \partial_a \gamma_0 (\partial_b \gamma_0 + \partial_b \gamma_1 X_0)] \\ &= \sigma^2 \partial_a \gamma_0 (\eta_0 \partial_b \gamma_0 + (\eta_0 \partial_b \gamma_1 + \eta_1 \partial_b \gamma_0) \nu_1 + \eta_1 \partial_b \gamma_1 \nu_2), \\ W_{22} &= E_{\mu_\theta} [\sigma^2(\eta_0 + \eta_1 X_0) (\partial_b \gamma_0 + \partial_b \gamma_1 X_0)^2] \\ &= \sigma^2 \left(\eta_0 (\partial_b \gamma_0)^2 + (\eta_1 (\partial_b \gamma_0)^2 + 2\eta_0 \partial_b \gamma_0 \partial_b \gamma_1) \nu_1 \right. \\ &\quad \left. + (\eta_0 (\partial_b \gamma_1)^2 + 2\eta_1 \partial_b \gamma_0 \partial_b \gamma_1) \nu_2 + (\eta_1 (\partial_b \gamma_1)^2) \nu_3 \right), \\ W_{13} &= E_{\mu_\theta} [(\mu_{30} + \mu_{31} X_0) \partial_a \gamma_0 (\eta_0 + \eta_1 X_0)] \\ &= \partial_a \gamma_0 (\mu_{30} \eta_0 + (\mu_{31} \eta_0 + \mu_{30} \eta_1) \nu_1 + \mu_{31} \eta_1 \nu_2), \\ W_{23} &= E_{\mu_\theta} [(\mu_{30} + \mu_{31} X_0) (\partial_b \gamma_0 + \partial_b \gamma_1 X_0) (\eta_0 + \eta_1 X_0)] \\ &= \mu_{30} \eta_0 \partial_b \gamma_0 + (\mu_{30} \eta_0 \partial_b \gamma_1 + \mu_{31} \eta_0 \partial_b \gamma_0 + \mu_{30} \eta_1 \partial_b \gamma_0) \nu_1 \\ &\quad + (\mu_{31} \eta_0 \partial_b \gamma_1 + \mu_{31} \eta_1 \partial_b \gamma_0 + \mu_{30} \eta_1 \partial_b \gamma_1) \nu_2 + \mu_{31} \eta_1 \partial_b \gamma_1 \nu_3, \\ W_{33} &= E_{\mu_\theta} [(\mu_{40} + \mu_{41} X_0 + \mu_{42} X_0^2 - \sigma^4(\eta_0 + \eta_1 X_0)^2) (\eta_0 + \eta_1 X_0)^2] \\ &= E_{\mu_\theta} [((\mu_{40} - \sigma^4 \eta_0^2) + (\mu_{41} - 2\sigma^4 \eta_0 \eta_1) X_0 + (\mu_{42} - \sigma^4 \eta_1^2) X_0^2) (\eta_0 + \eta_1 X_0)^2] \\ &= (\mu_{40} - \sigma^4 \eta_0^2)(\eta_0^2 + 2\eta_0 \eta_1 \nu_1 + \eta_1^2 \nu_2) + (\mu_{41} - 2\sigma^4 \eta_0 \eta_1)(\eta_0^2 \nu_1 + 2\eta_0 \eta_1 \nu_2 + \eta_1^2 \nu_3) \\ &\quad + (\mu_{42} - \sigma^4 \eta_1^2)(\eta_0^2 \nu_2 + 2\eta_0 \eta_1 \nu_3 + \eta_1^2 \nu_4). \end{aligned}$$

Die übrigen Einträge sind durch die Symmetrie von W festgelegt.

Für die Schätzfolge $(\tilde{\theta}_n)_n$ ist eine solche explizite Darstellung von $\tilde{V}(\theta)$ und $\tilde{W}(\theta)$ nicht möglich, da $E_{\mu_\theta} \left[\frac{1}{\eta_0 + \eta_1 X_0} \right]$ nicht in geschlossener Form berechnet werden kann. Jedoch ist

leicht zu sehen, daß auch die Matrix $\tilde{V}(\theta)$ invertierbar ist: offensichtlich gilt nämlich

$$\begin{aligned} \tilde{w}_v(x; \theta)[\nabla_\theta v(x; \theta)]^T &= \begin{pmatrix} 0 \\ 0 \\ \frac{1}{\eta_0(\phi) + \eta_1(\phi)x} \end{pmatrix} (\partial_a v(x; \theta), \partial_b v(x; \theta), \partial_{\sigma^2} v(x; \theta)) \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{\sigma^2 \partial_a \eta_0(\phi)}{(\eta_0(\phi) + \eta_1(\phi)x)} & \frac{\sigma^2 (\partial_b \eta_0(\phi) + \partial_b \eta_1(\phi)x)}{(\eta_0(\phi) + \eta_1(\phi)x)} & 1 \end{pmatrix} \end{aligned}$$

und somit

$$\tilde{V}(\theta) = -\mu_\theta (w_m(\cdot; \theta)[\nabla_\theta m(\cdot; \theta)]^T + \tilde{w}_v(\cdot; \theta)[\nabla_\theta v(\cdot; \theta)]^T) = \begin{pmatrix} V_{11}(\theta) & V_{12}(\theta) & 0 \\ V_{21}(\theta) & V_{22}(\theta) & 0 \\ * & * & -1 \end{pmatrix},$$

woraus sich

$$\det \tilde{V}(\theta) = -(V_{11}V_{22} - V_{12}^2) = -(\partial_a \gamma_0)^2 (\partial_b \gamma_1)^2 (\nu_2 - \nu_1^2) < 0$$

ergibt. ◆

3.2.2 Schätzung mittels optimaler quadratischer Schätzfunktion

Die im vorhergehenden Unterabschnitt konstruierten Schätzfolgen $(\hat{\theta}_n)_n$ und $(\tilde{\theta}_n)_n$ sind stark konsistente Lösungen quadratischer Martingal-Schätzgleichungen. Allerdings konnten die zugehörigen asymptotischen Kovarianzmatrizes nicht analytisch verglichen und somit auch nicht entschieden werden, welches die „bessere“ Schätzfolge ist. Wir können allerdings fragen, ob es in der Klasse der quadratischen Martingal-Schätzfunktionen ein optimales Element im Sinne von Abschnitt 2.5 gibt. Dazu sei \mathcal{G} definiert wie in Abschnitt 2.5, und es bezeichne wieder \mathcal{G}_2 die Teilmenge aller $g \in \mathcal{G}$, die quadratische Martingal-Schätzfunktionen induzieren. Wir möchten für den **CIR**-Prozeß optimale Gewichte w_m^* und w_v^* bestimmen. Nach Satz 2.5.2 ist eine mögliche Wahl von w_m^* und w_v^* gegeben durch

$$w_m^*(x; \theta) = \frac{(\mu_4(x; \theta) - v(x; \theta)^2) \nabla_\theta m(x; \theta) - \mu_3(x; \theta) \nabla_\theta v(x; \theta)}{v(x; \theta)(\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2} \quad (3.2.44)$$

$$w_v^*(x; \theta) = \frac{v(x; \theta) \nabla_\theta v(x; \theta) - \mu_3(x; \theta) \nabla_\theta m(x; \theta)}{v(x; \theta)(\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2}, \quad (3.2.45)$$

vorausgesetzt, die zugehörige Funktion

$$g_*(x, y; \theta) := (y - m(x; \theta))w_m^*(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta))w_v^*(x; \theta) \quad (3.2.46)$$

ist ein Element in \mathcal{G} . Wir erhalten folgendes Resultat, welches sowohl (3.2.38) als auch (3.2.39) verbessert:

Theorem 3.2.9 *Sei g_* definiert durch (3.2.46). Dann ist $g_* \in \mathcal{G}$ und somit g_* ein optimales Element in \mathcal{G}_2 . Außerdem existiert eine stark konsistente Lösung $(\hat{\theta}_n^*)_n$ der Schätzgleichung*

$$0 = G_n^*(\theta) = \sum_{k=1}^n g_*(X_{k-1}, X_k; \theta), \quad (3.2.47)$$

mit

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\theta}_n^* - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, W_{g_*}(\theta)^{-1}) \quad \text{unter } P_\theta. \quad (3.2.48)$$

Beweis Nach Definition von \mathcal{G} ist zu zeigen: g_* erfüllt die Bedingungen **C1** und **I** aus Abschnitt 2.3 sowie $g_*(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ für jedes $\theta \in \Theta$.

Wir beobachten zunächst, daß w_m^* und w_v^* wohldefiniert sind: die Funktion

$$h(x; \theta) := v(x; \theta) (\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2$$

im Nenner von (3.2.44) und (3.2.45) ist nach der Bemerkung am Ende von Abschnitt 2.5 stets positiv, da nach (3.1.1) $Q_\theta(x, dy) \sim \lambda(dy)$, das Maß $Q_\theta(x, dy)$ also keine Punktmassen hat. Wir zeigen jetzt, daß $h(\cdot; \theta)$ für festes $\theta \in \Theta$ auf $\mathbb{R}_{\geq 0}$ sogar nach unten beschränkt weg von 0 ist. Wegen der Gestalt von v , μ_3 und μ_4 (siehe (3.2.7) und (3.2.8)) ist $h(\cdot; \theta)$ für festes $\theta \in \Theta$ ein Polynom in x vom Grad 3 mit Koeffizienten, die stetig von θ abhängen, d.h. h hat die Gestalt

$$h(x; \theta) = \sum_{i=0}^3 a_i(\theta) x^i$$

mit stetigen $a_i(\cdot) : \Theta \rightarrow \mathbb{R}$, $i = 0, \dots, 3$. Es kann explizit nachgerechnet werden, daß $a_i(\theta) > 0$ für alle $\theta \in \Theta$, woraus die Zwischenbehauptung folgt. Man kann aber auch wie folgt argumentieren: da $h(\cdot; \theta) > 0$ ein Polynom ist, gilt $\lim_{x \rightarrow \infty} h(x; \theta) = \infty$, und aus Stetigkeitsgründen folgt

$$\inf_{x \geq 0} h(x; \theta) = \min_{x \geq 0} h(x; \theta) =: C(\theta) > 0.$$

Dabei ist $\theta \mapsto C(\theta)$ stetig: sei dazu $(\vartheta_n)_n \subseteq \Theta$ mit $\vartheta_n \xrightarrow{n \rightarrow \infty} \theta \in \Theta$. Weiterhin sei $x_* \geq 0$ mit $h(x_*; \theta) = C(\theta) = \min_{x \geq 0} h(x; \theta)$ sowie $(x_n)_n \subseteq \mathbb{R}_{\geq 0}$ mit $h(x_n; \vartheta_n) = C(\vartheta_n) = \min_{x \geq 0} h(x; \vartheta_n)$. Dann ist die Folge $(x_n)_n$ beschränkt. Wäre dies nämlich nicht der Fall, so gäbe es eine Teilfolge $(x_{n_k})_k \subseteq \mathbb{R}_{> 0}$ mit $x_{n_k} \rightarrow \infty$ für $k \rightarrow \infty$. Da $x_{n_k} > 0$ eine Minimalstelle von $h(\cdot; \vartheta_{n_k})$ ist, folgt $\frac{d}{dx} h(x_{n_k}; \vartheta_{n_k}) = 0$ für alle k . Andererseits gilt

$$\frac{d}{dx} h(x_{n_k}; \vartheta_{n_k}) = \sum_{i=1}^3 i \cdot \underbrace{a_i(\vartheta_{n_k})}_{\rightarrow a_i(\theta)} x_{n_k}^{i-1} \xrightarrow{k \rightarrow \infty} \infty,$$

Widerspruch. Folglich gibt es $K < \infty$ mit $x_*, x_n \leq K$ für alle $n \in \mathbb{N}$, und es folgt

$$|C(\vartheta_n) - C(\theta)| = \left| \min_{0 \leq x \leq K} h(x; \vartheta_n) - \min_{0 \leq x \leq K} h(x; \theta) \right| \leq \max_{0 \leq x \leq K} |h(x; \vartheta_{n_k}) - h(x; \theta)| \xrightarrow{n \rightarrow \infty} 0$$

und damit die Stetigkeit von $C(\cdot)$.

Nun ist der Zähler in (3.2.44) und (3.2.45) jeweils komponentenweise ein Polynom in x mit stetig von θ abhängigen Koeffizienten, also hat die i -te Komponente von g_* aus (3.2.46) die Gestalt

$$g_{i*}(x, y; \theta) = \frac{\tilde{q}_i(x, y; \theta)}{h(x; \theta)},$$

wobei \tilde{q}_i ein Polynom in (x, y) ist, dessen Koeffizienten stetig von θ abhängen. Daraus folgt dann

$$|g_{i*}(x, y; \theta)|^2 \leq \frac{1}{C(\theta)^2} |\tilde{q}_i(x, y; \theta)|^2.$$

und damit nach mittlerweile hinlänglich bekannten Argumenten $g_{i*}(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$, $i = 1, \dots, 3$. Außerdem ergibt sich für die partielle Ableitung nach θ_j

$$\partial_j g_{i*}(x, y; \theta) = \frac{\partial_j \tilde{q}_i(x, y; \theta) \cdot h(x; \theta) - \tilde{q}_i(x, y; \theta) \cdot \partial_j h(x; \theta)}{h(x; \theta)^2} = \frac{\tilde{p}_{ij}(x, y; \theta)}{h(x; \theta)^2}$$

mit einem Polynom \tilde{p}_{ij} , dessen Koeffizienten wiederum stetig von θ abhängen. Folglich können wir für $i, j = 1, \dots, 3$ abschätzen:

$$|\partial_j g_{i*}(x, y; \theta)| \leq \frac{1}{C(\theta)^2} \cdot |\tilde{p}_{ij}(x, y; \theta)|,$$

und hier ist die Funktion auf der rechten Seite lokal dominiert integrierbar, wobei wir die Stetigkeit von $\theta \mapsto C(\theta)$ verwenden. Damit erfüllt g_* die Bedingung **C1**.

Nun zur Bedingung **I**. Wir müssen zeigen, daß für jedes $\theta \in \Theta$ die Matrix $V_{g_*}(\theta) = \tilde{\mu}_\theta(D_\theta g_*(\theta))$ invertierbar ist. Fixiere dazu $\theta \in \Theta$. Da g_* nach Konstruktion die Voraussetzungen von Lemma 2.5.1 erfüllt, gilt $V_{g_*}(\theta) = -W_{g_*}(\theta) = -\tilde{\mu}_\theta(g_*(\theta)g_*(\theta)^T)$, siehe (2.5.4). Sei $u = (u_1, u_2, u_3)^T \in \mathbb{R}^3$ mit $V_{g_*}(\theta)u = 0$. Dann gilt auch

$$0 = u^T W_{g_*}(\theta) u = \sum_{i,j=1}^3 u_i \tilde{\mu}_\theta(g_{i*}(\cdot, \cdot; \theta) g_{j*}(\cdot, \cdot; \theta)) u_j = \tilde{\mu}_\theta \left(\left[\sum_{i=1}^3 u_i g_{i*}(\cdot, \cdot; \theta) \right]^2 \right),$$

also $\sum_{i=1}^3 u_i g_{i*}(\cdot, \cdot; \theta) = 0$ in $L^2(\tilde{\mu}_\theta)$ und damit auch für $\tilde{\mu}_\theta$ -fast alle $(x, y) \in E^2$. Wegen $\tilde{\mu}_\theta \sim \lambda^2$ (siehe (3.1.3)) legt $\tilde{\mu}_\theta$ positive Masse auf jede offene Teilmenge von E^2 , und da

$(x, y) \mapsto g_*(x, y; \theta)$ stetig ist, folgt

$$\begin{aligned}
\forall (x, y) \in E^2 : 0 &= \sum_{i=1}^3 u_i g_{i*}(x, y; \theta) \\
&= \sum_{i=1}^3 u_i [(y - m(x; \theta)) w_{m,i}^*(x; \theta) + ((y - m(x; \theta))^2 - v(x; \theta)) w_{v,i}^*(x; \theta)] \\
&= y^0 \cdot \sum_{i=1}^3 u_i [(m(x; \theta)^2 - v(x; \theta)) w_{v,i}^*(x; \theta) - m(x; \theta) w_{m,i}^*(x; \theta)] \\
&\quad + y^1 \sum_{i=1}^3 u_i [w_{m,i}^*(x; \theta) - 2m(x; \theta) w_{v,i}^*(x; \theta)] \\
&\quad + y^2 \sum_{i=1}^3 u_i w_{v,i}^*(x; \theta).
\end{aligned} \tag{3.2.49}$$

Durch Vergleich der Koeffizienten für y und y^2 sowie eine elementare Umformung erhalten wir hieraus, daß für alle $x \in E$

$$0 = \sum_{i=1}^3 u_i w_{m,i}^*(x; \theta) \tag{3.2.50}$$

$$0 = \sum_{i=1}^3 u_i w_{v,i}^*(x; \theta). \tag{3.2.51}$$

Multiplikation der ersten Gleichung mit $v(x; \theta)$ und der zweiten Gleichung mit $\mu_3(x; \theta)$ sowie anschließende Addition ergibt

$$\begin{aligned}
\forall x \in E : 0 &= \sum_{i=1}^3 u_i [v(x; \theta) w_{m,i}^*(x; \theta) + \mu_3(x; \theta) w_{v,i}^*(x; \theta)] \\
&= \sum_{i=1}^3 u_i \left[v(x; \theta) \frac{(\mu_4(x; \theta) - v(x; \theta)^2) \partial_i m(x; \theta) - \mu_3(x; \theta) \partial_i v(x; \theta)}{h(x; \theta)} \right. \\
&\quad \left. + \mu_3(x; \theta) \frac{v(x; \theta) \partial_i v(x; \theta) - \mu_3(x; \theta) \partial_i m(x; \theta)}{h(x; \theta)} \right] \\
&= \sum_{i=1}^3 u_i \frac{1}{h(x; \theta)} \underbrace{[v(x; \theta) (\mu_4(x; \theta) - v(x; \theta)^2) - \mu_3(x; \theta)^2]}_{=h(x; \theta)} \partial_i m(x; \theta) \\
&= \sum_{i=1}^3 u_i \partial_i m(x; \theta) \\
&= u_1 \partial_a \gamma_0(\theta) + u_2 \partial_b \gamma_0(\theta) + u_2 \partial_b \gamma_1(\theta) x
\end{aligned} \tag{3.2.52}$$

aufgrund der Gestalt von m aus (3.2.2) (beachte $\partial_3 m(x; \theta) = \partial_{\sigma^2} m(x; \theta) = 0$). Ein Koeffizientenvergleich zeigt dann $u_2 = 0 = u_1$, wegen $\partial_b \gamma_1(\theta), \partial_a \gamma_0(\theta) > 0$. Aus Gleichung (3.2.50) ergibt sich dann wiederum wegen $\partial_{\sigma^2} m(x; \theta) = 0$ sowie der Gestalt (3.2.7) von μ_3 , daß für alle $x \in E$

$$0 = u_3 w_{m,3}^*(x; \theta) = u_3 \frac{\mu_3(x; \theta) \partial_{\sigma^2} v(x; \theta) - 0}{h(x; \theta)} = u_3 \frac{1}{h(x; \theta)} \underbrace{(\mu_{30}(\theta) + \mu_{31}(\theta)x) \cdot (\eta_0(\theta) + \eta_1(\theta)x)}_{>0},$$

also auch $u_3 = 0$. Damit ist die Invertierbarkeit von $V_{g_*}(\theta)$ gezeigt. Es folgt $g_* \in \mathcal{G}$ und mit Satz 2.5.2 die Optimalität von g_* in der Klasse G_2 . Da der Parameter θ im **CIR**-Modell nach dem vorhergehenden Unterabschnitt 3.2.1 stark konsistent schätzbar ist, existiert auch eine stark konsistente Lösung $(\hat{\theta}_n^*)_n$ der Schätzgleichung (3.2.47), und es gilt asymptotische Normalität (3.2.48), vgl. die Bemerkung nach Lemma 2.5.1. Die Schätzfolge $(\hat{\theta}_n^*)_n$ realisiert also die untere Schranke $W_{g_*}(\theta)^{-1}$ für die asymptotische Varianz von Schätzfolgen, die als Lösungen quadratischer Martingal-Schätzfunktionen (d.h. solcher mit $g \in \mathcal{G}_2$) definiert sind. ■

3.3 Likelihood-Schätzung und LAN im CIR-Modell

Im letzten Abschnitt der Arbeit wollen wir schließlich das diskretisierte **CIR**-Modell auf lokal-asymptotische Normalität (im folgenden wieder: **LAN**) und auf die Existenz einer stark konsistenten Lösung der Likelihood-Gleichung hin untersuchen. Wir betrachten also wie in Abschnitt 2.4 die Folge der Experimente

$$\mathcal{E}_n := (\Omega, \mathcal{F}_n, \{P_{n,\theta} : \theta \in \Theta\})$$

mit $P_{n,\theta} := P_\theta|_{\mathcal{F}_n}$. Die Gültigkeit der **LAN**-Bedingung wurde in [OR] nachgewiesen, allerdings ohne die Invertierbarkeit der Fisher-Information und nur für das Submodell $\{\theta \in \Theta : 2a/\sigma^2 > 1\}$. Diese Einschränkung wird sich jedoch anders als im Fall der Beobachtung in stetiger Zeit (siehe [Ove], Abschnitt 3) als unnötig herausstellen.

Der diskretisierte **CIR**-Prozeß erfüllt die zusätzlichen Voraussetzungen aus Abschnitt 2.4: nach Abschnitt 1.2 ist

$$Q_\theta(x, dy) = p(x, y; \theta) dy$$

wobei die Übergangsdichte p nach (1.2.3) strikt positiv festgelegt werden kann, etwa als

$$p(x, y; \theta) := \mathbf{1}_{\{0\}}(y) + \mathbf{1}_{(0,\infty)}(y) c(\theta)^{q(\theta)+1} y^{q(\theta)} \exp(-c(\theta)y - c(\theta)e^{b\Delta}x) \cdot \sum_{j=0}^{\infty} \frac{(c(\theta)^2 y x e^{b\Delta})^j}{j! \Gamma(q(\theta) + j + 1)} > 0$$

für $x, y \in E = \mathbb{R}_{\geq 0}$ und $\theta \in \Theta$. Hierbei

$$q(\theta) = \frac{2a}{\sigma^2} - 1 > -1, \quad c(\theta) = -\frac{2b}{\sigma^2(1 - e^{b\Delta})} > 0. \quad (3.3.1)$$

Offenbar gilt $q(\cdot), c(\cdot) \in C^\infty(\Theta)$. Insbesondere folgt wie in Abschnitt 2.4, daß für jedes $n \in \mathbb{N}$ die Wahrscheinlichkeitsmaße $\{P_{n,\theta} : \theta \in \Theta\}$ untereinander äquivalent sind (anders im Fall zeitstetiger Beobachtung: siehe [Ove], Abschnitt 4). Der Logarithmus der Übergangsdichte ist gegeben durch

$$\begin{aligned} \ell(x, y; \theta) &:= \log p(x, y; \theta) \\ &= (q(\theta) + 1) \log c(\theta) + q(\theta) \log y - c(\theta)y - c(\theta)e^{b\Delta}x + \log \left(\sum_{j=0}^{\infty} \frac{(c(\theta)^2 y x e^{b\Delta})^j}{j! \Gamma(q(\theta) + j + 1)} \right) \end{aligned} \quad (3.3.2)$$

für $y > 0$ sowie $\ell(x, 0; \theta) = 0$ für $y = 0$. Wir schreiben nun wie in Abschnitt 2.4

$$L_n(\theta) := \sum_{i=1}^n \ell(X_{i-1}, X_i; \theta)$$

für die log-Likelihood-Funktion und erhalten die Likelihood-Gleichung

$$\nabla_\theta L_n(\theta) = \sum_{i=1}^n \nabla_\theta \ell(X_{i-1}, X_i; \theta) = 0. \quad (3.3.3)$$

Der Einfachheit halber gehen wir im folgenden davon aus, daß die Startverteilung des Prozesses unter P_θ durch die invariante Verteilung μ_θ gegeben ist, $\mathcal{L}(X_0 | P_\theta) = \mu_\theta$. Nach Abschnitt 2.4 kann hier aber $\{\mu_\theta\}_{\theta \in \Theta}$ durch jede Familie $\{\nu_\theta\}_{\theta \in \Theta}$ mit $\nu_\theta \sim \mathbb{M}$ ersetzt werden oder die Startverteilung auch als bekannt angenommen werden.

Als Hauptresultat dieses Abschnitts erhalten wir als Konsequenz von Theorem 2.4.2:

Theorem 3.3.1 *Für den diskretisierten CIR-Prozeß gilt LAN an jeder Stelle $\theta \in \Theta$ mit normierender Folge $\delta_n(\theta) = n^{-1/2}$ und Fisher-Information*

$$K(\theta) = [\tilde{\mu}_\theta (\partial_i \ell(\cdot, \cdot; \theta) \cdot \partial_j \ell(\cdot, \cdot; \theta))]_{i,j=1,\dots,3} \in \mathbb{R}^{3 \times 3} \quad (3.3.4)$$

mit ℓ aus (3.3.2). Außerdem gibt es eine eindeutige stark konsistente Lösung $(\hat{\vartheta}_n^*)_n$ der Likelihood-Gleichung (3.3.3) mit

$$\forall \theta \in \Theta : P_\theta \left(\liminf_{n \rightarrow \infty} \{ \omega : \hat{\vartheta}_n^*(\omega) \text{ ist lokale Maximalstelle von } L_n(\omega; \cdot) \} \right) = 1, \quad (3.3.5)$$

und es gilt

$$\forall \theta \in \Theta : \sqrt{n}(\hat{\vartheta}_n^* - \theta) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, K(\theta)^{-1}) \quad \text{unter } P_\theta. \quad (3.3.6)$$

Der Beweis erfolgt über den Nachweis der Voraussetzungen von Theorem 2.4.2, die wir im folgenden in einer Reihe von Lemmata sorgfältig nachrechnen.

Zunächst müssen wir den letzten Summanden auf der rechten Seite von (3.3.2) genauer untersuchen. Dazu definieren wir für $z \in \mathbb{R}$ und $q > -1$

$$B(z, q) := \sum_{j=0}^{\infty} \frac{z^j}{j! \Gamma(q + j + 1)}$$

sowie

$$C := \left(\min_{x>0} \Gamma(x) \right)^{-1} < \infty.$$

Wegen der Abschätzung

$$\forall j \in \mathbb{N}_0 : \left| \frac{z^j}{j! \Gamma(q + j + 1)} \right| \leq C \cdot \frac{|z|^j}{j!} \quad (3.3.7)$$

und der lokal gleichmäßigen Konvergenz der Exponentialreihe konvergiert die Reihe in (3.3) lokal gleichmäßig in (z, q) auf $\mathbb{R} \times \mathbb{R}_{>-1}$, und folglich ist $B(\cdot, \cdot)$ stetig. Offenbar gilt

$$\forall z \geq 0, q > -1 : B(z, q) \geq \frac{1}{\Gamma(q + 1)} > 0. \quad (3.3.8)$$

Aus der Funktionalgleichung der Gammafunktion

$$\Gamma(x + 1) = x\Gamma(x), \quad x > 0$$

ergibt sich

$$\Gamma(x + 1) \geq \Gamma(x), \quad x \geq 1,$$

also auch

$$\begin{aligned} \forall z \geq 0, q > -1 : 0 < B(z, q + 1) &= \frac{1}{\Gamma(q + 2)} + \sum_{j=1}^{\infty} \frac{z^j}{j! \Gamma(q + j + 2)} \\ &\leq C + \sum_{j=1}^{\infty} \frac{z^j}{j! \Gamma(q + j + 1)} \\ &\leq C + B(z, q). \end{aligned} \quad (3.3.9)$$

Wir untersuchen die Funktion $B(\cdot, \cdot)$ nun auf Differenzierbarkeit. Die partiellen Ableitungen nach den Variablen z bzw. q werden dabei mit ∂_z bzw. ∂_q bezeichnet.

Lemma 3.3.2 *Sei $B : \mathbb{R} \times \mathbb{R}_{>-1} \rightarrow \mathbb{R}$ definiert wie in (3.3). Dann gilt $B(\cdot, \cdot) \in C^2(\mathbb{R} \times \mathbb{R}_{>-1})$, und auf $\mathbb{R}_{\geq 0} \times \mathbb{R}_{>-1}$ sind sämtliche partiellen Ableitungen erster und zweiter Ordnung nach Division durch $B(\cdot, \cdot)$ beschränkt durch ein Polynom in z mit stetig von q abhängigen Koeffizienten, genauer: für alle $z \geq 0, q > -1$ gilt*

$$\left| \frac{\partial_z B(z, q)}{B(z, q)} \right| \leq a_0(q), \quad \left| \frac{\partial_z^2 B(z, q)}{B(z, q)} \right| \leq a_0(q + 1)a_0(q), \quad (3.3.10)$$

$$\left| \frac{\partial_q B(z, q)}{B(z, q)} \right| \leq b_0(q) + b_1(q)z, \quad \left| \frac{\partial_{qz}^2 B(z, q)}{B(z, q)} \right| \leq a_0(q) (b_0(q+1) + b_1(q+1)z), \quad (3.3.11)$$

$$\left| \frac{\partial_q^2 B(z, q)}{B(z, q)} \right| \leq c_0(q) + c_1(q)z + c_2(q)z^2 \quad (3.3.12)$$

mit stetigen Funktionen $a_i(\cdot), b_i(\cdot), c_i(\cdot) : \mathbb{R}_{>-1} \rightarrow \mathbb{R}_{\geq 0}$.

Beweis

- Wir zeigen zuerst, daß $B(\cdot, \cdot) \in C^2(\mathbb{R} \times \mathbb{R}_{>-1})$. Für festes $q > -1$ ist $B(\cdot, q)$ offensichtlich eine Potenzreihe in z mit Konvergenzradius ∞ , also existieren die partiellen Ableitungen beliebig hoher Ordnung nach z und können durch formale Differentiation berechnet werden. So ergibt sich

$$\begin{aligned} \partial_z B(z, q) &= \sum_{j=0}^{\infty} \frac{jz^{j-1}}{j! \Gamma(q+j+1)} \\ &= \sum_{j=1}^{\infty} \frac{z^{j-1}}{(j-1)! \Gamma(q+j+1)} \\ &= \sum_{j=0}^{\infty} \frac{z^j}{j! \Gamma((q+1)+j+1)} \\ &= B(z, q+1), \end{aligned} \quad (3.3.13)$$

also auch

$$\partial_z^2 B(z, q) = B(z, q+2). \quad (3.3.14)$$

Zur Berechnung von $\partial_q B(z, q)$ differenzieren wir die Reihe (3.3) zunächst formal nach q und erhalten

$$\sum_{j=0}^{\infty} \partial_q \left[\frac{z^j}{j! \Gamma(q+j+1)} \right] = - \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{\Gamma'(q+j+1)}{\Gamma(q+j+1)^2} = - \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{\psi(q+j+1)}{\Gamma(q+j+1)} \quad (3.3.15)$$

mit der Riemannsches ψ -Funktion

$$\psi(x) := (\log \Gamma(x))' = \frac{\Gamma'(x)}{\Gamma(x)}, \quad x > 0. \quad (3.3.16)$$

Es bleibt nun noch zu zeigen, daß die Reihe (3.3.15) lokal gleichmäßig in q konvergiert. Nach 1.7 (3) in [Bat] hat die Funktion ψ die Darstellung

$$\psi(x) = -\gamma + (x-1) \sum_{n=0}^{\infty} \frac{1}{(n+1)(x+n)},$$

wobei $0 < \gamma < 1$ die Euler-Mascheroni-Konstante ist. Für $x \geq 1$ erhalten wir hieraus mit

$$1 < K_0 := \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} < \infty$$

die Abschätzung

$$|\psi(x)| \leq \gamma + (x-1) \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \leq K_0 (1 + (x-1)) = K_0 x, \quad x \geq 1, \quad (3.3.17)$$

d.h. ψ wächst auf $[1, \infty)$ sublinear. Da außerdem $\frac{x}{\Gamma(x)} \rightarrow 0$ für $x \rightarrow \infty$, gilt

$$K_1 := \sup_{x \geq 1} \frac{x}{\Gamma(x)} < \infty. \quad (3.3.18)$$

Wegen $q > -1$ liefern nun die Abschätzungen (3.3.17) und (3.3.18)

$$\forall j \geq 1 : \left| \frac{z^j \psi(q+j+1)}{j! \Gamma(q+j+1)} \right| \leq \frac{|z|^j}{j!} K_0 \cdot \frac{(q+j+1)}{\Gamma(q+j+1)} \leq K_1 \cdot K_0 \frac{|z|^j}{j!}$$

und somit lokal gleichmäßige Konvergenz sogar in (z, q) der Reihe (3.3.15). Damit ist gezeigt: $\partial_q B(z, q)$ existiert und ist stetig auf $\mathbb{R} \times \mathbb{R}_{>-1}$, und es gilt

$$\partial_q B(z, q) = - \sum_{j=0}^{\infty} \frac{z^j \psi(q+j+1)}{j! \Gamma(q+j+1)}. \quad (3.3.19)$$

Hieraus und aus (3.3.13) ergibt sich sofort

$$\partial_{zq}^2 B(z, q) = \partial_q (\partial_z B(z, q)) = \partial_q B(z, q+1). \quad (3.3.20)$$

Da $\partial_q B(z, q)$ nach (3.3.19) wieder eine Potenzreihe in z mit Konvergenzradius ∞ ist, berechnet sich $\partial_{zq}^2 B(z, q)$ durch formale Differentiation zu

$$\begin{aligned} \partial_{zq}^2 B(z, q) &= - \sum_{j=0}^{\infty} \frac{j z^{j-1} \psi(q+j+1)}{j! \Gamma(q+j+1)} = - \sum_{j=0}^{\infty} \frac{z^j \psi(q+j+2)}{j! \Gamma(q+j+2)} = \partial_q B(z, q+1) \\ &= \partial_{zq}^2 B(z, q). \end{aligned}$$

Schließlich ergibt formale Differentiation von $\partial_q B(z, q)$ nach q

$$\begin{aligned} & - \sum_{j=0}^{\infty} \partial_q \left[\frac{z^j \psi(q+j+1)}{j! \Gamma(q+j+1)} \right] \\ &= - \sum_{j=0}^{\infty} \frac{z^j \psi'(q+j+1) \Gamma(q+j+1) - \psi(q+j+1) \Gamma'(q+j+1)}{j! \Gamma(q+j+1)^2} \\ &= - \sum_{j=0}^{\infty} \frac{z^j \psi'(q+j+1) - \psi^2(q+j+1)}{j! \Gamma(q+j+1)}. \end{aligned} \quad (3.3.21)$$

Für ψ' gilt nach 1.16 (9) in [Bat] die Reihenentwicklung

$$\psi'(x) = \sum_{n=0}^{\infty} \frac{1}{(x+n)^2}, \quad x > 0,$$

also gilt

$$|\psi'(x)| = \psi'(x) \leq \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} = K_0, \quad x \geq 1. \quad (3.3.22)$$

Wegen $\frac{x^2}{\Gamma(x)} \rightarrow 0$ für $x \rightarrow \infty$ ist auch

$$K_2 := \sup_{x \geq 1} \frac{x^2}{\Gamma(x)} < \infty. \quad (3.3.23)$$

Hieraus und aus (3.3.17) sowie (3.3.22) erhalten wir für $j \geq 1$

$$\begin{aligned} \left| \frac{z^j \psi'(q+j+1) - \psi^2(q+j+1)}{j! \Gamma(q+j+1)} \right| &\leq \frac{|z|^j}{j!} \left(\frac{\psi'(q+j+1)}{\Gamma(q+j+1)} + \frac{|\psi(q+j+1)|^2}{\Gamma(q+j+1)} \right) \\ &\leq \frac{|z|^j}{j!} \cdot (K_0 C + K_0^2 K_2) \end{aligned} \quad (3.3.24)$$

und somit lokal gleichmäßige Konvergenz in (z, q) der Reihe (3.3.21). Daher gilt für alle $z \in \mathbb{R}, q > -1$

$$\partial_q^2 B(z, q) = - \sum_{j=0}^{\infty} \frac{z^j \psi'(q+j+1) - \psi^2(q+j+1)}{j! \Gamma(q+j+1)}. \quad (3.3.25)$$

Da sämtliche Reihen (3.3.14), (3.3.20) und (3.3.25) lokal gleichmäßig in (z, q) konvergieren, folgt Stetigkeit der zweiten partiellen Ableitungen, und $B(\cdot, \cdot) \in C^2(\mathbb{R} \times \mathbb{R}_{>-1})$ ist gezeigt.

2. Wir zeigen nun, daß für die partiellen Ableitungen erster und zweiter Ordnung auf $\mathbb{R}_{\geq 0} \times \mathbb{R}_{>-1}$ Abschätzungen der Form (3.3.10) - (3.3.12) erfüllt sind. Seien also $z \geq 0, q > -1$ beliebig. Nach (3.3.13) und (3.3.14) haben wir wegen (3.3.8) und (3.3.9)

$$\left| \frac{\partial_z B(z, q)}{B(z, q)} \right| = \left| \frac{B(z, q+1)}{B(z, q)} \right| = \frac{B(z, q+1)}{B(z, q)} \leq \frac{C + B(z, q)}{B(z, q)} \leq C \cdot \Gamma(q+1) + 1$$

und

$$\begin{aligned} \left| \frac{\partial_z^2 B(z, q)}{B(z, q)} \right| &= \frac{B(z, q+2)}{B(z, q+1)} \cdot \frac{B(z, q+1)}{B(z, q)} \leq \frac{C + B(z, q+1)}{B(z, q+1)} \cdot \frac{C + B(z, q)}{B(z, q)} \\ &\leq (C \cdot \Gamma(q+2) + 1) (C \cdot \Gamma(q+1) + 1), \end{aligned}$$

also gilt (3.3.10) mit $a_0(q) := C\Gamma(q+1) + 1$.

Aus (3.3.17) ergibt sich für $j \geq 2$

$$\begin{aligned} \frac{|\psi(q+j+1)|}{j!\Gamma(q+j+1)} &\leq K_0 \frac{q+j+1}{j!\Gamma(q+j+1)} \\ &= K_0 \frac{1}{j!\Gamma(q+j)} \cdot \frac{q+j+1}{q+j} \\ &\leq K_0 \frac{1}{(j-1)!\Gamma(q+j)} \cdot \underbrace{\left(1 + \frac{1}{q+j}\right)}_{\leq 2, \text{ da } j \geq 2} \\ &\leq 2K_0 \frac{1}{(j-1)!\Gamma(q+j)}. \end{aligned}$$

Für $j = 1$ haben wir $q+j+1 = q+2 > 1$ und damit wegen (3.3.17)

$$\frac{|\psi(q+2)|}{\Gamma(q+2)} \leq K_0 \frac{q+2}{\Gamma(q+2)} \leq K_0 K_1.$$

Zusammen mit (3.3.19) ergibt sich aus diesen Abschätzungen

$$\begin{aligned} |\partial_q B(z, q)| &\leq \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{|\psi(q+j+1)|}{\Gamma(q+j+1)} \\ &\leq \frac{|\psi(q+1)|}{\Gamma(q+1)} + K_0 K_1 z + 2K_0 \sum_{j=2}^{\infty} \frac{z^j}{(j-1)!\Gamma(q+j)} \\ &\leq \frac{|\psi(q+1)|}{\Gamma(q+1)} + K_0 z (K_1 + 2B(z, q)), \end{aligned}$$

also wegen (3.3.8) auch

$$\begin{aligned} \left| \frac{\partial_q B(z, q)}{B(z, q)} \right| &\leq |\psi(q+1)| \underbrace{\frac{1}{\Gamma(q+1)B(z, q)}}_{\leq 1 \text{ nach (3.3.8)}} + K_0 z \left(\frac{K_1}{B(z, q)} + 2 \right) \\ &\leq |\psi(q+1)| + K_0 z (K_1 \Gamma(q+1) + 2) =: b_0(q) + b_1(q)z \end{aligned}$$

und damit unter Verwendung von (3.3.9) auch

$$\begin{aligned} \left| \frac{\partial_{qz}^2 B(z, q)}{B(z, q)} \right| &= \left| \frac{\partial_q B(z, q+1)}{B(z, q)} \right| = \left| \frac{\partial_q B(z, q+1)}{B(z, q+1)} \right| \cdot \frac{B(z, q+1)}{B(z, q)} \\ &\leq \left| \frac{\partial_q B(z, q+1)}{B(z, q+1)} \right| \cdot \frac{C + B(z, q)}{B(z, q)} \\ &\leq \left(|\psi(q+2)| + K_0 z (K_1 \Gamma(q+2) + 2) \right) \cdot \left(C\Gamma(q+1) + 1 \right). \end{aligned}$$

Damit ist (3.3.11) gezeigt.

Schließlich haben wir aus (3.3.25)

$$|\partial_q^2 B(z, q)| \leq \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{|\psi'(q+j+1)| + \psi^2(q+j+1)}{\Gamma(q+j+1)}.$$

Aus (3.3.17) folgt für $j \geq 3$

$$\begin{aligned} \frac{|\psi(q+j+1)|^2}{j!\Gamma(q+j+1)} &\leq K_0^2 \frac{(q+j+1)^2}{j!\Gamma(q+j+1)} \\ &= K_0^2 \frac{(q+j+1)^2}{j!(q+j)(q+j-1)\Gamma(q+j-1)} \\ &= K_0^2 \frac{1}{j!\Gamma(q+j-1)} \cdot \frac{q+j+1}{q+j} \cdot \frac{q+j+1}{q+j-1} \\ &= K_0^2 \frac{1}{j!\Gamma(q+j-1)} \cdot \underbrace{\left(1 + \frac{1}{q+j}\right)}_{\leq 2} \cdot \underbrace{\left(1 + \frac{2}{q+j-1}\right)}_{\leq 3, \text{ da } j \geq 3} \\ &\leq 6K_0^2 \frac{1}{(j-2)!\Gamma(q+j-1)}, \end{aligned}$$

und für $j = 1, 2$

$$\frac{\psi(q+j+1)^2}{\Gamma(q+j+1)} \leq K_0^2 \frac{(q+j+1)^2}{\Gamma(q+j+1)} \leq K_0^2 K_2$$

mit K_2 aus (3.3.23). Verwenden wir noch (3.3.22), so ergibt sich insgesamt

$$\begin{aligned} |\partial_q^2 B(z, q)| &\leq \sum_{j=0}^{\infty} \frac{z^j}{j!} \frac{|\psi'(q+j+1)| + \psi(q+j+1)^2}{\Gamma(q+j+1)} \\ &\leq \frac{|\psi'(q+1)|}{\Gamma(q+1)} + \sum_{j=1}^{\infty} \frac{z^j}{j!} \frac{K_0}{\Gamma(q+j+1)} + \frac{\psi(q+1)^2}{\Gamma(q+1)} + K_0^2 K_2 \left(z + \frac{1}{2}z^2\right) \\ &\quad + 6K_0^2 \sum_{j=3}^{\infty} \frac{z^j}{(j-2)!\Gamma(q+j-1)} \\ &\leq \frac{|\psi'(q+1)| + \psi(q+1)^2}{\Gamma(q+1)} + K_0 B(z, q) + K_0^2 K_2 \left(z + \frac{1}{2}z^2\right) + 6K_0^2 z^2 B(z, q), \end{aligned}$$

also

$$\left| \frac{\partial_q^2 B(z, q)}{B(z, q)} \right| \leq |\psi'(q+1)| + \psi(q+1)^2 + K_0 \left(1 + K_0 K_2 \Gamma(q+1) \left(z + \frac{1}{2}z^2\right) + 6K_0 z^2\right),$$

wobei wir nochmals (3.3.8) verwendet haben. Das zeigt (3.3.12), und die Behauptung ist bewiesen. ■

Bevor wir an den Beweis der einzelnen Voraussetzungen von Theorem 2.4.2 gehen, benötigen wir noch zwei kleine technische Hilfssätze.

Lemma 3.3.3 *Für jedes $x \in E$ und $\theta \in \Theta$ gibt es eine Konstante $C(x; \theta) < \infty$, so daß*

$$\forall y > 0 : p(x, y; \theta) \leq C(x; \theta) \cdot y^{q(\theta)} e^{-c(\theta)y/2}.$$

Dabei ist für jedes feste $x \in E$ die Abbildung

$$\Theta \ni \theta \mapsto C(x; \theta)$$

stetig.

Beweis Seien $x \in E$, $\theta \in \Theta$ beliebig. In (1.1.7) wurde schon gezeigt, daß

$$p(x, y; \theta) \leq C_0(x; \theta) \cdot c(\theta)^{q(\theta)+1} \exp(-c(\theta)e^{b\Delta}x) \cdot y^{q(\theta)} e^{-c(\theta)y/2}$$

mit

$$C_0(x; \theta) := \sum_{j=0}^{\infty} \frac{(2c(\theta)e^{b\Delta}x)^j}{\Gamma(q(\theta) + j + 1)} < \infty$$

aus (1.1.6). Wir zeigen, daß $\theta \mapsto C_0(x; \theta)$ stetig ist. Dazu definiere

$$\tilde{B}(z, q) := \sum_{j=0}^{\infty} \frac{z^j}{\Gamma(q + j + 1)}, \quad z \geq 0, q > -1. \quad (3.3.26)$$

Wegen $q + j + 1 > j$ und der Monotonie der Gammafunktion auf dem Intervall $[2, \infty)$ gilt die Abschätzung

$$\forall j \geq 2 : \frac{z^j}{\Gamma(q + j + 1)} \leq \frac{z^j}{\Gamma(j)} = z \frac{z^{j-1}}{(j-1)!},$$

also konvergiert die Reihe (3.3.26) lokal gleichmäßig in (z, q) , und insbesondere ist $\tilde{B}(\cdot, \cdot)$ stetig. Als Komposition stetiger Funktionen ist damit auch $\theta \mapsto C_0(x; \theta) = \tilde{B}(2c(\theta)e^{b\Delta}x, q(\theta))$ stetig. Mit $C(x; \theta) := C_0(x; \theta) \cdot c(\theta)^{q(\theta)+1} \exp(-c(\theta)e^{b\Delta}x)$ folgt nun die Behauptung. ■

Im Beweis von Theorem 4.1 in [OR] begründen die Autoren ihre Beschränkung auf das Submodell $\{\theta \in \Theta : 2a/\sigma^2 > 1\}$ damit, daß unter dieser Bedingung die Funktion $\log x$ integrierbar unter dem invarianten Maß μ_θ sei. Die Einschränkung auf den Fall $2a/\sigma^2 > 1$ ist jedoch unnötig. Hierzu stellen wir fest:

Feststellung 3.3.4 Seien $\alpha > 0$, $\beta > 0$ und $k \in \mathbb{N}_0$ beliebig. Dann ist

$$\int_0^\infty x^{\alpha-1} e^{-\beta x} |\log x|^k dx < \infty. \quad (3.3.27)$$

Aussage (3.3.27) folgt durch eine Substitution sofort aus $\int_0^\infty x^{\alpha-1} e^{-x} |\log x|^k dx < \infty$ für $\alpha > 0$ und $k \in \mathbb{N}_0$. In der Tat gilt

$$\int_0^\infty x^{\alpha-1} e^{-x} (\log x)^k dx = \Gamma^{(k)}(\alpha), \quad \alpha > 0,$$

siehe z.B. [Kab], (13.8). Hieraus erhalten wir nun:

Lemma 3.3.5 Für jedes $\theta \in \Theta$ und $k \in \mathbb{N}_0$ ist die Funktion $(x, y) \mapsto \log y$ in $L^k(\tilde{\mu}_\theta)$.

Beweis Wegen der Invarianz von μ_θ ist

$$\begin{aligned} \tilde{\mu}_\theta(|\log y|^k) &= \int_E \mu_\theta(dx) \int_E Q_\theta(x, dy) |\log y|^k = \int_E \mu_\theta(dx) |\log x|^k \\ &= \int_E \gamma(x; \theta) |\log x|^k dx = \frac{r(\theta)^{\alpha(\theta)}}{\Gamma(\alpha(\theta))} \int_0^\infty x^{\alpha(\theta)-1} e^{-r(\theta)x} |\log x|^k dx < \infty \end{aligned}$$

aufgrund von Feststellung 3.3.4, da nach (3.1.2)

$$\alpha(\theta) = \frac{2a}{\sigma^2} > 0, \quad r(\theta) = \frac{-2b}{\sigma^2} > 0. \quad \blacksquare$$

Wir gehen nun an den Beweis der einzelnen Voraussetzungen von Theorem 2.4.2. Das nächste Lemma zeigt, daß $\nabla_\theta \ell(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ für $\theta \in \Theta$ und daß $\nabla_\theta \ell$ die Bedingung (2.2.6) erfüllt. Dies stellt insbesondere sicher, daß die Folge $(\nabla_\theta L_n)_n$ der Scores eine Folge von Martingal-Schätzfunktionen ist.

Lemma 3.3.6 Sei $\theta \in \Theta$ beliebig. Dann sind für $i = 1, \dots, 3$ die Funktionen

$$E^2 \ni (x, y) \mapsto \partial_i \ell(x, y; \theta) \quad (3.3.28)$$

in $L^2(\tilde{\mu}_\theta)$, und es gilt

$$\forall x \in E : \int_E \partial_i p(x, y; \theta) dy = 0. \quad (3.3.29)$$

Beweis Fixiere ein beliebiges $\theta = (a, b, \sigma^2) \in \Theta$ und $(x, y) \in E^2$. Mit $z(\theta) := c(\theta)^2 e^{b\Delta} > 0$ und der Funktion $B(\cdot, \cdot)$ aus (3.3) schreibt sich ℓ gemäß (3.3.2) als

$$\ell(x, y; \theta) = (q(\theta) + 1) \log c(\theta) + q(\theta) \log y - c(\theta)y - c(\theta)e^{b\Delta}x + \log B(z(\theta)xy, q(\theta)). \quad (3.3.30)$$

Schreibe wieder ∂_z, ∂_q für die partiellen Ableitungen von $B(\cdot, \cdot)$ nach z und q . Betrachten wir zuerst den letzten Term in (3.3.30), so ergibt partielle Differentiation nach $\theta_i, i = 1, \dots, 3$

$$\begin{aligned} \partial_i \log B(z(\theta)xy, q(\theta)) &= \partial_z \log B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta) \cdot xy + \partial_q \log B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta) \\ &= \frac{\partial_z B(z(\theta)xy, q(\theta))}{B(z(\theta)xy, q(\theta))} \cdot \partial_i z(\theta) \cdot xy + \frac{\partial_q B(z(\theta)xy, q(\theta))}{B(z(\theta)xy, q(\theta))} \cdot \partial_i q(\theta). \end{aligned} \quad (3.3.31)$$

Aus Lemma 3.3.2 folgt

$$\begin{aligned} |\partial_i \log B(z(\theta)xy, q(\theta))| &\leq \left| \frac{\partial_z B(z(\theta)xy, q(\theta))}{B(z(\theta)xy, q(\theta))} \right| \cdot |\partial_i z(\theta)xy| + \left| \frac{\partial_q B(z(\theta)xy, q(\theta))}{B(z(\theta)xy, q(\theta))} \right| \cdot |\partial_i q(\theta)| \\ &\leq a_0(q(\theta)) \cdot |\partial_i z(\theta)xy| + \left(b_0(q(\theta)) + b_1(q(\theta))z(\theta)xy \right) \cdot |\partial_i q(\theta)|, \end{aligned}$$

wobei $a_0(\cdot), b_0(\cdot), b_1(\cdot) : \mathbb{R}_{>-1} \rightarrow \mathbb{R}$ stetig sind. Differenzieren wir auch die übrigen Terme in (3.3.30) und schätzen wir ab, so erhalten wir insgesamt

$$|\partial_i \ell(x, y; \theta)| \leq r_0^{(i)}(\theta) + r_1^{(i)}(\theta)x + r_2^{(i)}(\theta)y + r_3^{(i)}(\theta)xy + r_4^{(i)}(\theta)|\log y| =: h_i(x, y; \theta) \quad (3.3.32)$$

mit stetigen Funktionen $r_k^{(i)} : \Theta \rightarrow \mathbb{R}_{\geq 0}, k = 0, \dots, 4$. Da der Prozeß $(X_n)_n$ unter P_{μ_θ} Momente beliebig hoher Ordnung hat, sind die ersten vier Terme auf der rechten Seite von (3.3.32) in $L^2(\tilde{\mu}_\theta)$. Für den Term mit $|\log y|$ gilt dasselbe wegen Lemma 3.3.5. Damit ist $\partial_i \ell(\cdot, \cdot; \theta) \in L^2(\tilde{\mu}_\theta)$ gezeigt.

Zeige nun (3.3.29). Dazu beobachten wir zunächst, daß nach Lemma 3.3.3 gilt

$$p(x, y; \theta) \leq C(x; \theta) \cdot y^{q(\theta)} e^{-c(\theta)y/2},$$

wobei $\theta \mapsto C(x; \theta)$ stetig ist. Wegen $\partial_i \ell = \partial_i p/p$ erhalten wir aus (3.3.32) also

$$\begin{aligned} |\partial_i p(x, y; \theta)| &\leq p(x, y; \theta) \cdot h_i(x, y; \theta) \\ &\leq C(x; \theta) y^{q(\theta)} e^{-c(\theta)y/2} \cdot h_i(x, y; \theta). \end{aligned}$$

Fixiere nun $x \in E$ und $\theta \in \Theta$ wähle eine beliebige kompakte Umgebung $U \subseteq \Theta$ von θ . Aus der Stetigkeit von $q(\cdot), c(\cdot), C(x; \cdot)$ und den $r_k^{(i)}(\cdot)$ erhalten wir mit

$$-1 < q_* := \min_{\vartheta \in U} q(\vartheta) \leq \max_{\vartheta \in U} q(\vartheta) =: q^*, \quad c_* := \min_{\vartheta \in U} c(\vartheta) > 0$$

und

$$C := C(x) := \max_{\vartheta \in U} C(x; \vartheta) > 0, \quad C_k^{(i)} := \max_{\vartheta \in U} r_k^{(i)}(\vartheta) \geq 0,$$

daß für alle $y > 0$

$$\begin{aligned}
& \sup_{\vartheta \in U} |\partial_i p(x, y; \vartheta)| \\
& \leq \sup_{\vartheta \in U} \{C(x; \vartheta) y^{q(\vartheta)} e^{-c(\vartheta)y/2} \cdot h_i(x, y; \vartheta)\} \\
& \leq C (y^{q^*} + y^{q^*}) e^{-c^*/2} \cdot \left(C_0^{(i)} + C_1^{(i)} x + C_2^{(i)} y + C_3^{(i)} xy + C_4^{(i)} |\log y| \right)
\end{aligned} \tag{3.3.33}$$

Unter Verwendung von Feststellung 3.3.4 sieht man nun, daß die Funktion auf der rechten Seite von (3.3.33) (als Funktion von y) Lebesgue-integrierbar auf $[0, \infty)$ ist. Nach einem bekannten Satz aus der Lebesgueschen Integrationstheorie (siehe z.B. [Els], Satz IV.5.7) darf somit die Funktion $\theta \mapsto \int_E p(x, y; \theta) dy \equiv 1$ partiell unter dem Integralzeichen differenziert werden, und es folgt (3.3.29). ■

Im nächsten Lemma untersuchen wir die zweiten Ableitungen von ℓ :

Lemma 3.3.7 *Für alle $i, j = 1, \dots, 3$ sind die Funktionen*

$$E^2 \times \Theta \ni (x, y; \theta) \mapsto \partial_{ij}^2 \ell(x, y; \theta)$$

lokal dominiert integrierbar bezüglich der Familie $(\tilde{\mu}_\theta)_{\theta \in \Theta}$, und es gilt

$$\forall x \in E : \int_E \partial_{ij}^2 p(x, y; \theta) dy = 0. \tag{3.3.34}$$

Beweis Schreibe wieder $z(\theta) := c(\theta)^2 e^{b\Delta} > 0$, so daß also $z \in C^\infty(\Theta)$ und

$$\ell(x, y; \theta) = (q(\theta) + 1) \log c(\theta) + q(\theta) \log y - c(\theta)y - c(\theta)e^{b\Delta}x + \log B(z(\theta)xy, q(\theta)). \tag{3.3.35}$$

Wir differenzieren wieder zuerst den letzten Term in der Summe zweimal partiell: das ergibt

$$\begin{aligned}
& \partial_{ij}^2 \log B(z(\theta)xy, q(\theta)) \\
&= \frac{\partial_{ij}^2 B(z(\theta)xy, q(\theta)) \cdot B(z(\theta)xy, q(\theta)) - [\partial_i B(z(\theta)xy, q(\theta))] \cdot [\partial_j B(z(\theta)xy, q(\theta))]}{B(z(\theta)xy, q(\theta))^2} \\
&= \frac{\partial_i [\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_j z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_j q(\theta)]}{B(z(\theta)xy, q(\theta))} \\
&\quad - \frac{\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta)}{B(z(\theta)xy, q(\theta))} \\
&\quad \times \frac{\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_j z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_j q(\theta)}{B(z(\theta)xy, q(\theta))} \\
&= \frac{[\partial_z^2 B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta)xy + \partial_{zq}^2 B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta)] \cdot \partial_j z(\theta)xy}{B(z(\theta)xy, q(\theta))} \\
&\quad + \frac{\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_{ij}^2 z(\theta)xy}{B(z(\theta)xy, q(\theta))} \\
&\quad + \frac{[\partial_{zq}^2 B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta)xy + \partial_q^2 B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta)] \cdot \partial_j q(\theta)}{B(z(\theta)xy, q(\theta))} \\
&\quad + \frac{\partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_{ij}^2 q(\theta)}{B(z(\theta)xy, q(\theta))} \\
&\quad - \frac{\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta)}{B(z(\theta)xy, q(\theta))} \\
&\quad \times \frac{\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_j z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_j q(\theta)}{B(z(\theta)xy, q(\theta))}.
\end{aligned}$$

Ähnlich wie im Beweis von Lemma 3.3.6 folgert man nun aus den Abschätzungen (3.3.12) in Lemma 3.3.2, daß jeder der Summanden beschränkt ist durch ein Polynom in xy vom Grad ≤ 3 , dessen Koeffizienten stetig von θ abhängen. Durch zweimaliges partielles Differenzieren der restlichen Terme in (3.3.35) und Abschätzen sieht man dann: es gibt stetige Funktionen $\tilde{r}_k : \Theta \rightarrow \mathbb{R}_{\geq 0}$, $k = 0, \dots, 6$, so daß

$$\begin{aligned}
|\partial_{ij}^2 \ell(x, y; \theta)| &\leq \tilde{r}_0(\theta) + \tilde{r}_1(\theta)x + \tilde{r}_2(\theta)y + \tilde{r}_3(\theta)xy + \tilde{r}_4(\theta)x^2y^2 + \tilde{r}_5(\theta)x^3y^3 + \tilde{r}_6(\theta)|\log y| \\
&=: h_{ij}(x, y; \theta).
\end{aligned} \tag{3.3.36}$$

Nun ist jeder der Terme in obiger Summe von der Form $f(\theta) \cdot g(x, y)$ mit stetigem f und $g \in L^1(\tilde{\mu}_\theta)$, also lokal dominiert integrierbar. Das zeigt die erste Behauptung.

Zeige nun (3.3.34); auch hier geht der Beweis ähnlich wie in Lemma 3.3.6. Sei wieder h_i die Majorante für $\partial_i \ell$ aus (3.3.32), $i = 1, \dots, 3$. Wir haben

$$\partial_{ij}^2 \ell(x, y; \theta) = \frac{\partial_{ij}^2 p(x, y; \theta)}{p(x, y; \theta)} - \partial_i \ell(x, y; \theta) \cdot \partial_j \ell(x, y; \theta),$$

also

$$\begin{aligned} \left| \frac{\partial_{ij}^2 p(x, y; \theta)}{p(x, y; \theta)} \right| &\leq |\partial_{ij}^2 \ell(x, y; \theta)| + |\partial_i \ell(x, y; \theta) \cdot \partial_j \ell(x, y; \theta)| \\ &\leq h_{ij}(x, y; \theta) + h_i(x, y; \theta) h_j(x, y; \theta) \end{aligned}$$

und somit

$$|\partial_{ij}^2 p(x, y; \theta)| \leq p(x, y; \theta) \cdot (h_{ij}(x, y; \theta) + h_i(x, y; \theta) h_j(x, y; \theta)) \quad (3.3.37)$$

$$\leq C(x; \theta) \cdot y^{q(\theta)} e^{-c(\theta)y/2} \cdot (h_{ij}(x, y; \theta) + h_i(x, y; \theta) h_j(x, y; \theta)). \quad (3.3.38)$$

Nun ist $h_{ij}(x, y; \theta) + h_i(x, y; \theta) h_j(x, y; \theta)$ im wesentlichen von derselben Form wie die h_i aus (3.3.32), nämlich ein Polynom in x, y und $|\log y|$ mit stetig von θ abhängigen Koeffizienten. Analoge Abschätzungen und Überlegungen wie im Beweis von Lemma 3.3.6 unter Verwendung von Feststellung 3.3.4 ergeben dann, daß $\theta \mapsto \int_E \partial_i p(x, y; \theta) dy \equiv 0$ partiell nach θ_j unter dem Integral differenziert werden darf, und es folgt die Behauptung.

Schließlich bleibt noch die Invertierbarkeit der Fisher-Information zu beweisen:

Lemma 3.3.8 *Für alle $\theta \in \Theta$ ist die Matrix $K(\theta)$ aus (3.3.4) invertierbar.*

Beweis Der Beweis geht analog zum Nachweis der Invertierbarkeit von $V_{g_*}(\theta)$ aus Theorem 3.2.9. Sei $\theta = (a, b, \sigma^2) \in \Theta$ beliebig und $u = (u_1, u_2, u_3)^T \in \mathbb{R}^3$ mit $K(\theta)u = 0$. Genau wie im Beweis von Theorem 3.2.9 zeigt man, daß dann

$$\sum_{i=1}^3 u_i \partial_i \ell(x, y; \theta) = 0 \quad \text{für } \tilde{\mu}_\theta\text{-fast alle } (x, y) \in E^2.$$

Wegen $\tilde{\mu}_\theta \sim \lambda^2$ und da $p(x, y; \theta)$ stetig in (x, y) ist, erhalten wir, daß für alle $x, y > 0$

$$\begin{aligned} 0 &= \sum_{i=1}^3 u_i \partial_i \ell(x, y; \theta) \\ &= \sum_{i=1}^3 u_i \partial_i [(q(\theta) + 1) \log c(\theta) + q(\theta) \log y - c(\theta)y - c(\theta)e^{b\Delta}x + \log B(z(\theta)xy, q(\theta))] \\ &= \sum_{i=1}^3 u_i \left(\partial_i [(q(\theta) + 1) \log c(\theta)] + \partial_i q(\theta) \log y - \partial_i c(\theta)y - \partial_i [z(\theta)/c(\theta)] x + \frac{\partial_i B(z(\theta)xy, q(\theta))}{B(z(\theta)xy, q(\theta))} \right) \end{aligned}$$

Hier haben wir wieder $z(\theta) := c(\theta)^2 e^{b\Delta} > 0$ geschrieben; beachte $c(\theta)e^{b\Delta} = z(\theta)/c(\theta)$.

Durch Multiplikation mit $B(z(\theta)xy, q(\theta))$ ergibt sich, daß für alle $x, y > 0$

$$\begin{aligned}
0 &= B(z(\theta)xy, q(\theta)) \cdot \sum_{i=1}^3 u_i (\partial_i [(q(\theta) + 1) \log c(\theta)] + \partial_i q(\theta) \log y - \partial_i c(\theta) y - \partial_i [z(\theta)/c(\theta)] x) \\
&\quad + \sum_{i=1}^3 u_i (\partial_z B(z(\theta)xy, q(\theta)) \cdot \partial_i z(\theta)xy + \partial_q B(z(\theta)xy, q(\theta)) \cdot \partial_i q(\theta)) \\
&= B(z(\theta)xy, q(\theta)) \cdot \sum_{i=1}^3 u_i (\partial_i [(q(\theta) + 1) \log c(\theta)] + \partial_i q(\theta) \log y - \partial_i c(\theta) y - \partial_i [z(\theta)/c(\theta)] x) \\
&\quad + \sum_{i=1}^3 u_i \left(B(z(\theta)xy, q(\theta) + 1) \cdot \partial_i z(\theta)xy - \sum_{j=0}^{\infty} \frac{(z(\theta)xy)^j \psi(q + j + 1)}{j! \Gamma(q + j + 1)} \cdot \partial_i q(\theta) \right),
\end{aligned}$$

wobei wir die Formeln für $\partial_z B$ und $\partial_q B$ aus (3.3.13) und (3.3.19) verwendet haben. Im folgenden lassen wir das Argument θ der Funktionen $c(\cdot), q(\cdot)$ und $z(\cdot)$ der besseren Übersichtlichkeit halber weg. Sei nun $y > 0$ beliebig, aber fest. Dann können wir die rechte Seite als eine Potenzreihe in x mit von y abhängigen Koeffizienten schreiben. Zusammenfassung der x^j -Terme führt nach längerer Rechnung zu

$$\begin{aligned}
\forall x > 0 : 0 &= \frac{1}{\Gamma(q + 1)} \sum_{i=1}^3 u_i \left(\partial_i [(q + 1) \log c] - \partial_i c \cdot y + \partial_i q \cdot \log y - \psi(q + 1) \partial_i q \right) \\
&\quad + \sum_{j=1}^{\infty} x^j \cdot \frac{y^{j-1} z^{j-1}}{j! \Gamma(q + j + 1)} \left[-j(q + j) \sum_{i=1}^3 u_i \partial_i [z/c] \right. \\
&\quad \quad + y \sum_{i=1}^3 u_i \left(z \partial_i [(q + 1) \log c] + z \partial_i q (\log y - \psi(q + j + 1)) + j \partial_i z \right) \\
&\quad \quad \left. - y^2 \sum_{i=1}^3 u_i z \partial_i c \right].
\end{aligned} \tag{3.3.39}$$

Mit dem Identitätssatz für Potenzreihen folgt nun, daß alle Koeffizienten gleich 0 sind. Insbesondere ergibt sich für $j = 1$

$$\begin{aligned}
0 &= \frac{1}{\Gamma(q + 2)} \left[- (q + 1) \sum_{i=1}^3 u_i \partial_i [z/c] \right. \\
&\quad \left. + y \sum_{i=1}^3 u_i \left(z \partial_i [(q + 1) \log c] + z \partial_i q (\log y - \psi(q + 2)) + \partial_i z \right) - y^2 \sum_{i=1}^3 u_i z \partial_i c \right].
\end{aligned} \tag{3.3.40}$$

Da $y > 0$ beliebig war, gilt dies insbesondere für alle y in einer Umgebung U der 1, auf der eine Reihenentwicklung des Logarithmus existiert. Wir können dann die rechte Seite von (3.3.40) als Potenzreihe in y um 1 schreiben. Zusammenfassen der Terme führt nach einiger Algebra zu

$$\begin{aligned} \forall y \in U : 0 &= \sum_{i=1}^3 u_i (z \partial_i [(q+1) \log c] - z \psi(q+2) \partial_i q - z \partial_i c + \partial_i z - (q+1) \partial_i [z/c]) \\ &+ (y-1) \sum_{i=1}^3 u_i (z \partial_i [(q+1) \log c] + z \partial_i q (1 - \psi(q+2)) - 2z \partial_i c + \partial_i z) \\ &+ (y-1)^2 \sum_{i=1}^3 u_i \left(\frac{1}{2} z \partial_i q - z \partial_i c \right) \\ &+ \sum_{j=3}^{\infty} \frac{(-1)^j}{j(j-1)} (y-1)^j \cdot \sum_{i=1}^3 u_i z \partial_i q. \end{aligned}$$

Schreibe nun $s(\theta) := e^{b\Delta}$; dann ist $z(\theta) = c(\theta)^2 s(\theta)$ und $\partial_i z(\theta) = 2c(\theta) s(\theta) \partial_i c(\theta) + c(\theta)^2 \partial_i s(\theta)$. Eine erneute Anwendung des Identitätssatzes für Potenzreihen liefert für $j = 1, 2, 3$ das Gleichungssystem

$$0 = \sum_{i=1}^3 u_i \left(z \partial_i q (\log c + 1 - \psi(q+2)) + \partial_i c \left((q+1) \frac{z}{c} - 2z + 2cs \right) + c^2 \partial_i s \right) \quad (3.3.41)$$

$$0 = \sum_{i=1}^3 u_i (z \partial_i q - 2z \partial_i c) \quad (3.3.42)$$

$$0 = \sum_{i=1}^3 u_i z \partial_i q, \quad (3.3.43)$$

das sich durch elementare Umformungen auf die Gestalt

$$0 = \sum_{i=1}^3 u_i \partial_i s \quad (3.3.44)$$

$$0 = \sum_{i=1}^3 u_i \partial_i c \quad (3.3.45)$$

$$0 = \sum_{i=1}^3 u_i \partial_i q \quad (3.3.46)$$

bringen läßt. Nun sind aber die drei Vektoren $(\partial_i s, \partial_i c, \partial_i q)^T \in \mathbb{R}^3$, $i = 1, \dots, 3$ linear unabhängig, denn wegen

$$\partial_1 s = \partial_3 s = 0 = \partial_1 c$$

folgt

$$\det \begin{pmatrix} \partial_1 s & \partial_2 s & \partial_3 s \\ \partial_1 c & \partial_2 c & \partial_3 c \\ \partial_1 q & \partial_2 q & \partial_3 q \end{pmatrix} = \det \begin{pmatrix} 0 & \partial_2 s & 0 \\ 0 & \partial_2 c & \partial_3 c \\ \partial_1 q & \partial_2 q & \partial_3 q \end{pmatrix} = \partial_1 q \cdot \partial_2 s \cdot \partial_3 c = \frac{2}{\sigma^2} \Delta e^{b\Delta} \frac{2b}{e^{b\Delta} - 1} \left(-\frac{1}{\sigma^4} \right) < 0.$$

Damit folgt $u_1 = u_2 = u_3 = 0$, und die Behauptung ist bewiesen. \blacksquare

Beweis von Theorem 3.3.1 Da wir davon ausgegangen sind, daß die Startverteilung durch

$$\mu_\theta(dx) = \gamma(x; \theta) dx = \frac{r(\theta)^{\alpha(\theta)}}{\Gamma(\alpha(\theta))} x^{\alpha(\theta)-1} e^{-r(\theta)x} dx$$

parametrisiert ist und $\theta \mapsto \gamma(x; \theta)$ wegen der Gestalt von α und r (vgl. (3.1.2)) offensichtlich stetig ist, ist die erste Voraussetzung in Theorem 2.4.2 erfüllt. Weiterhin ist der Parameter $\theta \in \Theta$ stark konsistent schätzbar, wie in Abschnitt 3.2 gezeigt wurde. Die übrigen Voraussetzungen von Theorem 2.4.2 wurden in den Lemmata 3.3.6, 3.3.7 und 3.3.8 bewiesen. Dies schließt den Beweis von Theorem 3.3.1 ab.

Bemerkungen

- Sei wieder \mathcal{G} definiert wie in Abschnitt 2.5. Aus den Lemmata 3.3.6, 3.3.7 und 3.3.8 ergibt sich dann, daß $\nabla_\theta \ell \in \mathcal{G}$. Falls für jedes $g \in \mathcal{G}$ die Funktion

$$\theta \mapsto \int_E g(x, y; \theta) p(x, y; \theta) dy$$

komponentenweise partiell unter dem Integralzeichen differenziert werden darf, ist also $\nabla_\theta \ell$ ein optimales Element in \mathcal{G} , vgl. die dritte Bemerkung nach Lemma 2.5.1. Wir beschränken uns hier auf die Feststellung, daß $\nabla_\theta \ell$ jedenfalls in der Teilklasse $\tilde{\mathcal{G}}$ aller $g \in \mathcal{G}$, für die Differentiation unter dem Integralzeichen erlaubt ist, ein optimales Element ist. Man überzeugt sich mit Argumenten wie im Beweis von Lemma 3.3.6 ohne größere Schwierigkeiten davon, daß $g \in \tilde{\mathcal{G}}$, falls g komponentenweise durch ein Polynom in (x, y) mit stetig von θ abhängigen Koeffizienten majorisiert ist. Daher ist insbesondere das zur optimalen quadratischen Schätzfunktion gehörende g_* aus (3.2.46) in $\tilde{\mathcal{G}}$. Somit ist Theorem 3.3.1 eine Verbesserung von Theorem 2.5.2.

- Mit Theorem 3.3.1 gelten natürlich auch die in der Bemerkung nach Definition 2.4.1 kurz angesprochenen Implikationen der LAN-Bedingung. Wir bemerken noch kurz, daß sich die in [Hö], Kapitel 7.B angegebenen Zusatzvoraussetzungen für die Ein-Schritt-Modifikation von Schätzern im Falle des diskretisierten CIR-Modells ohne größeren Aufwand nachweisen lassen.

\blacklozenge

Anhang A

Maßtheoretische Hilfsresultate

In diesem Anhang stellen wir einige Hilfsresultate zur Meßbarkeit von Mengen und Funktionen, darunter ein „Measurable Selection Theorem“ für mengenwertige Abbildungen, zusammen.

Satz A.1 Sei (Ω, \mathcal{A}) ein meßbarer Raum, $\Theta \subseteq \mathbb{R}^d$ offen und $G : \Omega \times \Theta \rightarrow \mathbb{R}^d$ eine Funktion, so daß

$$\forall \theta \in \Theta : G(\cdot; \theta) : \Omega \rightarrow \mathbb{R}^d \text{ ist } \mathcal{A}\text{-meßbar,} \quad (\text{A.1})$$

$$\forall \omega \in \Omega : G(\omega; \cdot) : \Theta \rightarrow \mathbb{R}^d \text{ ist stetig.} \quad (\text{A.2})$$

Dann ist G eine $\mathcal{A} \otimes \mathcal{B}(\Theta)$ -meßbare Funktion. Ist außerdem $C \subseteq \mathbb{R}^d$ abgeschlossen und $U \subseteq \Theta$ σ -kompakt (z.B. offen oder abgeschlossen), so gilt

$$\{\omega \in \Omega : \exists \theta \in U \text{ mit } G(\omega; \theta) \in C\} \in \mathcal{A}. \quad (\text{A.3})$$

Insbesondere sieht man für $C = \{0\}$, daß die Menge aller $\omega \in \Omega$, für die die Gleichung $G(\omega; \theta) = 0$ eine Lösung $\theta \in U$ hat, meßbar ist:

$$\{\omega \in \Omega : \exists \theta \in U \text{ mit } G(\omega; \theta) = 0\} \in \mathcal{A}. \quad (\text{A.4})$$

Beweis Zum Beweis der $\mathcal{A} \otimes \mathcal{B}(\Theta)$ -Meßbarkeit von G siehe z.B. Lemma 6.7.3 in [Pfa]. Sei nun $U \subseteq \Theta$ σ -kompakt. Für jedes feste $\vartheta \in \Theta$ und $r \in \mathbb{N}$ gilt

$$\{\omega \in \Omega : \text{dist}(G(\omega; \vartheta), C) < 1/r\} \in \mathcal{A}$$

aufgrund der Meßbarkeit von $G(\cdot; \vartheta)$ und der Stetigkeit von $y \mapsto \text{dist}(y, C)$. Ist nun $K \subseteq U$ kompakt, so überzeugt man sich mit Hilfe des Satzes von Bolzano-Weierstraß davon, daß aufgrund der Abgeschlossenheit von C sowie der Stetigkeit von $G(\omega; \cdot)$ und $y \mapsto \text{dist}(y, C)$

$$\{\omega \in \Omega : \exists \theta \in K \text{ mit } G(\omega; \theta) \in C\} = \bigcap_{r \in \mathbb{N}} \bigcup_{\vartheta \in K \cap \mathbb{Q}^d} \{\omega \in \Omega : \text{dist}(G(\omega; \vartheta), C) < 1/r\} \in \mathcal{A}.$$

Wählen wir nun eine kompakte Ausschöpfung $(K_m)_{m \in \mathbb{N}}$ von U , d.h. $K_m \subseteq U$ kompakt für alle m , $K_m \uparrow U$, so folgt

$$\{\omega : \exists \theta \in U \text{ mit } G(\omega; \theta) \in C\} = \bigcup_{m \in \mathbb{N}} \{\omega : \exists \theta \in K_m \text{ mit } G_n(\omega; \theta) \in C\} \in \mathcal{A}.$$

■

Satz A.2 (Measurable Selection Theorem) Sei (Ω, \mathcal{A}) ein meßbarer Raum und $F : \Omega \rightarrow \mathcal{P}(\mathbb{R}^d)$ eine mengenwertige Abbildung, so daß $F(\omega)$ für alle $\omega \in \Omega$ nichtleer und kompakt ist. Es gelte

$$\forall O \subseteq \mathbb{R}^d \text{ offen} : F^-(O) := \{\omega \in \Omega : F(\omega) \cap O \neq \emptyset\} \in \mathcal{A}. \quad (\text{A.5})$$

Dann kann man konstruktiv eine meßbare Funktion $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ angeben, so daß

$$\forall \omega \in \Omega : f(\omega) \in F(\omega).$$

Bemerkung Der Satz gilt wesentlich allgemeiner als hier dargestellt: so kann \mathbb{R}^d durch einen beliebigen vollständigen separablen metrischen Raum ersetzt werden, und es genügt, daß $F(\omega)$ für jedes $\omega \in \Omega$ abgeschlossen und nichtleer ist, siehe [CV], Theorem III.6. Die obige Formulierung reicht jedoch für unsere Zwecke aus und hat den Vorteil, daß sich eine „Auswahlfunktion“ f besonders explizit konstruieren läßt. Diese kann nämlich durch „sukzessive komponentenweise Minimierung“ der Menge $F(\omega)$ konstruiert werden, wie der nun folgende Beweis zeigt. ◆

Beweis von Satz A.2 Seien $p_j : \mathbb{R}^d \rightarrow \mathbb{R}$ die Projektionen auf die j -te Komponente. Für $\omega \in \Omega$ setze

$$f_1(\omega) := \inf\{x_1 : \text{es gibt Punkte } x = (x_1, x_2, \dots, x_d) \in F(\omega)\} = \inf p_1(F(\omega)).$$

Da $F(\omega)$ kompakt und p_1 stetig ist, so ist auch $p_1(F(\omega))$ kompakt, also gilt

$$f_1(\omega) = \min p_1(F(\omega)) \in p_1(F(\omega)),$$

d.h. es gibt in $F(\omega)$ Punkte der Bauart $(f_1(\omega), x_2, \dots, x_d)$. Betrachte nun

$$f_2(\omega) := \inf\{x_2 : \text{es gibt Punkte } x = (f_1(\omega), x_2, \dots, x_d) \in F(\omega)\} = \inf p_2(F(\omega) \cap p_1^{-1}(f_1(\omega))).$$

Wegen der Stetigkeit von p_1 ist $p_1^{-1}(f_1(\omega))$ abgeschlossen, also $F(\omega) \cap p_1^{-1}(f_1(\omega))$ kompakt, und wie oben folgt dann aus der Stetigkeit von p_2

$$f_2(\omega) = \min p_2(F(\omega) \cap p_1^{-1}(f_1(\omega))) \in F(\omega) \cap p_1^{-1}(f_1(\omega)),$$

d.h. es gibt in $F(\omega)$ Punkte der Bauart $(f_1(\omega), f_2(\omega), x_3, \dots, x_d)$. Führt man in derselben Weise fort, so erhält man nach d Schritten schließlich die Existenz eines Elements $(f_1(\omega), \dots, f_d(\omega)) \in F(\omega)$ mit

$$\begin{aligned} f_j(\omega) &= \min\{x_j : \text{es gibt Punkte } x = (f_1(\omega), \dots, f_{j-1}(\omega), x_j, \dots, x_d) \in F(\omega)\} \\ &= \min p_j (F(\omega) \cap p_1^{-1}(f_1(\omega)) \cap p_2^{-1}(f_2(\omega)) \cap \dots \cap p_{j-1}^{-1}(f_{j-1}(\omega))). \end{aligned} \quad (\text{A.6})$$

Insbesondere gilt für alle $j = 1, \dots, d$ und $b \in \mathbb{R}$

$$\begin{aligned} \{\omega \in \Omega : f_j(\omega) \leq b\} &= \{\omega : \exists x \in F(\omega) \cap p_1^{-1}(f_1(\omega)) \cap \dots \cap p_{j-1}^{-1}(f_{j-1}(\omega)) : p_j(x) \leq b\} \\ &= \{\omega : F(\omega) \cap p_1^{-1}(f_1(\omega)) \cap \dots \cap p_{j-1}^{-1}(f_{j-1}(\omega)) \cap I^{(j)} \neq \emptyset\}, \end{aligned} \quad (\text{A.7})$$

wobei $I^{(j)} := \{x = (x_1, \dots, x_d) \in \mathbb{R}^d : x_j \leq b\}$.

Setze nun

$$f(\omega) := (f_1(\omega), \dots, f_d(\omega)), \quad \omega \in \Omega.$$

Zu zeigen bleibt, daß f meßbar ist. Dazu reicht es zu zeigen, daß alle Komponenten f_j meßbar sind. Dies wird nun durch (unvollständige) Induktion nach j gezeigt. Für eine beliebige Menge $M \subseteq \mathbb{R}^d$ und $r \in \mathbb{N}$ definieren wir die offene Menge

$$U_{1/r}(M) := \{x \in \mathbb{R}^d : \text{dist}(x, M) < 1/r\}.$$

Sei $b \in \mathbb{R}$ beliebig. Dann haben wir für $j = 1$ aufgrund der Kompaktheit von $F(\omega)$

$$\{\omega : f_1(\omega) \leq b\} = \{\omega : F(\omega) \cap I^{(1)} \neq \emptyset\} = \bigcap_{r \in \mathbb{N}} \underbrace{\{\omega : F(\omega) \cap U_{1/r}(I^{(1)}) \neq \emptyset\}}_{\in \mathcal{A}} \in \mathcal{A} \quad (\text{A.8})$$

nach Voraussetzung (A.5) an F . Sei nun für ein $j > 1$ schon Meßbarkeit von f_1, \dots, f_{j-1} gezeigt. Dann ist aufgrund von (A.7) und der Kompaktheit von $F(\omega)$

$$\begin{aligned} &\{\omega : f_j(\omega) \leq b\} \\ &= \{\omega : F(\omega) \cap I^{(j)} \cap p_1^{-1}(f_1(\omega)) \cap p_2^{-1}(f_2(\omega)) \cap \dots \cap p_{j-1}^{-1}(f_{j-1}(\omega)) \neq \emptyset\} \\ &= \bigcap_{r \in \mathbb{N}} \{\omega : U_{1/r}(F(\omega)) \cap U_{1/r}(I^{(j)}) \cap p_1^{-1}(B_{1/r}(f_1(\omega))) \cap \dots \cap p_{j-1}^{-1}(B_{1/r}(f_{j-1}(\omega))) \neq \emptyset\}. \end{aligned} \quad (\text{A.9})$$

Nun ist aufgrund der Stetigkeit der Projektionen für jedes feste $r \in \mathbb{N}$ die Menge

$$U_{1/r}(F(\omega)) \cap U_{1/r}(I^{(j)}) \cap p_1^{-1}(B_{1/r}(f_1(\omega))) \cap \dots \cap p_{j-1}^{-1}(B_{1/r}(f_{j-1}(\omega)))$$

als Schnitt endlich vieler offener Mengen wiederum offen. Folglich gilt mit $D_{r,j} := \mathbb{Q}^d \cap U_{1/r}(I^{(j)})$

$$\begin{aligned}
& \{\omega : U_{1/r}(F(\omega)) \cap U_{1/r}(I^{(j)}) \cap p_1^{-1}(B_{1/r}(f_1(\omega))) \cap \cdots \cap p_{j-1}^{-1}(B_{1/r}(f_{j-1}(\omega))) \neq \emptyset\} \\
&= \bigcup_{u \in D_{r,j}} \{\omega : u \in U_{1/r}(F(\omega)) \cap p_1^{-1}(B_{1/r}(f_1(\omega))) \cap \cdots \cap p_{j-1}^{-1}(B_{1/r}(f_{j-1}(\omega)))\} \\
&= \bigcup_{u \in D_{r,j}} \{\omega : \text{dist}(u, F(\omega)) < 1/r, |u_1 - f_1(\omega)| < 1/r, \dots, |u_{j-1} - f_{j-1}(\omega)| < 1/r\} \\
&= \bigcup_{u \in D_{r,j}} \{\omega : \text{dist}(u, F(\omega)) < 1/r\} \cap \bigcap_{i=1}^{j-1} \{\omega : |u_i - f_i(\omega)| < 1/r\} \\
&= \bigcup_{u \in D_{r,j}} \underbrace{\{\omega : F(\omega) \cap B_{1/r}(u) \neq \emptyset\}}_{\in \mathcal{A} \text{ nach Vor.}} \cap \bigcap_{i=1}^{j-1} \underbrace{\{\omega : |u_i - f_i(\omega)| < 1/r\}}_{\in \mathcal{A} \text{ nach I.V.}}.
\end{aligned} \tag{A.10}$$

Wegen (A.9) ist damit die Behauptung gezeigt.

Literaturverzeichnis

- [AS] AITCHISON, J. UND SILVEY, S. D.: *Maximum-Likelihood Estimation of Parameters Subject to Restraints*, Annals of Mathematical Statistics 29 (1958), 813 - 828
- [Bat] BATEMAN, H. UND ERDÉLYI, A. (HRSG.): Higher Transcendental Functions I, New York: McGraw-Hill, 1953 (California Institute of Technology Bateman Manuscript Project)
- [Bill1] BILLINGSLEY, P.: *Statistical Inference for Markov Processes*, Chicago: University of Chicago Press, 1961
- [Bill2] BILLINGSLEY, P.: *The Lindeberg-Lévy Theorem for Martingales*, Proceedings of the AMS 12 (1961), 788 - 792
- [BS] BIBBY, B. M. UND SØRENSEN, M.: *Martingale Estimation Functions for Discretely Observed Diffusion Processes*, Bernoulli 1 (1995), 17 - 39
- [CIR] COX, J. C., INGERSOLL, J. E. UND ROSS, ST. A.: *A Theory of the Term Structure of Interest Rates*, Econometrica 53 (1985), 385 - 408
- [CV] CASTAING, C. UND VALADIER, M.: *Convex Analysis and Measurable Multifunctions*, Berlin: Springer-Verlag, 1977
- [Els] ELSTRODT, J.: *Maß- und Integrationstheorie*, 3. Aufl., Berlin: Springer-Verlag 2002
- [GH] GODAMBE, V. P. UND HEYDE, C. C.: *Quasi-Likelihood and Optimal Estimation*, International Statistical Review 55 (1987), 231 - 244
- [HH] HALL, P. UND HEYDE, C. C.: *Martingale Limit Theory and its Applications*, New York: Academic Press, 1980
- [Hö] HÖPFNER, REINHARD: Mathematische Statistik. Vorlesungsskript, Mainz 2004
- [IH] IBRAGIMOV, I. A. UND HAS'MINSKII, R. Z.: *Statistical Estimation. Asymptotic Theory*, New York: Springer-Verlag, 1981

- [IW] IKEDA, N. UND WATANABE, S.: *Stochastic Differential Equations and Diffusion Processes*,
2. Aufl., Amsterdam: North-Holland, 1989
- [Kab] KABALLO, W.: *Einführung in die Analysis III*, Heidelberg: Spektrum Akademischer Verlag, 1999
- [KN] KLIMKO, L. A. UND NELSON, P. I.: *On Conditional Least Squares Estimation for Stochastic Processes*, *Annals of Statistics* 6 (1978), 629 - 642
- [Leh] LEHMANN, E. L.: *Theory of Point Estimation*, New York: Wiley, 1983
- [LL] LIEB, E. H. UND LOSS, M.: *Analysis*, 2. Aufl., Providence: American Mathematical Society, 2001
- [MT] MEYN, S. UND TWEEDIE, R.: *Markov Chains and Stochastic Stability*, London: Springer-Verlag, 1994
- [OR] OVERBECK, L. UND RYDÉN, T.: *Estimation in the Cox-Ingersoll-Ross Model*, *Econometric Theory* 13 (1997), 430 - 461
- [Ove] OVERBECK, L.: *Estimation for Continuous Branching Processes*, *Scandinavian Journal of Statistics* 25 (1998), 111 - 126
- [Pfa] PFANZAGL, J.: *Parametric Statistical Theory*, Berlin: de Gruyter, 1994
- [Rev] REVUZ, D.: *Markov Chains*, Amsterdam: North-Holland, 1984
- [Sør] SØRENSEN, M.: *On Asymptotics of Estimating Functions*, *Brazilian Journal of Probability and Statistics* 13 (1999), 111 - 136. Preprintversion 1998, <http://www.maphysto.dk/publications/MPS-misc/1998/2.pdf>
- [Wef] WEFELMEYER, W.: *Quasi-Likelihood Models and Optimal Inference*, *Annals of Statistics* 24 (1996), 405 - 422