Schätzer in periodisch beobachteten autoregressiven Modellen

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Kim Kuen Tang

aus Rheinbach

(M & S, Köln)

2007

Berichterstatter: Prof. Dr. Wolfgang Wefelmeyer

Prof. Dr. Josef G. Steinebach

Tag der letzten mündlichen Prüfung: 31.10.2007

Zusammenfassung

Gegeben sei eine lineare autoregressive Zeitreihe, die wir nicht vollständig, sondern zu periodisch sich wiederholenden Zeitpunkten beobachten. Ausgehend von diesen Beobachtungen konstruieren wir Schätzer für den Autoregressionsparameter und die Innovationsdichte der voll beobachteten Zeitreihe. Der Schätzer für die Innovationsdichte verwendet einen Dekonvolutionsschätzer. Eine einfachere Variante wurde schon in einem sogenannten "eingeschränkten" Dekonvolutionsproblem für unabhängige Beobachtungen eingeführt. Ferner zeigen wir die lokale asymptotische Normalität der periodisch beobachteten Zeitreihe. Effiziente Schätzer für den Autoregressionsparameter werden charakterisiert. Wir konstruieren einen effizienten Schätzer unter einer zusätzlichen strukturellen Annahme. Dazu verwenden wir den Kleinste-Quadrate-Schätzer als Startschätzer und verbessern ihn mit dem Newton-Rhapson-Verfahren.

Abstract

Let a linear autoregressive Process be given. Assume further that we observe only some of the realizations, in a deterministic pattern that repeats itself periodically. Based on these observations we construct estimators for the autoregression parameter and for the innovation density. The estimator for the innovation density uses a so-called deconvolution estimator. A simple variation of the deconvolution estimator was already introduced in a restricted deconvolution problem for independent observations. Furthermore we show that the periodically observed autoregression process is locally asymptotically normal. Efficient estimators of the autoregression parameter will be characterized and constructed under an additional structural assumption. For this purpose we take the least squares estimator as the initial estimator and improve it by the Newton-Rhapson procedure.

Inhaltsverzeichnis

1 Einführung **1**

1.1 Modellbeschreibung 4

1.2 Notationen 8

1.3 Schätzer für den Regressionsparameter 8

2 Das Dekonvolutionsproblem **11**

2.1 Die Konvergenzrate bei bekanntem Regressionsparameter 17

2.2 Die Konvergenzrate bei unbekanntem Regressionsparameter 27

2.3 Das verallgemeinerte Dekonvolutionsproblem 39

2.4 Schätzer der Innovationsdichte in periodisch beobachteten heteroskedastischen nichtlinearen Regressionsmodellen 43

3 Lokale asymptotische Normalität **45**

3.1 Einführung und Notation 46

3.2 LAN für autoregressive Prozesse 46

3.3 LAN für periodisch beobachtete autoregressive Prozesse 48

4 Faltungssatz und Effizienz **56**

5 Effizientes Schätzen des autoregressiven Parameters **57**

5.1 Die Konstruktion des effizienten Schätzers unter einer zusätzlichen strukturellen Annahme 60

5.2 Die q sample splitting Methode 65

6 Ausblick **70**

6.1 Erweiterung auf die ARMA(p, q)-Reihe 70

6.2 Faltung von Dekonvolutionsschätzern 71

6.3 Konstruktion eines effizienten Schätzers ohne strukturelle Annahme . 73

Literatur **75**

1 Einführung

Mit dem Symbol $< x_i >_{i \in I}$ meinen wir eine Folge oder Menge von Elementen x_i mit Indexmenge I. Häufig geben wir I nicht an, falls wir die Indexmenge noch nicht festgelegt haben. Sei $< X_i >$ ein stationärer und reelle Markovkette der ersten Ordnung mit Übergangskern $P(x, dy)$ und invarianter Verteilung π. Wir bezeichnen mit $P_2(dx, dy)$ die gemeinsame Verteilung von zwei sukzessiven Realisationen. Die Markovkette wird vollständig beschrieben durch den Übergangskern. Dieser hat für messbare Menge A die Darstellung

$$P(dx, A) = P_2(dx, A)/\pi(dx).$$

Weiterhin gilt die Beziehung $\pi(A) = P_2(A, \mathbb{R})$. Die Markovkette ist also durch P_2 eindeutig bestimmt. Der empirische Schätzer für P_2 ist das Verhältnis der günstigen Paare von sukzessiven Beobachtungen zur Gesamtanzahl der Paare der sukzessiven Beobachtungen. Unter einem günstigen Paar von sukzessiven Beobachtungen verstehen wir ein Paar von zwei nachfolgende Realisationen (X_i, X_{i+1}). Das ist auch nötig, da das Paar nicht sukzessiver Beobachtungen wie (X_i, X_{i+1+r}) das Maß $\pi \otimes P \otimes P^{(r)}$ hat, was nicht identisch mit P_2 ist. Für messbare Mengen A und B ist der empirische Schätzer für $P_2(A \times B)$ demnach

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(X_{i-1} \in A, X_i \in B).$$

Dieser ist nach Greenwood & Wefelmeyer (1995) sogar effizient.

Im Weiteren nehmen wir an, dass wir nicht alle Realisationen $X_0, \ldots, X_n$ beobachten können. Wir diskutieren in dieser Arbeit nun, ob wir weiterhin den Übergangskern schätzen können. Die Aufgabenstellung hängt maßgeblich davon ab, welche Realisationen wir nicht beobachten können. Anders ausgedrückt hängt die Schätzbarkeit des Übergangskerns davon ab, ob in den verbleibenden Realisationen weiterhin Paare von sukzessiven Beobachtungen vorkommen. Ist dies der Fall, dann können wir weiterhin mit

$$\frac{1}{N} \sum_{j=1}^{N} \mathbf{1}(X_{\iota_j} \in A\,,\; X_{\iota_j+1} \in B)$$

den Wert $P_2(A \times B)$ schätzen. Hier sind $\iota_1, \ldots, \iota_N$ die Zeitpunkte gemeint, bei denen wir sukzessive Beobachtungen haben, und N ist die Anzahl der Paare von sukzessiven Beobachtungen. Ein Nachteil hat der obige Schätzer trotzdem. Dieser benutzt nicht die Informationen, die in den Lücken liegen. Dazu betrachten wir als Beispiel, das wir nur die Paare $< (X_{3j-1}, X_{3j+1}) >_j$ beobachten. Hier sehen wir jede dritte Realisation nicht. Der empirische Schätzer für eine integrierbare Funktion h mit $E|h(X_0, X_1)| < \infty$ ist

$$\frac{1}{n}\sum_{j=1}^{n} h(X_{3j-2}, X_{3j-1}).$$

Dieser benutzt nicht die Information, die in dem Paar (X_{3j-1}, X_{3j+1}) liegt. Um das zu demonstrieren, schreiben wir (X, Y, Z) für die Zufallsvariable $(X_{3j-1}, X_{3j}, X_{3j+1})$. Die beiden bedingten Erwartungswerte

$$h_r(X, Z) = E[h(X, Y) \,|\, (X, Z)] \text{ und } h_l(X, Z) = E[h(Y, Z) \,|\, (X, Z)]$$

haben denselben Erwartungswert wie $Eh(X_0, X_1)$. Also eignen sich die beiden Schätzer

$$\frac{1}{n}\sum_{j=1}^{n} h_l(X_{3j-1}, X_{3j+1}) \text{ und } \frac{1}{n}\sum_{j=1}^{n} h_r(X_{3j-1}, X_{3j+1})$$

genauso gut wie der empirische Schätzer

$$\frac{1}{n}\sum_{j=1}^{n} h(X_{3j-2}, X_{3j-1}).$$

Wie man die beiden Funktionen h_r und h_l schätzen kann, wird in Greenwood & Wefelmeyer (1995)[Observing two out of three] diskutiert. Wir gehen nicht weiter auf das Beispiel ein. Dieses soll demonstrieren, dass die Schätzbarkeit des Übergangskerns davon abhängt, ob man Paare von sukzessiven Beobachtungen hatten. Im nächsten Abschnitt setzen wir uns mit der Situation aus, bei der wir keine Paare von sukzessiven Beobachtungen haben.

In diesem Abschnitt richtet sich unser Hauptaugenmerk auf den Fall, in dem wir keine sukzessiven Beobachtungen haben. Dazu nehmen wir an, dass wir nur jede r-te Realisationen $< X_{rj} >_j$ haben. Unter dieser Tatsache können wir immer noch die Übergangsverteilung $P^{(r)}(x, dy)$ schätzen. Ob wir nun von $P^{(r)}(x, dy)$ auf $P(x, dy)$ schließen können, hängt davon ab, ob $P(x, dy)$ eindeutig durch $P^{(r)}(x, dy)$ bestimmt ist. Dass dies nicht der Fall ist, sieht man schon, wenn der Zustandsraum endlich ist. Die Übergangsverteilung der periodisch beobachteten Markovkette ist dann eine Matrix, die wir mit W bezeichnen. Die Gleichung $Q^r = W$ ist für gegebene r und W nicht eindeutig lösbar. Die Anzahl der Lösungen hängt von der Dimension der Matrix W

ab und wächst mit der Dimension. Da unser Zustandsraum nicht endlich ist, müssen wir mit einer beliebigen Anzahl von Lösungen rechnen. Im nichtparametrischen Fall kann man von $P^{(r)}(x, dy)$ nicht auf $P(x, dy)$ schlussfolgern. Die Gleichung $Q^{(r)} = P^{(r)}$ hat für festes $P^{(r)}$ beliebig viele Lösungen. Im parametrischen Fall ist die Frage durch den Maximum-Likelihood-Schätzer beantwortet.

Der semiparametrische Fall stellt das interessanteste Phänomen dar. Dazu nehmen wir an, dass die vollbeobachtete Markovkette die Gleichung

$$X_{i+1} = \phi_\alpha(X_i, \varepsilon_{i+1})$$

erfüllt. Hier ist das Symbol ϕ_α eine Funktion, die wir bis auf den Parameter $\alpha \in \Delta \subseteq \mathbb{R}^d$ kennen. Der nicht beobachtete stochastische Prozess $< \varepsilon_i >$ ist unabhängig und identisch verteilt. Die Varianz ist endlich, und die eindimensionale Verteilung hat Verteilungsfunktion F. Wir diskutieren weiter, wie wir das Parameter-Paar (α, F) schätzen können. In der Literatur wird häufig angenommen, dass für jedes Element $w \in \Delta$ genau eine Pseudo-Inverse ψ_w existiert mit

$$\psi_w(\phi_w(X_i, \varepsilon_{i+1}), X_i) = \varepsilon_{i+1}.$$

Die Gleichung ist auflösbar nach den Innovationen. Weiterhin wird vorausgesetzt, dass das Infimum von $\inf_{w\in\Delta} \operatorname{Var}(\psi_w(X_1, X_0))$ genau bei $w = \alpha$ angenommen wird. Beispiele für solche Markovkette stellen die ARMA-,ARCH- und die Bilinearen Prozesse, siehe Brockwell & Davis (1987), dar. Da die Gleichheit $\psi_w(\phi_\alpha(X_i, \varepsilon_{i+1}), X_i) = \psi_w(X_{i+1}, X_i)$ gilt, ist

$$\hat{\alpha} = \arg\min_{w\in\Delta} \frac{1}{n} \sum_{k=0}^{n} \left(\psi_w(X_{k+1}, X_k) - \frac{1}{n} \sum_{k=0}^{n} \psi_w(X_{k+1}, X_k) \right)^2$$

ein plausibler Schätzer. Unter zusätzlicher Annahme, siehe zum Beispiel van der Vaart (1998)[§5], ist $\hat{\alpha}$ sogar $\sqrt{n}$-konsistent. Weiterhin können wir mit $\psi_{\hat{\alpha}}(X_{i+1}, X_i)$ die Residuen von ε_{i+1} definieren. Unter geeigneter Voraussetzungen kann der Schätzer

$$\frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(\psi_{\hat{\alpha}}(X_{i+1}, X_i) \leq t)$$

für den Wert $F(t)$ sogar $\sqrt{n}-$konsistent sein.

Die oben beschriebene Methode eignet sich nur, falls die Gleichung nach den Innovationen auflösbar ist. Wir untersuchen nun die Problematik der Schätzbarkeit des Parameter-Paares (α, F), falls wir die Markovkette periodisch beobachten. Die periodisch beobachtete Zeitreihe $< X_{rn} >$ erfüllt die Gleichung

$$X_{r(n+1)} = \phi_{r,\alpha}(X_{rn}; \varepsilon_{rn+1}, \ldots, \varepsilon_{rn+r})$$

mit

$$\phi_{1,\alpha} = \phi_\alpha \text{ und } \phi_{r,\alpha}(x; y_1, \ldots, y_r) = \phi_\alpha(\phi_{r-1,\alpha}(x; y_1, \ldots, y_{r-1}), y_r).$$

Die Funktion $\phi_{r,\alpha}$ bildet von der Menge $\mathbb{R}^{r+1}$ nach $\mathbb{R}$ ab. Für die Inverse, falls sie existiert, muss dann $\psi_{r,\alpha}(X_{rn+r}, X_{rn}) = (\varepsilon_{rn+1}, \ldots, \varepsilon_{rn+r})$ gelten. Das kommt eher selten vor, da wir von zwei Elementen X_{rn+r}, X_{rn} auf r Elemente $\varepsilon_{rn+1}, \ldots, \varepsilon_{rn+r}$ schließen mussten. Wir brauchen aber nur eine Funktion $\psi_{W(\alpha)}$ mit

$$\psi_{W(\alpha)}(X_{rn+r}, X_{rn}) = G_\alpha(\varepsilon_{rn+1}, \ldots, \varepsilon_{rn+r})$$

und $G_\alpha : \mathbb{R}^r \to \mathbb{R}$. Ob wir nun α schätzen können, hängt davon ab, ob W injektiv ist.

Als Beispiel- und Untersuchungsmodell nehmen wir die Funktion $\phi_\alpha(x, y) = \alpha x + y$. Dann haben wir

$$\phi_{r,\alpha}(x, y_1 \ldots, y_r) = \phi_{\alpha^r}(x, G_\alpha(y_1 \ldots, y_r)),$$

$$G_\alpha(y_1 \ldots, y_r) = \sum_{k=0}^{r-1} y_{r-j}\alpha^j \text{ und } W(\alpha) = \alpha^r.$$

Durch die oben beschriebene Methode können wir den Parameter α^r schätzen. Ist r ungerade, dann ist sogar α schätzbar. Ist r gerade, dann können wir α bis auf das Vorzeichen schätzen. Der Grund liegt in der Tatsache, dass die Funktion $x \to x^r$ nicht auf allen Teilmengen von $\mathbb{R}$ injektiv ist. Ist die Funktion W auf Δ injektiv, dann ist α eindeutig identifizierbar. Im Weiteren werden wir den Kleinste-Quadrate-Schätzer für α vorstellen, dann werden wir einen Dekonvolutions-Schätzer für die Dichte von ε konstruieren.

1.1 Modellbeschreibung

Als Beispiel- und Untersuchungsmodell nehmen wir die einfache lineare Autoregression. Dabei genügen die Realisationen der Gleichung

$$X_{n+1} = \alpha X_n + \varepsilon_{n+1}$$

mit Regressionsparameter $|\alpha| < 1$, und die Innovationen $< \varepsilon_i >$ sind eine Folge von unabhängigen und identisch verteilten Zufallsvariablen. Die Charakterisierung des effizienten Schätzers setzt lokale asymptotische Normalität des Prozesses voraus. Hinreichend dafür ist, dass die Zufallsvariable ε eine absolut stetig differenzierbare

Dichte hat und die zugehörige Lageparameter-Familie endliche Fisher-Information besitzt. Darum setzen wir voraus, dass die Zufallsvariable ε eine differenzierbare Dichte f mit endlicher Fisher-Information

$$\int \left(\frac{f'}{f}\right)^2 f < \infty$$

hat. Hier ist $\int f = \int f(x)dx$. Bezeichnet $\mathcal{D}$ die Menge der Dichten mit endlicher Fisher-Information und Varianz, dann erfüllt der Parameterraum die Gleichung

$$\Theta = (-1, 1) \times \mathcal{D}.$$

Dieses Modell ist semiparametrisch, da der Parameterraum sowohl einen endlichen als auch einen unendlichen Parameter enthält. Zusätzlich zum Parameterraum führen wir den Regressionsraum $\mathcal{R}$ und Innovationsraum $\mathcal{I}$ ein. In diesem Modell ist $\mathcal{R} = (-1, 1)$ und $\mathcal{I} = \mathcal{D}$. Diese Unterteilung der Räume brauchen wir, falls wir nur jede r-te Realisation zur Verfügung haben. Die periodisch beobachtete Zeitreihe $< X_{rn} >_n$ für $r \in \mathbb{N}$, bei der wir nur jede r-te Realisation beobachten, genügt sie der Gleichung

$$X_{r(n+1)} = \alpha^r X_{rn} + \sum_{j=0}^{r-1} \alpha^j \varepsilon_{r(n+1)-j}. \qquad (T_r\text{-Modell})$$

Eine r-periodisch beobachtete AR(1)-Reihe ist wieder eine AR(1)-Reihe mit Autoregressionsparameter α^r und Innovations-Prozess $< \sum_{j=0}^{r-1} \alpha^j \varepsilon_{r(n+1)-j} >_n$. Da das Modell aus der teilbeobachteten linearen Autoregression stammt, nennen wir es T_r-Modell. Hier ist der Parameterraum Θ noch derselbe, und der Regressionsraum ist

$$\mathcal{R} = \{a^r \,:\, a \in (-1, 1)\} = \begin{cases} (-1, 1) & \text{für ungerade } r \\ (0, 1) & \text{für gerade } r. \end{cases}$$

Der Parameter im Regressionsraum kann man mit dem Kleinste-Quadrate-Schätzer schätzen. Das bewirkt, dass man für gerade r den Parameter α bis auf das Vorzeichen schätzen kann. Der Innovationsraum ist

$$\mathcal{I} = \{L(\alpha, f) \,:\, (\alpha, f) \in \Theta\}$$

mit

$$L(\alpha, f)(x) = \int f\left(x - \sum_{j=1}^{r-1} \alpha^j y_j\right) \prod_{k=1}^{r-1} f(y_k) dy_k.$$

Hier ist der Innovationsraum $\mathcal{I}$ eine echte Teilmenge des Grundraums $\mathcal{D}$. Der Grund liegt darin, dass eine Funktion aus dem Innovationsraum $\mathcal{I}$ mindestens r-mal stetig differenzierbar ist, während eine Funktion aus dem Grundraum $\mathcal{D}$ nur einmal stetig differenzierbar sein muss. Um zu verstehen, ob der Innovationsraum $\mathcal{I}$ mit der speziellen Struktur $L(\alpha, f)$ eine Auswirkung auf die Optimierung des Schätzens vom Parameter α hat, führen wir noch parallel das Obermodell

$$X_{r(n+1)} = \alpha^r X_{rn} + \eta_{n+1}$$

ein. Dabei ist $|\alpha| < 1$, und $< \eta_i >$ ist eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit endlicher Varianz und endlicher Fisher-Information. Hier ist der Regressionsraum $\mathcal{R} = (-1, 1)$ und der Innovationsraum $\mathcal{I} = \mathcal{D}$. Das Modell ist die allgemeine AR(1)-Zeitreihe mit Autoregressionsparameter α^r. Da die Innovationen keine strukturelle Annahme besitzen, nennen wir es K_r-Modell. Der Innovationsraum des K_r-Modells ist größer als der Innovationsraum des T_r-Modells. Also ist der effiziente Schätzer im T_r-Modell im Allgemeinen besser als der effiziente Schätzer im K_r-Modell. Für das Funktional $\kappa(\alpha, f) = \alpha$ werden wir Bedingungen angeben, unter denen die beiden effizienten Schätzer im T_r- und im K_r-Modell übereinstimmen. Dennoch müssen wir für das Funktional $\kappa(\alpha, f) = \alpha$ den Regressionsraum einschränken, da in dem T_r-Modell für gerade r keine Möglichkeit besteht, den Parameter α zu schätzen (Siehe Müller *et al.* (2007)[1.3 Linear Regression]). Wir werden daher $\Theta = (-1, 0) \times \mathcal{D}$ oder $\Theta = (0, 1) \times \mathcal{D}$ nehmen. Wir geben hier eine Tabelle an, um die Unterschiede zwischen dem T_r- und K_r-Modell zusammenzufassen.

	T_r-Modell mit Realisationen $< X_{rn} >_n$	K_r-Modell mit Realisationen $< X_{rn} >$
Definierende Gleichung	$X_{r(n+1)} = \alpha^r X_{rn} + \sum_{j=0}^{r-1} \alpha^j \varepsilon_{r(n+1)-j}$	$X_{r(n+1)} = \alpha^r X_n + \eta_{n+1}$
Innovationen	$\sum_{j=0}^{r-1} \alpha^j \varepsilon_{r(n+1)-j}$	η_{n+1}
Parameterraum Θ	$(0,1)\times\mathcal{D}$ oder $(-1,0)\times\mathcal{D}$	$(0,1)\times\mathcal{D}$ oder $(-1,0)\times\mathcal{D}$
Innovationsraum $\mathcal{I}$	$\{L(\alpha, f) : (\alpha, f) \in \Theta\}$	$\mathcal{D}$
Regressionsraum $\mathcal{R}$	$(-1, 0)$ oder $(0, 1)$	$(-1, 0)$ oder $(0, 1)$

Tabelle 1: Unterschiede des Parameterraumes zwischen den Modellen

Später werden wir die Tabelle um die Spalten Tangente und kanonischer Gradient erweitern.

Bemerkung 1.1 (uniforme Ergodizität)**.** Da für den Regressionsparameter $|\alpha| < 1$ gilt, und die Innovation endliche Varianz hat, hat die vollbeobachtete Markovkette

$< X_n >_n$ eine MA(∞)-Darstellung

$$X_n = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{n-j}.$$

Bezeichnen wir mit π die Verteilung von X_0, dann ist diese die invariante Verteilung der Markovkette. Da die Innovation eine Dichte f hat, hat die Markovkette eine Übergangsdichte $q(x,y) = f(y - \alpha x)$. Somit hat die invariante Verteilung auch eine Dichte $h(y) = \int f(y - \alpha x)\pi(dx)$. Die n-te Übergangsdichte genügt der Darstellung

$$q^{(n)}(x,y) = Ef\left(y - \alpha^n x - \sum_{j=1}^{n-1} \alpha^j \varepsilon_j\right).$$

Daraus folgt die Gleichung

$$\int |h(y) - q^{(n)}(x,y)|dy = \int \left| Ef\left(y - \alpha^n w - \sum_{j=1}^{n-1} \alpha^j \varepsilon_j\right) - f\left(y - \alpha^n x - \sum_{j=1}^{n-1} \alpha^j \varepsilon_j\right)\right| \\ h(w)dwdy.$$

Benutzen wir die Eigenschaft, dass die Dichte f stetig differenzierbar ist und die Darstellung

$$f(x+w) - f(x) = \int_0^1 f'(x + aw)daw$$

hat, dann gilt

$$\int |h(y) - q^{(n)}(x,y)|dy \le |\alpha^n| \left(\int |w|h(w)dw + |x|\right) \int |f'|(w)dw.$$

Da die Fisher-Information der Dichte f endlich ist, ist das Integral $\int |f'|(x)dx$ endlich. Weiterhin hat die invariante Verteilung endliche Varianz. Die rechte Seite der obigen Ungleichung ist somit endlich, und sie konvergiert mit der Rate $|\alpha^r|^n$ gegen Null. Da $|\alpha| < 1$ gilt, folgt nach Meyn & Tweedie (1993)[Theorem 16.0.2], dass die Markovkette $< X_n >_n$ uniform ergodisch ist. Wir werden annehmen, dass X_0 nach der invariante Verteilung verteilt ist. Somit ist der Prozess $< X_n >$ strikt stationär. Die Annahme ist keine Einschränkung, da die Markovkette uniform ergodisch ist.

Bemerkung 1.2 (ρ-mischend)**.** Die Markovkette $< X_n >_n$ ist strikt stationär, also können wir dazu die Autokovarianzfunktion definieren. Diese hat nach Brockwell & Davis (1987)[Theorem 4.4.2] eine Spektraldichte, welche rational ist. Mit Kolmogorov & Rozanov (1960) und Lin & Lu (1996)[Theorem 1.1.2] wissen wir, dass der Prozess $< X_n >$ ρ-mischend ist. Die ρ-mischenden Koeffizienten $< \rho(n) >$ genügen ferner der Ungleichung $\rho(n) \le \exp(-cn)$ für ein $c > 0$. Die Definition der ρ-mischenden Koeffizienten kann man in Lin & Lu (1996)[Definition 1.1.2] nachlesen.

1.2 Notationen

Für eine Folge von Maßen $< \mu_n >$ und eine Folge $< a_n >$ von reellen Zahlen bezeichnen wir mit $o_{\mu_n}(a_n)$ eine Folge von Zufallsvariablen ξ_n mit $\mu_n(|a_n^{-1}||\xi_n| > \varepsilon) \to 0$ für alle $\varepsilon > 0$. Entsprechend bezeichnen wir mit $O_{\mu_n}(a_n)$ eine Folge von Zufallsvariablen, die straff ist.

Mit $\mathcal{N}(\mu, \Sigma)$ bezeichnen wir die mehrdimensionale Normalverteilung mit Erwartungswert-Vektor μ und Kovarianz-Matrix Σ. Schreibe abkürzend $\mathcal{N}$ für $\mathcal{N}(0,1)$.

Für eine Zufallsvariable Ξ und ein Maß ν bezeichnen wir mit $\mathcal{L}(\Xi \,|\, \nu)$ dessen Verteilung unter dem Maß ν. Weiterhin benutzen wir $\Rightarrow$ für die schwache Konvergenz. Ferner bezeichnen wir mit λ meistens das Lebesguemaß, λ-f-s steht für die Aussage λ-fast-sicher.

Für zwei Zufallsvariablen X und Y benutzen wir das Symbol $\overset{\mathcal{D}}{=}$ für die Gleichheit in Verteilung.

Für zwei positive Zahlenfolge $< a_n >$ und $< b_n >$ führen wir das Symbol $\leq_*$ ein und schreiben $a_n \leq_* b_n$, falls $a_n/b_n = O(1)$ ist.

1.3 Schätzer für den Regressionsparameter

Wir geben nun den Schätzer für das Funktional $\kappa(\alpha, f) = \alpha$ an. Für eine natürliche Zahl r sei

$$\kappa_r(x_0, x_1, \ldots, x_n) = \left(\frac{\sum_{j=1}^n \hat{x}_j \hat{x}_{j-1}}{\sum_{j=1}^n \hat{x}_{j-1}^2} \right)^{1/r} \quad \text{mit} \;\; \hat{x}_j = x_j - \frac{1}{n} \sum_{k=1}^n x_k.$$

Im K_r-Modell ist $\hat{\kappa}_r = \kappa_r(X_0, X_r, \ldots, X_{rn})$ der Kleinste-Quadrate-Schätzer. Wir zeigen nun, dass der Schätzer $\hat{\kappa}_r$ $n^{1/2}$-konsistent ist.

Lemma 1.3. Seien $Y_n = X_{rn}$ die Realisationen einer periodisch beobachteten AR(1)-Zeitreihe. Wir haben dann

$$n^{1/2}\left(\hat{\kappa}_r - \alpha\right) = \frac{r\alpha^{r-1}}{\operatorname{Var}(Y)\sqrt{n}} \sum_{j=0}^n (\eta_{j+1} - E\eta)(Y_j - EY) + o_P(1).$$

Hier ist das Maß $P = P_{(\alpha,f)}$ gemeint. Hat die Innovation sogar viertes Moment, dann erhalten wir

$$P\left(|\hat{\kappa}_r^r - \alpha^r| > \delta_n\right) \leq_* n^{-1}\delta_n^{-2}$$

für alle Nullfolgen $< \delta_n >$. Speziell gilt

$$P\left(|\hat{\kappa}_r - \alpha| > \delta_n\right) \leq_* \frac{1}{n} + n^{-1}\delta_n^{-2},$$

und für ein $1 > \varsigma > 0$ haben wir

$$P\left(\left|\hat{\kappa}_r^k - \alpha^k\right| > \delta_n\right) \leq_* \frac{1}{n} + \frac{k^2 \varsigma^{2k-2}}{\delta_n^2 n}$$

für alle k aus $\mathbb{N}$.

Beweis. Seien $m = EY_0$ und $n \cdot \hat{m} = \sum_{j=1}^n Y_j$. Wir haben dann

$$\begin{aligned}
(\hat{\kappa}_r^r - \alpha^r)\left(\frac{1}{n}\sum_{j=1}^n \hat{Y}_j^2\right) =& n^{-1}\sum_{j=0}^n (\hat{Y}_{j+1} - \alpha^r \hat{Y}_j)\hat{Y}_j \\
=& \frac{1}{n}\sum_{j=0}^n (\eta_{j+1} - (1-\alpha^r)m)(Y_j - m) \\
&+ \frac{m - \hat{m}}{n}\sum_{j=0}^n (\eta_{j+1} - (1-\alpha^r)m).
\end{aligned}$$

Mit $\frac{1}{n}\sum_{j=1}^n \hat{Y}_j^2 = \frac{1}{n}\sum_{j=1}^n (Y_j - m)^2 - (\hat{m} - m)^2$ und Meyn & Tweedie (1993)[Theorem 17.1.7] erhalten wir

$$\frac{1}{n}\sum_{j=1}^n \hat{Y}_j^2 = \mathrm{Var}(Y) + o_P(1).$$

Da $\frac{m-\hat{m}}{n}\sum_{j=0}^n (\eta_{j+1} - (1-\alpha^r)m) = O_P(n^{-1})$, folgt

$$n^{1/2}(\hat{\kappa}_r^r - \alpha^r) = n^{1/2}\frac{1}{\mathrm{Var}(Y)n}\sum_{j=0}^n (\eta_{j+1} - E\eta)(Y_j - EY) + o_P(1).$$

Mit Brockwell & Davis (1987)[Proposition 6.1.5] erhalten wir dann die Behauptung der ersten Aussage.

Für allgemeine messbare Mengen A, B, C gilt die Ungleichung

$$P(A) \leq P(A \cap B \cap C) + P(B^c) + P(C^c).$$

Wählen wir nun $A = \{|\hat{\kappa}_r^r - \alpha^r| > \delta_n\}$, $B = \left\{\left|n^{-1}\sum_{j=1}^n \hat{Y}_j^2\right| > 4\,\mathrm{Var}(Y)\right\}$ und $C = \left\{\left|n^{-1}\sum_{j=0}^n (\hat{Y}_{j+1} - \alpha^r \hat{Y}_j)\hat{Y}_j\right| > 2\,\mathrm{Var}(Y)\delta_n\right\}$, dann verschwindet der Term $P(A \cap B \cap C)$. Für die Wahrscheinlichkeit $P(B^c)$ betrachten wir

$$\begin{aligned}
P\left(|n^{-1}\sum_{j=1}^n \hat{Y}_j^2| > 4\,\mathrm{Var}(Y)\right) \leq& P\left(|\frac{1}{n}\sum_{j=1}^n (Y_j - m)^2 - \mathrm{Var}(Y)| > \mathrm{Var}(Y)\right) \\
&+ P\left((m - \hat{m})^2 > 2\,\mathrm{Var}(Y)\right).
\end{aligned}$$

Da das vierte Moment endlich ist, können wir die rechte Seite durch die Varianz der Summe abschätzen. Mit Bemerkung 1.2 und Lin & Lu (1996)[Lemma 2.2.2] folgt die Aussage

$$P\left(\left|n^{-1}\sum_{j=1}^{n}\hat{Y}_j^2\right| > 4\,\mathrm{Var}(Y)\right) \leq_* n^{-1}.$$

Mit derselben Begründung folgt die Aussage auch für die Menge C^c, und es gilt

$$P\left(\left|n^{-1}\sum_{j=0}^{n}(\hat{Y}_{j+1}-\alpha^r\hat{Y}_j)\hat{Y}_j\right| > 4\,\mathrm{Var}(Y)\delta_n\right) \leq_* \frac{1}{\delta_n^2 n}.$$

Für die vorletzte Aussage benutzen wir die Gleichung

$$\hat{\kappa}_r - \alpha = \frac{\hat{\kappa}_r^r - \alpha^r}{r}\int_0^1 (\alpha^r + t(\hat{\kappa}_r^r - \alpha^r))^{1/r-1}dt.$$

Wir benutzen weiterhin die Ungleichung $P(A) \leq P(A \cap B) + P(B^c)$ für

$$A = \{|\hat{\kappa}_r - \alpha| > \delta_n\} \text{ und } B = \{|\hat{\kappa}_r^r - \alpha^r| < 1/2|\alpha^r|\}.$$

Damit können wir nun den Integranden abschätzen, und wir erhalten die Aussage.

Für die letzte Aussage, wählen wir ein $\iota > 0$ mit $|\alpha| + \iota < 1$. Dann gilt

$$\hat{\kappa}_r^k - \alpha^k = (\hat{\kappa}_r - \alpha)k\int_0^1 (\alpha + t(\hat{\kappa}_r - \alpha))^{k-1}dt$$

und

$$P\left(\left|\hat{\kappa}_r^k - \alpha^k\right| > \delta_n\right) \leq P\left(|\hat{\kappa}_r - \alpha| > \iota\right) + P\left(k(|\alpha|^r + \iota)^{k-1}|\hat{\kappa}_r - \alpha| > \delta_n\right).$$

□

Das Schätzen des Regressionsparameters ist nicht schwierig. Dafür reicht es, den Ausdruck $a \to E(Y_1 - a^r Y_0)^2$ zu minimieren. Dieser liefert den M-Schätzer $\hat{\kappa}_r^r$ (siehe van der Vaart (1998)[5.1 Introduction]). Der Schätzer $\hat{\kappa}_r^r$ ist auch bekannt als der Kleinste-Quadrate-Schätzer. Wir werden feststellen, dass der Kleinste-Quadrate-Schätzer nicht effizient ist. Dieser hat nicht die minimalste Varianz unter den regulären Schätzern. Wir können aber mit dem Kleinste-Quadrate-Schätzer einen effizienten Schätzer konstruieren, indem wir den Kleinste-Quadrate-Schätzer einen Schritt verbessern. Dieses Verfahren wird auch Newton-Raphson-Verfahren genannt und liefert im Allgemeinen auch einen effizienten Schätzer. Im nächsten Kapitel stellen wir Schätzer vor, mit denen wir die Dichte der Verteilung ε schätzen können. Diese werden wir brauchen, falls wir den effizienten Schätzer in einem ungünstigen Fall (siehe Kapitel 6.3, Fall F_3) konstruieren wollen.

2 Das Dekonvolutionsproblem

Beim Schätzen der Dichte f der Verteilung ε und deren Ableitungen in einem periodisch beobachteten AR(1)-Reihe (K_r-Modell) sind wir mit dem Problem konfrontiert, dass wir die Zufallsvariablen $< \varepsilon_i >$ nicht sehen und auch nicht annähern können. Das hat zur Folge, dass wir keinen Kernschätzer benutzen können. Die Existenz eines Schätzers für die Dichte f ist nicht gesichert. Darum zeigen wir hier, dass es möglich ist, die Dichte f zu schätzen.

Ein Problem besteht darin, dass wir keine Realisationen zur Verfügung haben, mit denen wir direkt die Parameterdichte f durch einen Kernschätzer schätzen können. Die Realisationen $< X_{rn} >_n$ genügen der folgenden Gleichheit in Verteilung:

$$X \stackrel{\mathcal{D}}{=} \varepsilon + \alpha X,$$

wobei ε unabhängig von X ist. Also haben wir

$$f_X(x) = \int_{\mathbb{R}} f_\varepsilon(x - \alpha y) f_X(y) dy,$$

wobei f_X die Dichte von X und $f_\varepsilon = f$ ist. Wir müssen nun versuchen, die Funktion f_X zu entfalten oder zu dekonvolieren. Der Lösungsansatz besteht darin, einen Schätzer $\hat{\phi}_\varepsilon$ für die charakteristische Funktion ϕ_ε zu suchen. Dann können wir mittels Rücktransformation von ϕ_ε (Fouriersynthese) einen Schätzer

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} (-it)^l \hat{\phi}_\varepsilon(t) dt$$

für $f^{(l)}(x)$ definieren. Der obige Schätzer ist aber nicht konsistent, da im Allgemeinen die Funktion $t \to E|\hat{\phi}_\varepsilon(t) - \phi_\varepsilon(t)|^2$ nicht integrierbar ist. Um das Problem zu lösen, wird das Verfahren der Kerndichteschätzung adaptiert. Wir integrieren die Fouriertransformation ϕ_K eines geeigneten Kerns K mit und erhalten als Schätzer

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} (-it)^l \phi_K(th_n) \hat{\phi}_\varepsilon(t) dt. \qquad (F\text{-Schätzer})$$

Die Bandbreite $< h_n >_n$ konvergiert gegen Null. Wir nennen diesen Schätzer den F-Schätzer. Der F-Schätzer hat den Vorteil, dass der Fehler

$$\begin{aligned} \hat{f}^{(l)}(x) - f^{(l)}(x) =& \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} (-it)^l \phi_K(th_n) \left(\hat{\phi}_\varepsilon(t) - \phi_\varepsilon(t) \right) dt \\ &+ \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} (-it)^l \phi_K(th_n) \phi_\varepsilon(t) dt - f^{(l)}(x) \end{aligned}$$

in zwei Terme zerfällt. Da das Integral $\int K(y)dy = 1$ gilt, vereinfacht sich der letzte Term zu

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}(-it)^l \phi_K(th_n)\phi_\varepsilon(t)dt - f^{(l)}(x) = \int_{\mathbb{R}} \left(f^{(l)}(x-h_n y) - f^{(l)}(x)\right) K(y)dy,$$

und dieser Fehler hängt nur von der Glätte der Dichte f ab. Hier sind wir erstmal daran interessiert, den Wert $f^{(l)}(x)$ für $l = 0, 1, \ldots$ zu schätzen. Darum nehmen wir im Weiteren an, dass die Dichte f im Punkt x für eine natürliche Zahl m m-mal stetig differenzierbar ist. Die m-te Ableitung $f^{(m)}$ ist im Punkt x Hölder-stetig

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x-y|^\delta$$

für alle $y \in \mathbb{R}$, und $0 < \delta < 1$. Da K eine Funktion ist, die wir wählen können, fordern wir nun, dass die Integrale

$$\int K(y)dy = 1\,, \int y^j K(y)dy = 0\,, \int \left|y^j K(y)\right| dy < \infty \text{ und } \int |y^{m+\delta}K(y)|dy < \infty$$

für $j = 1, 2, \ldots, m-1$ gelten. Mit den Eigenschaften eines Kerns können wir nun den letzten Term geschickt abschätzen, siehe Wefelmeyer (2007)[Satz 22].

Um den Fehler

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}(-it)^l \phi_K(th_n)\left(\hat{\phi}_\varepsilon(t) - \phi_\varepsilon(t)\right) dt$$

abschätzen zu können, brauchen wir Bedingungen an die Konvergenzgeschwindigkeit von $\phi_\varepsilon(t)$ für t gegen unendlich. Gilt

$$\int |t|^m |\phi_\varepsilon|(t)dt < \infty,$$

dann ist das hinreichend dafür, dass

$$f^{(l)}(x) = \frac{1}{2\pi}\int e^{-itx}(-it)^l \phi_\varepsilon(t)dt$$

für $l = 0, \ldots, m-1$ gilt, siehe Lukacs (1970)[Theorem 3.2.2]. Das hat zur Folge, dass die Funktion $\phi_\varepsilon(t)$ nicht zu langsam gegen Null konvergieren darf. Wir fordern deshalb, dass positive Konstanten G, d, μ existieren, so dass die Ungleichung

$$|\phi_\varepsilon|(t) \leq d|t|^{-1-\mu-m}$$

gilt.

Vollständigkeit halber erwähnen wir schon, dass die Bandbreite $< h_n >$ daran gekoppelt ist, wie schnell die Funktion ϕ_ε gegen Null konvergiert. Um eine breite Klasse von Dichten zu erwischen, fordern wir deshalb, dass es positive Konstanten β und γ mit der Eigenschaft

$$|\phi_\varepsilon(t)| \geq \exp(-|t|^\beta \gamma)$$

für $|t| > G$ existieren.

Nun diskutieren wir, wie wir die charakteristische Funktion ϕ_ε schätzen können. Um ϕ_ε zu schätzen, müssen wir ϕ_ε in ϕ_X und ϕ_η zerlegen, wobei ϕ_η die charakteristische Funktion von η ist. Die charakteristische Funktion ϕ_η können wir durch die Zufallsvariablen $\eta_{i+1} = X_{r(i+1)} - \alpha^r X_{ri}$ schätzen. Hat die charakteristische Funktion ϕ_ε nirgends eine Nullstelle, dann haben wir

$$\frac{\phi_\eta(t)}{\phi_\eta(\alpha t)} = \frac{\prod_{k=0}^{r-1} \phi_\varepsilon(\alpha^k t)}{\prod_{k=1}^{r} \phi_\varepsilon(\alpha^k t)} = \frac{\phi_\varepsilon(t)}{\phi_\varepsilon(\alpha^r t)}$$

und

$$\prod_{l=0}^{N-1} \frac{\phi_\eta(\alpha^{rl} t)}{\phi_\eta(\alpha^{rl+1} t)} = \prod_{l=0}^{N-1} \frac{\phi_\varepsilon(\alpha^{rl} t)}{\phi_\varepsilon(\alpha^{r(l+1)} t)} = \frac{\phi_\varepsilon(t)}{\phi_\varepsilon(\alpha^{rN} t)}.$$

Die erste geschlossene Darstellung für ϕ_ε ist

$$\phi_\varepsilon(t) = \frac{\phi_X(t)}{\phi_X(\alpha t)}$$

und die zweite ist

$$\phi_\varepsilon(t) = \prod_{l=0}^{N-1} \frac{\phi_\eta(\alpha^{rl} t)}{\phi_\eta(\alpha^{rl+1} t)} \frac{\phi_X(\alpha^{rN} t)}{\phi_X(\alpha^{rN+1} t)} \quad \text{für } N = 0, 1, \ldots.$$

Für bekannten Autoregressionsparameter ist der Schätzer für die Funktion $f^{(l)}(x)$ nun

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int e^{-itx} (-it)^l \phi_K(th_n) \hat{\phi}_\varepsilon(t) dt.$$

Hier sind

$$\hat{\phi}_\eta(t) = \frac{1}{n} \sum_{j=1}^{n} e^{it\eta_j}, \quad \hat{\phi}_X(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_{rj}},$$

$$\hat{\phi}_\varepsilon(t) = \prod_{l=0}^{N_n-1} \frac{\hat{\phi}_\eta(\alpha^{rl} t)}{\hat{\phi}_\eta(\alpha^{rl+1} t)} \frac{\hat{\phi}_X(\alpha^{rN_n} t)}{\hat{\phi}_X(\alpha^{rN_n+1} t)}. \qquad (\phi_\varepsilon\text{-Schätzer})$$

Aus technischen Gründen führen wir noch eine obere Beschränktheit für den Schätzer $\hat{\phi}_\varepsilon(t)$ ein und integrieren nur bis $1/h_n$. Für eine Konstante $M > 0$ nennen wir den Schätzer

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int_{|t| \leq 1/h_n} e^{-itx}(-it)^l \phi_K(th_n) \hat{\phi}_\varepsilon(t) \mathbf{1}_{|\hat{\phi}_\varepsilon(t)| \leq M} dt \qquad (V\text{-Schätzer})$$

den V-Schätzer-Schätzer. Ist α unbekannt, so definieren wir $\hat{\eta}_{i+1} = X_{r(i+1)} - \hat{\alpha}^r X_{ri}$,

$$\tilde{\phi}_\eta(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(it\hat{\eta}_j),$$

$$\tilde{\phi}_\varepsilon(t) = \prod_{l=0}^{N_n - 1} \frac{\tilde{\phi}_\eta(\hat{\alpha}^{rl} t)}{\tilde{\phi}_\eta(\hat{\alpha}^{rl+1} t)} \frac{\hat{\phi}_X(\hat{\alpha}^{rN_n} t)}{\hat{\phi}_X(\hat{\alpha}^{rN_n+1} t)}$$

für einen Schätzer $\hat{\alpha}$. Der Schätzer für $f^{(l)}(x)$, falls α unbekannt ist, ist

$$\tilde{f}^{(l)}(x) = \frac{1}{2\pi} \int_{|t| \leq 1/h_n} e^{-itx}(-it)^l \phi_K(th_n) \tilde{\phi}_\varepsilon(t) \mathbf{1}_{|\tilde{\phi}_\varepsilon(t)| \leq M} dt. \qquad (\tilde{V}\text{-Schätzer})$$

Wir zeigen nun im nächsten Kapitel, dass der Fehler $E|\hat{f}^{(l)} - f^{(l)}|(x)$ für den V-Schätzer mit der Rate $\log(\log(n))$ gegen Null konvergiert, falls N_n konstant ist. Insbesondere ist der primitive Schätzer

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int e^{-itx}(-it)^l \phi_K(th_n) \frac{\hat{\phi}_X(t)}{\hat{\phi}_X(\alpha t)} dt$$

schlecht. Lassen wir dagegen N_n gegen unendlich laufen, dann erhalten wir $E|\hat{f}^{(l)} - f^{(l)}|(x) = O(\log(n)^{-\delta})$ für ein $\delta > 0$.

Nun geben wir einige Lemmata an, die wir für die nächsten beiden Kapiteln benötigen.

Lemma 2.1. Sei $\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}$ die empirische charakteristische Funktion einer strikt stationären Zeitreihe $< X_j >$ mit ρ-mischenden Koeffizienten $\rho(n)$. Gilt $\rho(n) \leq_* e^{-cn}$, dann haben wir

$$P\left(\sup_{|t| \leq a} |\hat{\phi}(t) - \phi(t)| > b \right) \leq_* \frac{1}{n} + \frac{a}{b} e^{-d\sqrt{n^{1/2} b}}$$

und

$$\mathrm{Var}(\hat{\phi}(t)) \leq_* \frac{1}{n} \mathrm{Var}(e^{itX_1})$$

mit Konstanten a,b und $d > 0$. Die Konstanten sind nicht vom Symbol $\leq_*$ abhängig.

Bemerkung 2.2. Das obige Lemma wurde schon von Devroye (1994) für unabhängige und identisch verteilte Zufallsvariablen formuliert und bewiesen. Später wurde dann das Lemma in Belomestny (2003) weiter verschärft. Nun erweitern wir es auf ρ-mischende Zeitreihe. Wir wollen den Leser darauf aufmerksam machen, dass in Devroye (1994) die Abschätzung

$$|\hat{\phi}(t) - \hat{\phi}(s)| \leq |t-s| \left| \frac{1}{n} \sum_{k=0}^{n} Y_k \right|$$

benutzt wurde. Diese ist aber falsch, da für quadratintegrierbare und zentrierte Zufallsvariablen die rechte Seite für $n \to \infty$ P-fast-sicher gegen Null konvergiert, während die linke Seite gegen den Wert $|\phi(t) - \phi(s)|$ strebt. Der Fehler selbst hat nur einen technischen Effekt auf den Beweis und ist leicht korrigierbar. Die Aussage des Lemmas bleibt gleich.

Beweis. Sei $-a = t_{-k} < t_{-k+1} < \ldots < t_k = a$ mit $|t_i - t_{i+1}| < \gamma$ für ein $\gamma > 0$ und $0 < k < 1 + a/\gamma$. Sei $\beta_1 = E|X_0|$ das erste Moment. Wir werden $\gamma = b/(6\beta_1)$ wählen. Es gilt dann

$$\begin{aligned}
P\left(\sup_{|t|\leq a} |\hat{\phi}(t) - \phi(t)| > b\right) &\leq P\left(\bigcup_{i=-k}^{k} \{\sup_{t_i \leq t \leq t_{i+1}} |\hat{\phi}(t) - \phi(t)| > b\}\right) \\
&\leq P\left(\bigcup_{i=-k}^{k} \{\sup_{t_i \leq t \leq t_{i+1}} |\hat{\phi}(t) - \hat{\phi}(t_i)| > b/3\}\right) \\
&\quad + P\left(\bigcup_{i=-k}^{k} \{\sup_{t_i \leq t \leq t_{i+1}} |\hat{\phi}(t_i) - \phi(t_i)| > b/3\}\right) \\
&\quad + P\left(\bigcup_{i=-k}^{k} \{\sup_{t_i \leq t \leq t_{i+1}} |\phi(t_i) - \phi(t)| > b/3\}\right).
\end{aligned}$$

Nun müssen wir die letzten drei Terme abschätzen. Mit der Ungleichung

$$|\phi(t) - \phi(s)| \leq |t-s| \cdot |E(\int_0^1 \exp(isX + a(t-s)X)X da)| \leq |t-s|\beta_1$$

erhalten wir dann

$$\begin{aligned}
P\left(\bigcup_{i=-k}^{k} \{\sup_{t_i \leq t \leq t_{i+1}} |\phi(t_i) - \phi(t)| > b/3\}\right) \leq & P\left(\sup_{|t-s|<\gamma} |\phi(s) - \phi(t)| > b/3\right) \\
\leq & P\left(\gamma\beta_1 > b/3\right) = 0.
\end{aligned}$$

Für den ersten Term betrachte man für $|t-s| \leq \gamma$ die Ungleichung

$$|\hat{\phi}(t)-\hat{\phi}(s)| \leq \frac{\gamma}{n}\sum_{j=1}^{n}(|X_j|-\beta_1)+\gamma\beta_1.$$

Diese Ungleichung erhalten wir durch die Taylorentwicklung der Funktion

$$t \to \exp(itx).$$

Wir haben

$$\begin{aligned}
P\left(\bigcup_{i=-k}^{k}\{\sup_{t_i\leq t\leq t_{i+1}}|\hat{\phi}(t)-\hat{\phi}(t_i)|>b/3\}\right) \leq & P\left(\sup_{|t-s|\leq\gamma}|\hat{\phi}(t)-\hat{\phi}(s)|>b/3\right) \\
\leq & P\left(|\frac{\gamma}{n}\sum_{j=1}^{n}|X_j|-\beta_1|>(b/3-\beta_1\gamma)\right) \\
\leq & \operatorname{Var}\left(\frac{1}{n}\sum_{j=1}^{n}|X_j|-\beta_1\right)\left(\frac{\gamma}{b/3-\beta_1\gamma}\right)^2 \\
= & \frac{1}{n}\operatorname{Var}\left(\frac{1}{n}\sum_{j=1}^{n}|X_j|-\beta_1\right)\frac{n}{\beta_1}.
\end{aligned}$$

Per Definition ist die Zeitreihe $<|X_j|-\beta_1>$ zentriert und ρ-mischend mit demselben Koeffizient wie $< X_j >$. Mit Bradley (1999)[Theorem A] ist der Term

$$n\operatorname{Var}\left(\frac{1}{n}\sum_{j=1}^{n}|X_j|-\beta_1\right)$$

nach oben beschränkt, und es gilt

$$P\left(\bigcup_{i=-k}^{k}\{\sup_{t_i\leq t\leq t_{i+1}}|\hat{\phi}(t)-\hat{\phi}(t_i)|>b/3\}\right) \leq_* n^{-1}.$$

Für den zweiten Term haben wir

$$\begin{aligned}
P\left(\bigcup_{i=-k}^{k}\{\sup_{t_i\leq t\leq t_{i+1}}|\hat{\phi}(t_i)-\phi(t_i)|>b/3\}\right) \leq & \sum_{i=-k}^{k}P\left(|\Im\hat{\phi}(t_i)-\Im\phi(t_i)|>b/6\right) \\
& + \sum_{i=-k}^{k}P\left(|\Re\hat{\phi}(t_i)-\Re\phi(t_i)|>b/6\right),
\end{aligned}$$

wobei $\Im$ den Imaginärteil und $\Re$ den Realteil bezeichnet. Nach Doukhan *et al.* (1984) haben wir dann

$$
\begin{aligned}
P\left(\bigcup_{i=-k}^{k}\{\sup_{t_i\le t\le t_{i+1}}|\hat{\phi}(t_i)-\phi(t_i)|>b/3\}\right) \le& 2C(1+a/\gamma)\exp(-d\sqrt{n^{1/2}b})\\
\le& C_1\frac{a}{b}\exp(-d\sqrt{n^{1/2}b})
\end{aligned}
$$

Für die Varianz bemerken wir, dass die Ungleichung

$$
\operatorname{Var}(\hat{\phi}(t)) \le \frac{1}{n}\operatorname{Var}(e^{itX_1}) + \frac{2}{n^2}\sum_{1\le k<l\le n}\left|\operatorname{Cov}(e^{itX_k},e^{itX_l})\right|
$$

gilt. Mit Lin & Lu (1996)[Lemma 1.2.7] können wir den zweiten Term abschätzen und erhalten

$$
\frac{2}{n^2}\sum_{1\le k<l\le n}\left|\operatorname{Cov}(e^{itX_k},e^{itX_l})\right| \le \frac{2}{n^2}\sum_{1\le k<l\le n}4\operatorname{Var}(e^{itX_1})\sqrt{\rho(l-k)}.
$$

Es bleibt dann zu zeigen, dass die Reihe $n^{-1}\sum_{1\le k<l\le n}\sqrt{\rho(l-k)}$ für wachsende n beschränkt ist. Mit den Eigenschaften von $\rho(n)$ erhalten wir für ein $0<\delta<1$

$$
\begin{aligned}
\sum_{1\le k<l\le n}\sqrt{\rho(l-k)} \le \sum_{1\le k<l\le n}\delta^{l-k} =& \sum_{k=1}^{n}\delta^{-k}\sum_{l=k+1}^{n}\delta^{l}\\
=& \sum_{k=1}^{n}\delta^{-k}\delta^{k+1}\sum_{l=0}^{n-k-1}\delta^{l} = \delta\sum_{k=1}^{n}\frac{1-\delta^{n-k}}{1-\delta}\\
=& \frac{\delta}{1-\delta}\left(n-\frac{1-\delta^{n}}{1-\delta}\right).
\end{aligned}
$$

□

2.1 Die Konvergenzrate bei bekanntem Regressionsparameter

Damit wir die Darstellung

$$
\phi_\varepsilon(t) = \prod_{k=0}^{\infty}\frac{\phi_\eta(\alpha^{rk}t)}{\phi_\eta(\alpha^{rk+1}t)}
$$

benutzen können, müssen wir voraussetzen, dass $\phi_\varepsilon(t)\neq 0$ für alle t gilt.

Für die beiden Nullfolgen $< b_n >$ und $< h_n >$ definieren wir die Menge A_n und B_n durch

$$\begin{aligned}A_n =& \{ \sup_{|t|<1/h_n} |\hat{\phi}_\eta(t)/\phi_\eta(t)| < 1 + b_n\} \cap \{ \sup_{|t|<1/h_n} |\phi_\eta(t)/\hat{\phi}_\eta(t)| < 1 + b_n\} \\ &\cap \{ \sup_{|t|<1/h_n} |\hat{\phi}_X(\alpha^{rN_n}t)/\phi_X(\alpha^{rN_n}t)| < 1 + b_n\} \\ &\cap \{ \sup_{|t|<1/h_n} |\phi_X(\alpha^{rN_n}t)/\hat{\phi}_X(\alpha^{rN_n}t)| < 1 + b_n\}\end{aligned}$$

und

$$B_n = \{\tau_\eta > 1/h_n\} \cap \{\tau_X > 1/h_n\}.$$

Hier sind $\tau_\eta = \inf\{|t| : \hat{\phi}_\eta(t) = 0\}$ und $\tau_X = \inf\{|t| : \hat{\phi}_X(\alpha^{N_n}t) = 0\}$ die ersten Nullstellen der empirischen charakteristischen Funktionen.

Bemerkung 2.3. Auf der Menge A_n haben wir

$$\begin{aligned}|\hat{\phi}_\varepsilon|(t) = |\hat{\phi}_\varepsilon/\phi_\varepsilon| \cdot |\phi_\varepsilon|(t) =& \left| \prod_{l=0}^{N_n-1} \frac{\hat{\phi}_\eta(\alpha^{rl}t)\phi_\eta(\alpha^{rl+1}t)}{\phi_\eta(\alpha^{rl}t)\hat{\phi}_\eta(\alpha^{rl+1}t)} \frac{\hat{\phi}_X(\alpha^{rN_n}t)\phi_X(\alpha^{rN_n+1}t)}{\phi_X(\alpha^{rN_n}t)\hat{\phi}_X(\alpha^{rN_n+1}t)} \right| |\phi_\varepsilon(t)| \\ &\leq (1 + b_n)^{N_n} \leq M,\end{aligned}$$

falls die Folge $N_n b_n = O(1)$ beschränkt ist. Bei geeigneter Wahl von N_n und b_n verschwindet $\mathbf{1}_{|\hat{\phi}_\varepsilon|(t)<M}$ auf A_n nirgends.

Lemma 2.4. Für die Menge A_n haben wir

$$\begin{aligned}P(A_n^c) \leq_* & \frac{1}{n} + \frac{1/h_n}{b_n \inf_{|t|<1/h_n} |\phi_\eta(t)|} \exp\left(-d\sqrt{n^{1/2} b_n \inf_{|t|<1/h_n} |\phi_\eta(t)|} \right) \\ &+ \frac{1/h_n}{b_n \inf_{|t|<|\alpha|^{rN_n}1/h_n} |\phi_X(t)|} \exp\left(-d\sqrt{n^{1/2} b_n \inf_{|t|<|\alpha|^{rN_n}1/h_n} |\phi_X(t)|} \right)\end{aligned}$$

für ein $d > 0$.

Bemerkung 2.5. Lemma 2.4 liefert schon die erste Bedingung an die Bandbreite $< h_n >$. Konvergiert die Bandbreite zu schnell gegen Null, dann besteht die Gefahr, dass die Folge $n^{1/2} \inf_{|t|<1/h_n} |\phi_\eta(t)|$ auch gegen Null statt gegen Unendlich konvergiert.

Beweis. Es gilt

$$1 + b_n \leq \sup_{|t|<1/h_n} \left| \frac{\hat{\phi}_\eta(t)}{\phi_\eta(t)} \right| \leq 1 + \sup_{|t|<1/h_n} \left| \frac{\hat{\phi}_\eta(t) - \phi_\eta(t)}{\phi_\eta(t)} \right|.$$

Also haben wir

$$\left\{1 + b_n \leq \sup_{|t|<1/h_n} \left|\frac{\hat{\phi}_\eta(t)}{\phi_\eta(t)}\right|\right\} \subseteq \left\{\sup_{|t|<1/h_n} |\hat{\phi}_\eta(t) - \phi_\eta(t)| \geq b_n \inf_{|t|<1/h_n} |\phi_\eta(t)|\right\}.$$

Für die andere Menge betrachten wir

$$1 + b_n \leq \sup_{|t|<1/h_n} \left|\frac{\phi_\eta(t)}{\hat{\phi}_\eta(t)}\right| \leq 1 + \sup_{|t|<1/h_n} \left|\frac{\hat{\phi}_\eta(t) - \phi_\eta(t)}{\hat{\phi}_\eta(t)}\right|.$$

Diese können wir benutzen und liefert weiter

$$\begin{aligned}\sup_{|t|<1/h_n} |\hat{\phi}_\eta(t) - \phi_\eta(t)| &\geq b_n \inf_{|t|<1/h_n} |\hat{\phi}_\eta(t)| \\ &\geq b_n \inf_{|t|<1/h_n} |\phi_\eta(t)| - b_n \sup_{|t|<1/h_n} |\hat{\phi}_\eta(t) - \phi_\eta(t)|.\end{aligned}$$

Es folgt dann

$$\left\{1 + b_n \leq \sup_{|t|<1/h_n} \left|\frac{\phi_\eta(t)}{\hat{\phi}_\eta(t)}\right|\right\} \subseteq \left\{\sup_{|t|<1/h_n} |\hat{\phi}_\eta(t) - \phi_\eta(t)| \geq b_n \inf_{|t|<1/h_n} |\phi_\eta(t)|\right\}.$$

Da

$$\begin{aligned}P(A_n^c) \leq &P(\sup_{|t|<1/h_n} |\hat{\phi}_\eta(t)/\phi_\eta(t)| > 1 + b_n) + P(\sup_{|t|<1/h_n} |\phi_\eta(t)/\hat{\phi}_\eta(t)| > 1 + b_n) \\ &+ P(\sup_{|t|<|\alpha|^{rN_n}/h_n} |\hat{\phi}_X(t)/\phi_X(t)| > 1 + b_n) \\ &+ P(\sup_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)/\hat{\phi}_X(t)| > 1 + b_n),\end{aligned}$$

folgt dann die Aussage mit Lemma 2.1 und Bemerkung 1.2. □

Lemma 2.6. Seien $\tau_\eta = \inf\{|t| : \hat{\phi}_\eta(t) = 0\}$ und $\tau_X = \inf\{|t| : \hat{\phi}_X(\alpha^{N_n} t) = 0\}$ die beiden Stoppzeiten. Für jede beliebige Nullfolge $< h_n >$ gilt:

$$P(\tau_\eta \leq 1/h_n) \leq_* \frac{1}{n} + \frac{1/h_n}{\inf_{|t|\leq 1/h_n} |\phi_\eta|(t)} \exp\left(-d_1 \sqrt{n^{1/2} \inf_{|t|\leq 1/h_n} |\phi_\eta|(t)}\right)$$

und

$$P(\tau_X \leq 1/h_n) \leq_* \frac{1/h_n}{\inf_{|t|\leq 1/h_n} |\phi_X|(\alpha^{N_n} t)} \exp\left(-d_2 \sqrt{n^{1/2} \inf_{|t|\leq 1/h_n} |\phi_X|(\alpha^{N_n} t)}\right)$$

für positive Konstanten d_1 und d_2.

Beweis. Da $P(\tau_\eta \leq 1/h_n) \leq P\left(\inf_{|t|\leq 1/h_n} |\hat{\phi}_\eta(t)| = 0\right)$ gilt, haben wir

$$P(\tau_\eta \leq 1/h_n) \leq P\left(\sup_{|t|\leq 1/h_n} |\hat{\phi}_\eta - \phi_\eta|(t) \geq \inf_{|t|\leq 1/h_n} |\phi_\eta|(t)\right) \quad \text{und}$$

$$P(\tau_X \leq 1/h_n) \leq P\left(\sup_{|t|\leq 1/h_n} |\hat{\phi}_X - \phi_X|(\alpha^{N_n} t) \geq \inf_{|t|\leq 1/h_n} |\phi_X|(\alpha^{N_n} t)\right).$$

Mit Lemma 2.1 folgt dann die Aussage. □

Lemma 2.7. Ist $\hat{\phi}_\varepsilon(t)$ der ϕ_ε-Schätzer, dann haben wir

$$E\int_{|t|\leq 1/h_n} |t|^l |\hat{\phi}_\varepsilon - \phi_\varepsilon| \mathbf{1}_{A_n} \mathbf{1}_{B_n}(t) dt \leq_*$$
$$n^{-1/2}\left(\frac{\sum_{k=0}^{N_n-1}(1+b_n)^{2k}|\alpha|^{rk}}{\inf_{|t|\leq 1/h_n}|\phi_\eta(t)|} + \frac{|\alpha^{rN_n}|}{\inf_{|t|\leq |\alpha|^{rN_n}/h_n}|\phi_X(t)|}\right)\int_{|t|\leq 1/h_n} |t|^{l+1}|\phi_\varepsilon(t)| dt.$$

Beweis. Für $|t| \leq 1/h_n$ und mit der Gleichung

$$\frac{A}{B} - \frac{a}{b} = \frac{A-a}{b}\frac{b}{B} - \frac{a}{b}\frac{b}{B}(b-B)$$

erhalten wir

$$E\left|\frac{\hat{\phi}_\eta(\alpha^{rl}t)}{\hat{\phi}_\eta(\alpha^{rl+1}t)} - \frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right| \mathbf{1}_{C_n} \leq \left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right| \frac{E|\hat{\phi}_\eta(\alpha^{rl}t) - \phi_\eta(\alpha^{rl}t)|}{|\phi_\eta(\alpha^{rl}t)|}(1+b_n)$$
$$+ \left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right| \frac{E|\hat{\phi}_\eta(\alpha^{rl+1}t) - \phi_\eta(\alpha^{rl+1}t)|}{|\phi_\eta(\alpha^{rl+1}t)|}(1+b_n).$$

Hier ist $C_n = A_n \cap B_n$.

Da $n(E|\hat{\phi}_\eta(t) - \phi_\eta(t)|)^2 \leq 1 - |\phi_\eta(t)|^2$ und nach Ushakov (1999)[Theorem 2.3.2] auch $1 - |\phi_\eta(t)|^2 \leq \mathrm{Var}(\eta) \cdot |t|^2$ gilt, haben wir

$$E\left|\frac{\hat{\phi}_\eta(\alpha^{rl}t)}{\hat{\phi}_\eta(\alpha^{rl+1}t)} - \frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right| \mathbf{1}_{C_n} \leq 2\frac{|t\alpha^{rl}|}{n^{1/2}}\left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right| \frac{|(1+b_n)\alpha|\sqrt{\mathrm{Var}(\eta)}}{\inf_{|t|\leq 1/h_n}|\phi_\eta(t)|}.$$

Mit Lemma 2.1 und oben gemachten Überlegungen erhalten wir auch

$$E\left|\frac{\hat{\phi}_X(\alpha^{rN_n}t)}{\hat{\phi}_X(\alpha^{rN_n+1}t)} - \frac{\phi_X(\alpha^{rN_n}t)}{\phi_X(\alpha^{rN_n+1}t)}\right| \mathbf{1}_{C_n} \leq 2\frac{|t\alpha^{rN_n}|}{n^{1/2}}\left|\frac{\phi_X(\alpha^{rN_n}t)}{\phi_X(\alpha^{rN_n+1}t)}\right|$$
$$\cdot \frac{|(1+b_n)\alpha|\sqrt{\mathrm{Var}(X)}}{\inf_{|t|\leq |\alpha|^{rN_n}/h_n}|\phi_X(t)|}.$$

Das Resultat folgt nun mit der Ungleichung

$$\begin{aligned}\left|\prod_{k=0}^{N_n} E_k - \prod_{k=0}^{N_n} F_k\right| \mathbf{1}_{C_n} &\leq \sum_{k=0}^{N_n} \left|\prod_{i=0}^{k-1} E_i\right| |E_k - F_k| \left|\prod_{i=k+1}^{N_n} F_i\right| \mathbf{1}_{C_n} \\ &\leq \sum_{k=0}^{N_n} (1+b_n)^{2k} \left|\prod_{i=0}^{k-1} F_i\right| |E_k - F_k| \left|\prod_{i=k+1}^{N_n} F_i\right| \mathbf{1}_{C_n}.\end{aligned}$$

□

Wir fixieren nun ein Punkt x und formulieren die Rate des Fehlers $E|\hat{f}^{(l)} - f^{(l)}|(x)$.

Theorem 2.8. Die Dichte f sei im Punkt x m-mal stetig differenzierbar, und die Funktion $f^{(m)}$ sei im Punkt x Hölder-stetig mit

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x-y|^{\delta}$$

für ein $0 < \delta < 1$. Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d, β, γ, G und μ die folgende Eigenschaft

$$\exp(-|t|^{\beta}\gamma) \leq |\phi_\varepsilon|(t) \leq d|t|^{-m-1-\mu} \text{ für alle } |t| > G.$$

Weiterhin sei K ein Kern mit integrierbarer Fouriertransformation ϕ_K und den Eigenschaften

$$\int K(y)dy = 1\,, \int y^j K(y)dy = 0\,, \int \left|y^j K(y)\right| dy < \infty \text{ und } \int |y^{m+\delta} K(y)| dy < \infty$$

für $j = 1, 2, \ldots, m-1$. Wählen wir $N_n = \ln n$, $b_n = 1/\ln n$ und $h_n = (a(\ln n))^{-1/\beta}$ für eine positive Zahl a mit $(1-|\alpha|^{\beta})/(\gamma 2(1-|\alpha|^{r\beta})) > a$, dann gilt

$$\inf_{|t|<1/h_n} |\phi_\eta(t)| \geq n^{-a\gamma(1-|\alpha|^{\beta r})/(1-|\alpha|^{\beta})} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int_{|t|\leq 1/h_n} e^{-itx}(-it)^l \phi_K(th_n)\hat{\phi}_\varepsilon(t)\mathbf{1}_{|\hat{\phi}_\varepsilon|(t)<M} dt$$

der V-Schätzer, dann folgt

$$\begin{aligned}E|\hat{f}^{(l)} - f^{(l)}|(x) \leq_* &1/(nh_n^{l+1}) + 1/(n^{1/2} \inf_{|t|<1/h_n} |\phi_\eta(t)|) + h_n^{\mu+m-l} + h_n^{\delta+m-l} \\ =&O(1/(\ln n)^{(\min(\mu,\delta)+m-l)/\beta}).\end{aligned}$$

Bemerkung 2.9. Es ist nicht verwunderlich, dass die Rate des Fehlers $E|\hat{f}^{(l)} - f^{(l)}|(x)$ nicht polynomiell ist. Der Grund liegt in der Ungleichung

$$\exp(-|t|^{\beta}\gamma) \leq |\phi_{\varepsilon}|(t) \leq d|t|^{-m-1-\mu}.$$

Diese lässt auch charakteristische Funktionen zu, die hohe Volalitäten des Tails haben. Ist der Tail kontrollierbar, dann erreichen wir auch polynomielle Fehlerrate. Diesen Sachverhalt werden wir in den nächsten beiden Theoreme belegen.

Beweis. Die erste Aussage ist leicht zu zeigen. Die charakteristische Funktion ϕ_{η} genügt der Ungleichung

$$|\phi_{\eta}(t)| = \prod_{k=0}^{r-1} |\phi_{\varepsilon}(\alpha^{j}t)| \geq \exp(-\gamma|t|^{\beta}(1 - |\alpha|^{r\beta})/(1 - |\alpha|^{\beta})).$$

Da $|\alpha|^{rN_n}/h_n$ eine Nullfolge ist, folgt die erste Aussage.

Zur Übersichtlichkeit lassen wir den Punkt x weg und setzen $C_n = A_n \cap B_n$. Der Fehler ist beschränkt durch

$$\begin{aligned} E|\hat{f}^{(l)} - f^{(l)}| \leq \quad & E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \\ & + E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\leq 1/h_n} e^{-itx}(-it)^{l}\phi_K(th_n)\phi_{\varepsilon}(t)dt\right|\mathbf{1}_{C_n} \\ & + \left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^{l}\phi_K(th_n)\phi_{\varepsilon}(t)dt\right| \\ & + \left|\frac{1}{2\pi}\int e^{-itx}(-it)^{l}(\phi_K(th_n) - 1)\phi_{\varepsilon}(t)dt\right|. \end{aligned}$$

Wir schätzen nun die vier Terme ab.

Der Schätzer $\hat{f}^{(l)}$ ist weiterhin beschränkt durch

$$|\hat{f}^{(l)}| \leq M\int_{|t|\leq 1/h_n} |t|^{l}|\phi_K|(th_n)dt \leq_{*} 1/h_n^{l+1}.$$

Mit Lemma 2.4 und Lemma 2.6 erhalten wir für den ersten Term die Ungleichung

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \leq h_n^{-l-1}(P(A_n^c) + P(B_n^c)) \leq_{*} 1/(nh_n^{l+1}).$$

Das Lemma 2.7 liefert für den zweiten Term

$$E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\leq 1/h_n} e^{-itx}(-it)^{l}\phi_K(th_n)\phi_{\varepsilon}(t)dt\right|\mathbf{1}_{C_n} \leq n^{-1/2}/\inf_{|t|<1/h_n}|\phi_{\eta}(t)|.$$

Wir benutzen die Ungleichung $|\phi_\varepsilon|(t) \leq d|t|^{-m-1-\mu}$ für den dritten Term. Diese liefert

$$\left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^l \phi_K(th_n)\phi_\varepsilon(t)dt\right| \leq \frac{1}{2\pi}\int |K|(x)dx \cdot h_n^{\mu+m-l}$$

Letztes vereinfacht sich der vierte Term

$$\left|\frac{1}{2\pi}\int e^{-itx}(-it)^l(\phi_K(th_n)-1)\phi_\varepsilon(t)dt\right| = \left|\int (f^{(l)}(x) - f^{(l)}(x-yh_n))K(y)dy\right|.$$

Nach Wefelmeyer (2007)[Satz 22] haben wir dann

$$\left|\frac{1}{2\pi}\int e^{-itx}(-it)^l(\phi_K(th_n)-1)\phi_\varepsilon(t)dt\right| \leq_* h_n^{m-l+\delta}$$

□

Wir geben nun weitere Theoreme für charakteristische Funktionen mit kleineren Volatilitäten an. Die charakteristische Funktion soll entweder polynomiell oder exponentiell gegen Null konvergieren. Das nächste Theorem ist für charakteristische Funktion mit polynomieller Rate gedacht.

Theorem 2.10 (Die Fouriertransformation der Dichte konvergiert mit polynomieller Rate.)**.** Die Dichte f sei im Punkt x m-mal stetig differenzierbar, und die Funktion $f^{(m)}$ sei im Punkt x Hölder-stetig mit

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x-y|^\delta$$

für ein $0 < \delta < 1$. Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d_0, d_1, ν, G und μ die folgende Eigenschaft

$$d_0|t|^{-\nu} \leq |\phi_\varepsilon|(t) \leq d_1|t|^{-m-1-\mu} \text{ für alle } |t| > G.$$

Weiterhin sei K ein Kern mit integrierbarer Fouriertransformation ϕ_K und den Eigenschaften

$$\int K(y)dy = 1\,, \int y^j K(y)dy = 0\,, \int \left|y^j K(y)\right| dy < \infty \text{ und } \int |y^{m+\delta}K(y)|dy < \infty$$

für $j = 1, 2, \ldots, m-1$. Wählen wir $N_n = (\ln n)^2$, $b_n = (1/\ln n)^2$ und $h_n = n^{-a}$ für eine Zahl a mit $1/(2r\nu) > a > 0$, dann gilt

$$\inf_{|t|<1/h_n} |\phi_\eta(t)| \geq h_n^{r\nu} = n^{-ar\nu} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist

$$\hat{f}^{(l)}(x) = \frac{1}{2\pi} \int_{|t|\le 1/h_n} e^{-itx}(-it)^l \phi_K(th_n)\hat{\phi}_\varepsilon(t)\mathbf{1}_{|\hat{\phi}_\varepsilon|(t)<M} dt$$

der V-Schätzer, dann folgt

$$\begin{aligned} E|\hat{f}^{(l)} - f^{(l)}|(x) \le_* & 1/(nh_n^{l+1}) + 1/(n^{1/2} \inf_{|t|<1/h_n} |\phi_\eta(t)|) + h_n^{\mu+m-l} + h_n^{\delta+m-l} \\ = & n^{a(l+1)-1} + n^{ar\nu-1/2} + n^{-a(\mu+m-l)} + n^{-a(\delta+m-l)}. \end{aligned}$$

Beweis. Die erste Aussage folgt mit der Voraussetzung.

Wie beim Beweis des Theorems 2.8 lassen wir den Punkt x weg und setzen $C_n = A_n \cap B_n$. Der Fehler ist beschränkt durch

$$\begin{aligned} E|\hat{f}^{(l)} - f^{(l)}| \le \quad & E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \\ & + E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\le 1/h_n} e^{-itx}(-it)^l\phi_K(th_n)\phi_\varepsilon(t)dt\right| \mathbf{1}_{C_n} \\ & + \left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^l\phi_K(th_n)\phi_\varepsilon(t)dt\right| \\ & + \left|\frac{1}{2\pi}\int e^{-itx}(-it)^l(\phi_K(th_n)-1)\phi_\varepsilon(t)dt\right|. \end{aligned}$$

Wir schätzen nun die vier Terme ab.

Der Schätzer $\hat{f}^{(l)}$ ist weiterhin beschränkt durch

$$|\hat{f}^{(l)}| \le M \int_{|t|\le 1/h_n} |t|^l |\phi_K|(th_n)dt \le_* 1/h_n^{l+1}.$$

Mit Lemma 2.4 und Lemma 2.6 erhalten wir für den ersten Term die Ungleichung

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \le h_n^{-l-1}(P(A_n^c) + P(B_n^c)) \le_* 1/(nh_n^{l+1}).$$

Das Lemma 2.7 liefert für den zweiten Term

$$E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\le 1/h_n} e^{-itx}(-it)^l\phi_K(th_n)\phi_\varepsilon(t)dt\right| \mathbf{1}_{C_n} \le n^{-1/2}/\inf_{|t|<1/h_n}|\phi_\eta(t)|.$$

Wir benutzen die Ungleichung $|\phi_\varepsilon|(t) \le d|t|^{-m-1-\mu}$ für den dritten Term. Diese liefert

$$\left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^l\phi_K(th_n)\phi_\varepsilon(t)dt\right| \le \frac{1}{2\pi}\int |K|(x)dx\cdot h_n^{\mu+m-l}.$$

Letztes vereinfacht sich der vierte Term

$$\left|\frac{1}{2\pi}\int e^{-itx}(-it)^l(\phi_K(th_n)-1)\phi_\varepsilon(t)dt\right| = \left|\int (f^{(l)}(x)-f^{(l)}(x-yh_n))K(y)dy\right|.$$

Nach Wefelmeyer (2007)[Satz 22] haben wir dann

$$\left|\frac{1}{2\pi}\int e^{-itx}(-it)^l(\phi_K(th_n)-1)\phi_\varepsilon(t)dt\right| \le_* h_n^{m-l+\delta}$$

□

Nun kommt ein Theorem für charakteristische Funktion mit exponentieller Rate.

Theorem 2.11 (Die Fouriertransformation der Dichte konvergiert mit exponentieller Rate.)**.** Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d, β, γ_1, G und γ_2 die folgende Eigenschaft

$$\exp(-|t|^\beta\gamma_1) \le |\phi_\varepsilon|(t) \le d\exp(-|t|^\beta\gamma_2) \text{ für } |t| > G.$$

Wählen wir $N_n = \ln n$, $b_n = 1/\ln n$ und $h_n = 1/(a\ln n)^{1/\beta}$ für eine positive Zahl a mit $(1-|\alpha|^\beta)/(2\gamma_1(1-|\alpha|^{r\beta})) > a$, dann gilt

$$\inf_{|t|<1/h_n} |\phi_\eta(t)| \ge n^{-a\gamma_1(1-|\alpha|^{r\beta})/(1-|\alpha|^\beta)} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist $\hat\phi_\varepsilon$ der ϕ_ε-Schätzer und

$$\hat f^{(l)}(x) = \frac{1}{2\pi}\int_{|t|\le 1/h_n} e^{-itx}(-it)^l\hat\phi_\varepsilon(t)\mathbf{1}_{|\hat\phi_\varepsilon|(t)<M}dt,$$

dann folgt

$$\begin{aligned} E|\hat f^{(l)} - f^{(l)}|(x) &\le_* \frac{1}{nh_n^{l+1}} + n^{a\gamma_1(1-|\alpha|^{r\beta})/(1-|\alpha|^\beta)-1/2} + n^{-a\gamma_2}\\ &\le_* n^{a\gamma_1(1-|\alpha|^{r\beta})/(1-|\alpha|^\beta)-1/2} + n^{-a\gamma_2}. \end{aligned}$$

Aus der Voraussetzung folgt, dass die Funktion $|\phi_\varepsilon(t)||t|^l$ für alle l integrierbar ist. Die Dichte f ist daher beliebig oft differenzierbar. Aus diesem Grund sieht der Schätzer $\hat{f}^{(l)}$ anders als die anderen aus. Die Integration mit der Fouriertransformation ϕ_K würde die Fehler-Rate nur verlangsamen.

Beweis. Die erste Aussage ist klar.

Wie beim Beweis des Theorems 2.8 lassen wir den Punkt x weg und setzen $C_n = A_n \cap B_n$. Der Fehler ist beschränkt durch

$$\begin{aligned} E|\hat{f}^{(l)} - f^{(l)}| \leq \quad & E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \\ & + E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\leq 1/h_n} e^{-itx}(-it)^l\phi_\varepsilon(t)dt\right|\mathbf{1}_{C_n} \\ & + \left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^l\phi_\varepsilon(t)dt\right|. \end{aligned}$$

Wir schätzen nun die drei Terme ab.

Der Schätzer $\hat{f}^{(l)}$ ist weiterhin beschränkt durch

$$|\hat{f}^{(l)}| \leq M\int_{|t|\leq 1/h_n} |t|^l|\phi_K|(th_n)dt \leq_* 1/h_n^{l+1}.$$

Mit Lemma 2.4 und Lemma 2.6 erhalten wir für den ersten Term die Ungleichung

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{C_n^c} \leq h_n^{-l-1}(P(A_n^c) + P(B_n^c)) \leq_* 1/(nh_n^{l+1}).$$

Das Lemma 2.7 liefert für den zweiten Term

$$E\left|\hat{f}^{(l)} - \frac{1}{2\pi}\int_{|t|\leq 1/h_n} e^{-itx}(-it)^l\phi_K(th_n)\phi_\varepsilon(t)dt\right|\mathbf{1}_{C_n} \leq n^{-1/2}/\inf_{|t|<1/h_n}|\phi_\eta(t)|.$$

Wir benutzen die Ungleichung $|\phi_\varepsilon|(t) \leq d\exp(-|t|^\beta\gamma_2)$ für den dritten Term. Diese liefert

$$\begin{aligned} \left|\frac{1}{2\pi}\int_{|t|>1/h_n} e^{-itx}(-it)^l\phi_\varepsilon(t)dt\right| \leq_* & \int_{t>1/h_n} \exp(-t^\beta\gamma_2)\beta t^{\beta-1}\gamma_2 dt \\ & = \exp(-h_n^{-\beta}\gamma_2) = n^{-a\gamma_2}. \end{aligned}$$

□

Im Weiteren werden wir die Konvergenzrate des Fehlers $E|\hat{f}^{(l)} - f^{(l)}|(x)$ angeben, falls α unbekannt ist.

2.2 Die Konvergenzrate bei unbekanntem Regressionsparameter

In diesem Kapitel sei $< X_n >$ eine AR(1)-Zeitreihe mit Innovation-Prozess $< \varepsilon_i >_i$ und unbekanntem Regressionsparameter α. Ferner hat die Zufallsvariable $\varepsilon = \varepsilon_0$ eine Dichte f. Die zugehörige Lageparameter-Familie besitzt endliche Fisher-Information. Weiterhin beobachten wir periodisch nur jede r-te Realisation X_{rn}. Wir präsentieren Schätzer für die Dichte f, die auf die Realisationen $X_0, X_r, \ldots, X_{rn}$ basieren. Hier ersetzen wir α^r durch den Kleinste-Quadrate-Schätzer $\hat{\kappa}_r^r$. Da das Lemma 1.3 die Existenz des vierten Momentes der Innovationen voraussetzt, nehmen wir nun an, dass die Parameterdichte f ein endliches viertes Moment hat. Wir werden nun die Konvergenzrate des Fehlers von $E|\tilde{f}^{(l)} - f^{(l)}|(x)$ angeben, wobei $\tilde{f}^{(l)}$ der $\tilde{V}$-Schätzer ist.

Da der Regressionsparameter α nicht bekannt ist, kennen wir die Zufallsvariablen

$$\eta_{i+1} = X_{r(i+1)} - \alpha^r X_{ri} \text{ für } i = 0, 1 \ldots,$$

nicht. Wir definieren nun die Residuen $\hat{\eta}_{i+1}$ durch

$$\hat{\eta}_{i+1} = X_{r(i+1)} - \hat{\kappa}_r^r X_{ri} \text{ für } i = 0, 1 \ldots$$

und die zugehörige empirische charakteristische Funktion durch

$$\tilde{\phi}_\eta(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(it\hat{\eta}_i).$$

Wieder für zwei Nullfolge $< b_n >$ und $< h_n >$ definieren wir die Menge A_n und B_n durch

$$\begin{aligned} A_n =& \{ \sup_{|t|<1/h_n} |\hat{\phi}_{\hat{\eta}}(t)/\phi_\eta(t)| < 1 + b_n \} \cap \{ \sup_{|t|<1/h_n} |\phi_\eta(t)/\hat{\phi}_{\hat{\eta}}(t)| < 1 + b_n \} \\ & \cap \{ \sup_{|t|<1/h_n} |\hat{\phi}_X(\hat{\kappa}_r^{rN_n} t)/\phi_X(\alpha^{rN_n} t)| < 1 + b_n \} \\ & \cap \{ \sup_{|t|<1/h_n} |\phi_X(\alpha^{rN_n} t)/\hat{\phi}_X(\hat{\kappa}_r^{rN_n} t)| < 1 + b_n \} \end{aligned}$$

und

$$B_n = \{ \tau_{\hat{\eta}} \geq 1/h_n \} \cap \{ \tau_{\hat{X}} \geq 1/h_n \}.$$

Hier ist $\tau_{\hat{\eta}}$ durch $\tilde{\phi}_\eta$ und $\tau_{\hat{X}}$ durch $\hat{\phi}_X(\hat{\kappa}_r^{rN_n} t)$ definiert.

Die Menge A_n sichert, dass die empirische charakteristische Funktion in der Nähe der charakteristischen Funktion bleibt. Die Menge B_n hindert uns daran, durch Null zu teilen. Die beide Mengen reichen aber trotzdem noch nicht aus, um den Fehler abzuschätzen, da wir nicht in der Lage sind, den Wert $E\hat{\kappa}_r^r$ auszurechnen. Aus diesem Grund führen wir nun die dritte Menge C_n ein mit

$$\begin{aligned}C_n =& \bigcap_{k=0}^{N_n}\{\sup_{|t|\le 1/h_n} |\phi_\eta(\hat{\kappa}_r^k t)/\phi_\eta(\alpha^k t)| < 1 + b_n\} \bigcap_{k=0}^{N_n}\{\sup_{|t|\le 1/h_n} |\phi_\eta(\alpha^k t)/\phi_\eta(\hat{\kappa}_r^k t)| < 1 + b_n\} \\ & \bigcap\{\sup_{|t|\le 1/h_n} |\phi_X(\hat{\kappa}_r^{rN_n} t)/\phi_X(\alpha^{rN_n} t)| < 1 + b_n\} \\ & \bigcap\{\sup_{|t|\le 1/h_n} |\phi_X(\alpha^{rN_n} t)/\phi_X(\hat{\kappa}_r^{rN_n} t)| < 1 + b_n\}.\end{aligned}$$

Diese sichert, dass die Funktion $\phi_\eta(\hat{\kappa}_r^k t)$ in einer Umgebung von $\phi_\eta(\alpha^k t)$ bleibt. Mit der Ungleichung

$$\begin{aligned}\mathbf{1} &= (\mathbf{1}_{A_n} + \mathbf{1}_{A_n^c})(\mathbf{1}_{B_n} + \mathbf{1}_{B_n^c})(\mathbf{1}_{C_n} + \mathbf{1}_{C_n^c}) \\ &\le \mathbf{1}_{A_n}\mathbf{1}_{B_n}\mathbf{1}_{C_n} + 4\mathbf{1}_{A_n^c} + 4\mathbf{1}_{B_n^c} + 4\mathbf{1}_{C_n^c},\end{aligned}$$

können wir den Fehler $E|\hat{f}^{(l)} - f^{(l)}|(x)$ abschätzen durch

$$\begin{aligned}E|\hat{f}^{(l)} - f^{(l)}|(x) \le& E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{A_n}\mathbf{1}_{B_n}\mathbf{1}_{C_n}(x) \\ & + 4E|\hat{f}^{(l)} - f^{(l)}|(\mathbf{1}_{A_n^c} + \mathbf{1}_{B_n^c} + \mathbf{1}_{C_n^c})(x).\end{aligned}$$

Nun geben wir Lemmata an, um die beiden Terme auf der rechten Seite der Ungleichung abzuschätzen.

Bemerkung 2.12. Mit einer Taylorentwicklung haben wir

$$\tilde{\phi}_\eta(t) = \phi_\eta(t) + \hat{\phi}_\eta(t) - \phi_\eta(t) + R_n(t)$$

mit Rest $|R_n(t)| \le |t||\hat{\kappa}_r^r - \alpha^r|\frac{1}{n}\sum_{j=1}^n |X_{r(j-1)}|$.

Für die andere charakteristische Funktion haben wir auch

$$\hat{\phi}_X(\hat{\kappa}_r^{rN_n} t) = \phi_X(\alpha^{rN_n} t) + \hat{\phi}_X(\alpha^{rN_n} t) - \phi_X(\alpha^{rN_n} t) + R_n(t)$$

mit Rest $|R_n(t)| \le |t||\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}|\frac{1}{n}\sum_{j=1}^n |X_{rj}|$.

Lemma 2.13. Für ein $0 < \varsigma < 1$ haben wir, dass

$$\begin{aligned}P(\tau_{\hat{X}} < 1/h_n) \le_* & \frac{1}{n} + \frac{1/h_n}{\inf_{|t|\le 1/h_n} |\phi_X(\alpha^{rN_n} t)|} \exp\left(-d\sqrt{n^{1/2} \inf_{|t|\le 1/h_n} |\phi_X(\alpha^{rN_n} t)|}\right) \\ & \frac{1/h_n^2}{\inf_{|t|\le 1/h_n} |\phi_X(\alpha^{rN_n} t)|^2}\varsigma^{N_n}\frac{N_n}{n} + \frac{1/h_n}{n\inf_{|t|\le 1/h_n} |\phi_X(\alpha^{rN_n} t)|}\end{aligned}$$

und

$$P(\tau_{\hat{\eta}} < 1/h_n) \leq_* \frac{1}{n} + \frac{1/h_n}{\inf_{|t|\leq 1/h_n} |\phi_\eta(t)|} \exp\left(-d\sqrt{n^{1/2} \inf_{|t|\leq 1/h_n} |\phi_\eta(t)|} \right)$$
$$+ \frac{1/h_n}{\inf_{|t|\leq 1/h_n} |\phi_\eta(t)| n}.$$

Insbesondere gilt für die Menge B_n, dass

$$P(B_n^c) \leq_* P(\tau_{\hat{\eta}} < 1/h_n) + P(\tau_{\hat{X}} < 1/h_n).$$

Beweis. Mit der Ungleichung

$$\begin{aligned} 0 = \inf_{|t|\leq 1/h_n} |\hat{\phi}_X(\hat{\kappa}_r^{rN_n} t)| \geq & \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \\ & - \sup_{|t|\leq 1/h_n} |\hat{\phi}_X - \phi_X|(\alpha^{rN_n} t) \\ & - |1/h_n||\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| \frac{1}{n} \sum_{j=1}^n |X_{rj}| \end{aligned}$$

schätzen wir die Wahrscheinlichkeit ab durch

$$\begin{aligned} P(\tau_{\hat{X}} < 1/h_n) \leq & P\left(2 \sup_{|t|\leq 1/h_n} |\hat{\phi}_X - \phi_X|(\alpha^{rN_n} t) > \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & + P\left(2|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| \frac{1}{n} \sum_{j=1}^n |X_{rj}| > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right). \end{aligned}$$

Mit der Ungleichung $P(|UV| > a) \leq 2P(2|U| > \sqrt{a}) + 2P(2|V| > \sqrt{a})$ für Zufallsvariablen U und V und eine positive Zahl a erhalten wir weiter

$$\begin{aligned} & P\left(2|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| \frac{1}{n} \sum_{j=1}^n |X_{rj}| > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & \leq P\left(4|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| \frac{1}{n} \sum_{j=1}^n (|X_{rj}| - E|X|) > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & \quad + P\left(4|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| E|X| > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & \leq 2P\left(4|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}|^2 > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & \quad + 2P\left(4|\frac{1}{n} \sum_{j=1}^n |X_{rj}| - E|X||^2 > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right) \\ & \quad + P\left(4|\hat{\kappa}_r^{rN_n} - \alpha^{rN_n}| E|X| > h_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right). \end{aligned}$$

Nun schätzen wir die letzten drei Terme mit der Varianz bzw. mit Lemma 1.3 ab, und erhalten die erste Aussage.

Für die zweite Aussage beachten wir, dass gilt

$$
\begin{aligned}
0 = \inf_{|t|\leq 1/h_n} |\tilde{\phi}_\eta|(t) \geq \inf_{|t|\leq 1/h_n} |\phi_\eta|(t) - \sup_{|t|\leq 1/h_n} |\hat{\phi}_\eta - \phi_\eta|(t) \\
- 1/h_n |\hat{\kappa}_r^r - \alpha^r| \frac{1}{n} \sum_{j=1}^{n} |X_{r(j-1)}|.
\end{aligned}
$$

Wir können nun die Wahrscheinlichkeit $P(\tau_{\hat{\eta}} < 1/h_n)$ abschätzen und erhalten

$$
\begin{aligned}
P\left(\tau_{\hat{\eta}} < 1/h_n\right) \leq & P\left(3 \sup_{|t|\leq 1/h_n} |\hat{\phi}_\eta - \phi_\eta|(t) > \inf_{|t|\leq 1/h_n} |\phi_\eta|(t)\right) \\
& + 2P\left(6|\hat{\kappa}_r^r - \alpha^r|^2 > h_n \inf_{|t|\leq 1/h_n} |\phi_\eta|(t)\right) \\
& + 2P\left(6\left(\frac{1}{n}\sum_{j=1}^{n} |X_{r(j-1)}| - E|X|\right)^2 > h_n \inf_{|t|\leq 1/h_n} |\phi_\eta|(t)\right).
\end{aligned}
$$

Mit Lemma 1.3 und Lemma 2.1 folgt wir dann die zweite Aussage. □

Nun wollen wir die Wahrscheinlichkeit der Menge A_n abschätzen.

Lemma 2.14. Für die Wahrscheinlichkeit der Menge A_n haben wir die folgende Abschätzung:

$$
\begin{aligned}
P(A_n^c) \leq_* & \frac{1}{n} + \frac{4/h_n}{b_n \inf_{|t|\leq 1/h_n} |\phi_\eta(t)|} \exp\left(-d\sqrt{n^{1/2} b_n \inf_{|t|\leq 1/h_n} |\phi_\eta(t)|}\right) \\
& + \frac{4/h_n}{b_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{N_n} t)|} \exp\left(-d\sqrt{n^{1/2} b_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{N_n} t)|}\right) \\
& + \frac{4/h_n^2 N_n (|\alpha|^r + \delta)^{N_n - 1}}{n b_n^2 \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{N_n} t)|^2} + \frac{4/h_n^2}{n b_n^2 \inf_{|t|\leq 1/h_n} |\phi_\eta(t)|^2}
\end{aligned}
$$

für ein $\delta > 0$ mit $|\alpha|^r + \delta < 1$

Beweis. Wieder mit der Überlegung $\tilde{\phi}_\eta = \hat{\phi}_\eta + R_n$, wobei

$$
|R_n(t)| \leq |t||\alpha^r - \hat{\kappa}_r^r| \frac{1}{n} \sum_{j=1}^{n} |X_{rj-1}|,
$$

erhalten wir

$$P\left(\sup_{|t|\le 1/h_n}|\tilde{\phi}_\eta/\phi_\eta|(t)>1+b_n\right)\le P\left(\inf_{|t|\le 1/h_n}|\phi_\eta(t)|\le 2b_n\sup_{|t|\le 1/h_n}|\hat{\phi}_\eta-\phi_\eta|(t)\right)$$
$$+P\left(h_n\inf_{|t|\le 1/h_n}|\phi_\eta(t)|\le 2b_n|\alpha^r-\hat{\kappa}_r^r|\frac{1}{n}\sum_{j=1}^{n}|X_{rj-1}|\right),$$

da

$$\left\{\sup_{|t|\le 1/h_n}|\tilde{\phi}_\eta/\phi_\eta|(t)>1+b_n\right\}\subseteq\left\{b_n\inf_{|t|\le 1/h_n}|\phi_\eta(t)|\le\sup_{|t|\le 1/h_n}|\tilde{\phi}_\eta-\phi_\eta|(t)\right\}$$

gilt.

Für die andere Menge $\left\{\sup_{|t|\le 1/h_n}|\phi_\eta/\tilde{\phi}_\eta|(t)>1+b_n\right\}$ haben wir auch

$$\left\{\sup_{|t|\le 1/h_n}|\phi_\eta/\tilde{\phi}_\eta|(t)>1+b_n\right\}\subseteq\left\{\sup_{|t|\le 1/h_n}|(\phi_\eta-\tilde{\phi}_\eta)/\tilde{\phi}_\eta|(t)>b_n\right\}$$
$$\subseteq\left\{\sup_{|t|\le 1/h_n}|\phi_\eta-\tilde{\phi}_\eta|(t)>\frac{b_n}{1+b_n}\inf_{|t|\le 1/h_n}|\phi_\eta(t)|\right\}.$$

Mit dem Beweis von Lemma 2.13 folgt dann die Aussage. □

Lemma 2.15. Eine obere Schranke für die Wahrscheinlichkeit der Menge C_n^c ist für eine $0<\varsigma<1$ gegeben durch die Ungleichung

$$P(C_n^c)\le_*\frac{1}{n}+\frac{1/h_n^2}{b_n^2\inf_{|t|\le 1/h_n}|\phi_\eta(t)|^2 n}+\frac{1/h_n^2}{b_n^2\inf_{|t|\le 1/h_n}|\phi_X(\alpha^{rN_n}t)|^2 n}N_n^2\varsigma^{2N_n}.$$

Beweis. Mit $d_n=\inf_{|t|\le 1/h_n}|\phi_\eta(t)|$ erhalten wir

$$P\left(\bigcup_{k=0}^{N_n}\left\{\sup_{|t|\le 1/h_n}|\phi_\eta(\hat{\kappa}_r^k t)/\phi_\eta(\alpha^k t)|\ge 1+b_n\right\}\right)$$
$$\le\sum_{k=0}^{N_n}P\left(\sup_{|t|\le 1/h_n}|\phi_\eta(\hat{\kappa}_r^k t)-\phi_\eta(\alpha^k t)|\ge b_n d_n\right).$$

Da $E|\eta|=\beta_1$ beschränkt ist, gilt

$$|\phi_\eta(\hat{\kappa}_r^k t)-\phi_\eta(\alpha^k t)|\le\beta_1|t||\hat{\kappa}_r^k-\alpha^k|,$$

und mit Lemma 1.3 folgt

$$P\left(\bigcup_{k=0}^{N_n}\left\{\sup_{|t|\le 1/h_n}|\phi_\eta(\hat{\kappa}_r^k t)/\phi_\eta(\alpha^k t)|\ge 1+b_n\right\}\right)\le_*\frac{N_n}{n}+\frac{1/h_n^2}{nd_n^2b_n^2}\sum_{k=0}^{N_n}k^2\varsigma^{2k}$$

für ein $0 < \varsigma < 1$. Entsprechend können wir die Wahrscheinlichkeit der Menge

$$\bigcup_{k=0}^{N_n} \left\{ \sup_{|t|\leq 1/h_n} |\phi_\eta(\alpha^k t)/\phi_\eta(\hat{\kappa}_r^k t)| \geq 1 + b_n \right\}$$

abschätzen, da $\inf_{|t|\leq 1/h_n} |\phi_\eta(\hat{\kappa}_r^k t)| \geq \inf_{|t|\leq 1/h_n} |\phi_\eta(t)|$.

Es bleibt dann die Wahrscheinlichkeit der Menge

$$\begin{aligned} &\left\{ \sup_{|t|\leq 1/h_n} |\phi_X(\hat{\kappa}_r^{rN_n} t)/\phi_X(\alpha^{rN_n} t)| \geq 1 + b_n \right\} \\ \bigcup &\left\{ \sup_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)/\phi_X(\hat{\kappa}_r^{rN_n} t)| \geq 1 + b_n \right\} \end{aligned}$$

auszurechnen. Wir haben, dass

$$1 + b_n \leq \sup_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)/\phi_X(\hat{\kappa}_r^{rN_n} t)| \leq 1 + \sup_{|t|\leq 1/h_n} |\frac{\phi_X(\alpha^{rN_n} t) - \phi_X(\hat{\kappa}^{rN_n} t)}{\phi_X(\hat{\kappa}^{rN_n} t)}|,$$

und

$$|\phi_X(\alpha^{rN_n} t) - \phi_X(\hat{\kappa}^{rN_n} t)| \leq |t| E|X| |\alpha^{rN_n} - \hat{\kappa}^{rN_n}|.$$

Mit den Vorüberlegungen erhalten wir dann

$$\begin{aligned} &P\left(\sup_{|t|\leq 1/h_n} |\phi_X(\hat{\kappa}_r^{rN_n} t)/\phi_X(\alpha^{rN_n} t)| \geq 1 + b_n \right) \\ &+ P\left(\sup_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)/\phi_X(\hat{\kappa}_r^{rN_n} t)| \geq 1 + b_n \right) \\ &\leq P\left(1/h_n E|X| |\alpha^{rN_n} - \hat{\kappa}^{rN_n}| \geq b_n \inf_{|t|\leq 1/h_n} |\phi_X(\alpha^{rN_n} t)| \right), \end{aligned}$$

und das Lemma folgt dann mit Lemma 1.3. □

Lemma 2.16. Ist $\tilde{\phi}_\varepsilon(t)$ der ϕ_ε-Schätzer, dann haben wir

$$\begin{aligned} &E|\int_{|t|\leq 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon| \mathbf{1}_{A_n} \mathbf{1}_{B_n} \mathbf{1}_{C_n}(t) dt \leq_* \\ &n^{-1/2} \left(\frac{\sum_{k=0}^{N_n - 1} (1 + b_n)^{4k} |\alpha|^{rk}}{\inf_{|t|\leq 1/h_n} |\phi_\eta(t)|} + \frac{|\alpha^{rN_n}|}{\inf_{|t|\leq |\alpha|^{rN_n}/h_n} |\phi_X(t)|} \right) \int_{1/h_n} |t|^{l+1} |\phi_\varepsilon(t)| dt. \end{aligned}$$

Beweis. Für $|t| \leq 1/h_n$ und mit der Gleichung

$$\frac{A}{B} - \frac{a}{b} = \frac{A-a}{b}\frac{b}{B} - \frac{a}{b}\frac{b}{B}(b-B)$$

erhalten wir für $A_n \cap B_n \cap C_n = D_n$

$$E\left|\frac{\tilde{\phi}_\eta(\hat{\kappa}^{rl}t)}{\tilde{\phi}_\eta(\hat{\kappa}^{rl+1}t)} - \frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right|\mathbf{1}_{D_n} \leq \left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right|(1+b_n)^2\frac{E|\hat{\phi}_\eta(\alpha^{rl}t) - \phi_\eta(\alpha^{rl}t)|}{|\phi_\eta(\alpha^{rl}t)|}$$
$$- \left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right|(1+b_n)^2\frac{E|\hat{\phi}_\eta(\alpha^{rl+1}t) - \phi_\eta(\alpha^{rl+1}t)|}{|\phi_\eta(\alpha^{rl+1}t)|}.$$

Da $n(E|\hat{\phi}_\eta(t) - \phi_\eta(t)|)^2 \leq 1 - |\phi_\eta(t)|^2$ und nach Ushakov (1999)[Theorem 2.3.2] auch $1 - |\phi_\eta(t)|^2 \leq \mathrm{Var}(\eta) \cdot |t|^2$, haben wir

$$E\left|\frac{\tilde{\phi}_\eta(\hat{\kappa}^{rl}t)}{\tilde{\phi}_\eta(\hat{\kappa}^{rl+1}t)} - \frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right|\mathbf{1}_{D_n} \leq 2\frac{|t\alpha^{rl}|}{n^{1/2}}\left|\frac{\phi_\eta(\alpha^{rl}t)}{\phi_\eta(\alpha^{rl+1}t)}\right|\frac{|(1+b_n)^2\alpha|\sqrt{\mathrm{Var}(\eta)}}{\inf_{|t|\leq 1/h_n}|\phi_\eta(t)|}. \qquad (1)$$

Mit Lemma 2.1 und oben gemachten Überlegungen folgt auch

$$E\left|\frac{\hat{\phi}_X(\hat{\kappa}^{rN_n}t)}{\hat{\phi}_X(\hat{\kappa}^{rN_n+1}t)} - \frac{\phi_X(\alpha^{rN_n}t)}{\phi_X(\alpha^{rN_n+1}t)}\right|\mathbf{1}_{D_n} \leq 2\frac{|t\alpha^{rN_n}|}{n^{1/2}}\left|\frac{\phi_X(\alpha^{rN_n}t)}{\phi_X(\alpha^{rN_n+1}t)}\right|$$
$$\cdot \frac{|(1+b_n)^2\alpha|\sqrt{\mathrm{Var}(X)}}{\inf_{|t|\leq|\alpha|^{rN_n}/h_n}|\phi_X(t)|}.$$

Das Resultat erhalten wir mit der Ungleichung

$$\left|\prod_{k=0}^{N_n} E_k - \prod_{k=0}^{N_n} F_k\right|\mathbf{1}_{D_n} \leq \sum_{k=0}^{N_n}\left|\prod_{i=0}^{k-1} E_i\right||E_k - F_k|\left|\prod_{i=k+1}^{N_n} F_i\right|\mathbf{1}_{D_n}$$
$$\leq \sum_{k=0}^{N_n}(1+b_n)^{4k}\left|\prod_{i=0}^{k-1} F_i\right||E_k - F_k|\left|\prod_{i=k+1}^{N_n} F_i\right|\mathbf{1}_{D_n}.$$

□

Theorem 2.17. Die Dichte f der Innovationsverteilung ε habe viertes Moment und sei im Punkt x m-mal stetig differenzierbar, und die Funktion $f^{(m)}$ sei im Punkt x Hölder-stetig mit

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x-y|^\delta$$

für ein $0 < \delta < 1$. Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d, β, γ, G und μ die folgende Eigenschaft

$$\exp(-|t|^\beta \gamma) \le |\phi_\varepsilon|(t) \le d|t|^{-m-1-\mu} \text{ für alle } |t| > G.$$

Weiterhin sei K ein Kern mit integrierbarer Fouriertransformation ϕ_K und den Eigenschaften

$$\int K(y)dy = 1\,, \int y^j K(y)dy = 0\,, \int \left|y^j K(y)\right| dy < \infty \text{ und } \int |y^{m+\delta} K(y)|dy < \infty$$

für $j = 1, 2, \ldots, m-1$. Wählen wir $N_n = \ln n$, $b_n = 1/\ln n$ und $h_n = (a(\ln n))^{-1/\beta}$ für ein a mit $(1-|\alpha|^\beta)/(\gamma 2(1-|\alpha|^{r\beta})) > a > 0$, dann gilt

$$\inf_{|t|<1/h_n} |\phi_\eta(t)| \ge n^{-a\gamma(1-|\alpha|^{\beta r})/(1-|\alpha|^\beta)} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist

$$\hat{\eta}_j = X_{rj} - \hat{\kappa}^r X_{r(j-1)}, \quad \tilde{\phi}_\eta(t) = \frac{1}{n}\sum_{j=1}^n \exp(it\hat{\eta}_j),$$

$$\tilde{\phi}_\varepsilon(t) = \prod_{l=0}^{N_n-1} \frac{\tilde{\phi}_\eta(\hat{\kappa}^{rl}t)}{\tilde{\phi}_\eta(\hat{\kappa}^{rl+1}t)} \frac{\hat{\phi}_X(\hat{\kappa}^{rN_n}t)}{\hat{\phi}_X(\hat{\kappa}^{rN_n+1}t)}$$

und

$$\tilde{f}^{(l)}(x) = \frac{1}{2\pi}\int_{|t|\le 1/h_n} e^{-itx}(-it)^l \phi_K(th_n)\tilde{\phi}_\varepsilon(t)\mathbf{1}_{|\tilde{\phi}_\varepsilon|(t)<M} dt$$

der V-Schätzer, dann folgt

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}|(x) \le_* & h_n^{-l-3}/(b_n^2 d_n^2 n) + n^{-1/2}/d_n + h_n^{m-l+\mu} + h_n^{m-l+\delta} \\ = & O(1/(\ln n)^{(\min(\mu,\delta)+m-l)/\beta}). \end{aligned}$$

Beweis. Der Übersichtlichkeit halber lassen wir die reelle Zahl x weg. Es gilt, dass

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}| \le & E|\tilde{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \\ & + E\int_{|t|\le 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon|\mathbf{1}_{D_n}(t)dt \\ & + \left|\int_{|t|\ge 1/h_n} e^{-itx}\phi_K(t)\phi_\varepsilon(t)dt\right| \\ & + \left|\left(f^{(l)}(\cdot) * \frac{1}{h_n}K(\cdot/h_n)\right)(x) - f^{(l)}(x)\right| \end{aligned}$$

und $D_n = A_n \cap B_n \cap C_n$.

Der Schätzer $\tilde{f}^{(l)}$ ist beschränkt durch

$$|\tilde{f}^{(l)}| \le \frac{1}{2\pi} \int_{|t|\le 1/h_n} |t|^l |\phi_K|(th_n)|\tilde{\phi}_\varepsilon|(t)\mathbf{1}_{|\tilde{\phi}_\varepsilon(t)|\le M} dt \le_* h_n^{-l-1}.$$

Darum können wir den ersten Term abschätzen mit

$$E|\tilde{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \le_* h_n^{-l-1} P(D_n^c).$$

Die rechte Seite hängt dann nur vom Wert $\inf_{|t|\le 1/h_n} |\phi_\eta(t)|$ ab. Dieser ist nach Voraussetzung

$$d_n = \inf_{|t|\le 1/h_n} |\phi_\eta(t)| \ge_* n^{-a\gamma(1-|\alpha|^{\beta r})/(1-|\alpha|^\beta)} = o(n^{-1/2}).$$

Mit der Wahl von a folgt dann

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \le & h_n^{-l-1} \left(P(A_n^c) + P(B_n^c) + P(C_n^c)\right) \\ \le_* & h_n^{-l-1} \left(\frac{1}{n} + \frac{1/h_n^2}{b_n^2 d_n^2 n} + \frac{1/h_n}{d_n n}\right) \le_* h_n^{-l-3}/(b_n^2 d_n^2 n) \end{aligned}$$

nach Lemma 2.13, 2.14 und Lemma 2.15.

Für den Term $E\int_{|t|\le 1/h_n} |t|^l|\tilde{\phi}_\varepsilon - \phi_\varepsilon|\mathbf{1}_{D_n}(t)dt$ wenden wir Lemma 2.16 an, und es folgt

$$E\int_{|t|\le 1/h_n} |t|^l|\tilde{\phi}_\varepsilon - \phi_\varepsilon|\mathbf{1}_{D_n}(t)dt \le_* n^{-1/2}/d_n.$$

Nach Voraussetzung ist $|\phi_\varepsilon|(t) \le d|t|^{-m-1-\mu}$ für alle $|t| > G$. Folglich ist

$$\left|\int_{|t|\ge 1/h_n} e^{-itx}(-it)^l \phi_K(t)\phi_\varepsilon(t)dt\right| \le_* h_n^{m-l+\mu}.$$

Wir benutzen jetzt Wefelmeyer (2007)[Satz 22] und erhalten

$$\left|\left(f^{(l)}(\cdot) * \frac{1}{h_n}K(\cdot/h_n)\right)(x) - f^{(l)}(x)\right| \le_* h_n^{m-l+\delta}.$$

□

Das obige Theorem besagt, dass der Schätzer $\tilde{f}^{(l)}(x)$ konsistent ist. Und es ist nicht verwunderlich, dass der Fehler $E|\tilde{f}^{(l)} - f^{(l)}|$ logarithmische Rate hat. Die Voraussetzungen über die Differenzierbarkeit der Dichte und über den Tails der Fouriertransformation sind nicht stark. Dadurch ist eine Vielzahl von Dichten zulässig, die wir schätzen können.

Speziell rechnen wir den Fehler $E|\tilde{f}^{(l)} - f^{(l)}|$ aus, falls der Tail der Fouriertransformation entweder polynomiell oder exponentiell gegen Null konvergiert.

Theorem 2.18 (Die Fouriertransformation der Dichte konvergiert mit polynomieller Rate.). Die Dichte f der Innovationsverteilung ε habe viertes Moment und sei im Punkt x m-mal stetig differenzierbar, und die Funktion $f^{(m)}$ sei im Punkt x Hölder-stetig mit

$$|f^{(m)}(x) - f^{(m)}(y)| \leq L|x-y|^{\delta}$$

für ein $0 < \delta < 1$. Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d_0, d_1, ν, G und μ die folgende Eigenschaft

$$d_0|t|^{-\nu} \leq |\phi_\varepsilon|(t) \leq d_1|t|^{-m-1-\mu} \text{ für alle } |t| > G.$$

Weiterhin sei K ein Kern mit integrierbarer Fouriertransformation ϕ_K und den Eigenschaften

$$\int K(y)dy = 1\,, \int y^j K(y)dy = 0\,, \int \left|y^j K(y)\right| dy < \infty \text{ und } \int |y^{m+\delta} K(y)|dy < \infty$$

für $j = 1, 2, \ldots, m-1$. Wählen wir $N_n = (\ln n)^2$, $b_n = (1/\ln n)^2$ und $h_n = n^{-a}$ für ein a mit $1/(2r\nu) > 1/(2r\nu + l + 3) > a > 0$, dann gilt

$$\inf_{|t|<1/h_n} |\phi_\eta(t)| \geq_* n^{-ar\nu} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist

$$\hat{\eta}_j = X_{rj} - \hat{\kappa}^r X_{r(j-1)}, \quad \tilde{\phi}_\eta(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(it\hat{\eta}_j),$$

$$\tilde{\phi}_\varepsilon(t) = \prod_{l=0}^{N_n-1} \frac{\tilde{\phi}_\eta(\hat{\kappa}^{rl}t)}{\tilde{\phi}_\eta(\hat{\kappa}^{rl+1}t)} \frac{\hat{\phi}_X(\hat{\kappa}^{rN_n}t)}{\hat{\phi}_X(\hat{\kappa}^{rN_n+1}t)}$$

und

$$\tilde{f}^{(l)}(x) = \frac{1}{2\pi}\int_{|t|\leq 1/h_n} e^{-itx}(-it)^l \phi_K(th_n)\tilde{\phi}_\varepsilon(t)\mathbf{1}_{|\tilde{\phi}_\varepsilon|(t)<M} dt$$

der V-Schätzer, dann folgt

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}|(x) \leq_* & h_n^{-l-3}/(b_n^2 d_n^2 n) + n^{-1/2}/d_n + h_n^{m-l+\mu} + h_n^{m-l+\delta} \\ = & \log(n)^2 n^{a(2r\nu+l+3)-1} + n^{-a(m-l+\mu)} + n^{-a(m-l+\delta)}. \end{aligned}$$

Beweis. Wie wir gesehen haben, zerfällt der Fehler $E|\hat{f}^{(l)} - f^{(l)}|$ in die vier Terme

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}| \le & E|\tilde{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \\ & + E\int_{|t|\le 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon| \mathbf{1}_{D_n}(t)dt \\ & + \left|\int_{|t|\ge 1/h_n} e^{-itx}(-it)^l \phi_K(t)\phi_\varepsilon(t)dt\right| \\ & + \left|\left(f^{(l)}(\cdot) * \frac{1}{h_n}K(\cdot/h_n)\right)(x) - f^{(l)}(x)\right| \end{aligned}$$

mit $D_n = A_n \cap B_n \cap C_n$.

Der Schätzer $\hat{f}^{(l)}$ ist beschränkt durch

$$|\hat{f}^{(l)}| \le \frac{1}{2\pi}\int_{\mathbb{R}} |t|^l |\phi_K|(th_n)|\hat{\phi}_\varepsilon|(t)\mathbf{1}_{|\hat{\phi}_\varepsilon(t)|\le M} dt \le_* h_n^{-l-1}.$$

Darum können wir den ersten Term gut abschätzen mit

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \le_* h_n^{-l-1}P(D_n^c).$$

Die rechte Seite hängt dann nur vom Wert $\inf_{|t|\le 1/h_n}|\phi_\eta(t)|$ ab. Dieser ist

$$d_n = \inf_{|t|\le 1/h_n}|\phi_\eta(t)| \ge_* n^{-ar\nu} = o(n^{-1/2}).$$

Mit der Wahl von a folgt dann

$$\begin{aligned} P(D_n^c) \le & P(A_n^c) + P(B_n^c) + P(C_n^c) \\ & \le_* 1/(nd_n^2 b_n^2 h_n^2) + 1/(nh_n d_n) + 1/(nd_n^2 b_n^2 h_n^2) \end{aligned}$$

nach Lemma 2.13, 2.14 und Lemma 2.15.

Also folgt

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \le_* 1/(nb_n^2 d_n^2 h_n^{2+l+1}).$$

Weiter wenden wir Lemma 2.16 an und schätzt den zweiten Term entsprechend ab und erhalten

$$E\int_{|t|\le 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon|\mathbf{1}_{D_n}(t)dt \le 1/(n^{1/2}d_n).$$

Der dritte Term wird entsprechend der Voraussetzung $|\phi_\varepsilon|(t) \le d_1|t|^{-m-1-\mu}$ abgeschätzt. Dieser liefert

$$\left|\int_{|t|\ge 1/h_n} e^{-itx}(-it)^l \phi_K(t)\phi_\varepsilon(t)dt\right| \le h_n^{m-l+\mu}.$$

Für den letzten Term benutzen wir die Eigenschaften des Kerns K zusammen mit der Taylorentwicklung. Dieser erzeugt die Abschätzung

$$\left|\left(f^{(l)}(\cdot) * \frac{1}{h_n} K(\cdot/h_n)\right)(x) - f^{(l)}(x)\right| \leq h_n^{m-l+\delta}.$$

$\square$

Theorem 2.19 (Die Fouriertransformation der Dichte konvergiert mit exponentieller Rate.)**.** Die Fouriertransformation ϕ_ε der Dichte f erfüllt für positive Konstanten d, β, γ_1, G und γ_2 die folgende Eigenschaft

$$\exp(-|t|^\beta \gamma_1) \leq |\phi_\varepsilon|(t) \leq d \exp(-|t|^\beta \gamma_2) \text{ für } |t| > G.$$

Wählen wir $N_n = \ln n$, $b_n = 1/\ln n$ und $h_n = 1/(a \ln n)^{1/\beta}$ für ein a mit $(1 - |\alpha|^\beta)/(2\gamma_1(1 - |\alpha|^{r\beta})) > a$, dann gilt

$$d_n = \inf_{|t|<1/h_n} |\phi_\eta(t)| \geq n^{-a\gamma_1(1-|\alpha|^{r\beta})/(1-|\alpha|^\beta)} = o(n^{-1/2})$$

und

$$\inf_{|t|<|\alpha|^{rN_n}/h_n} |\phi_X(t)| \to 1.$$

Ist

$$\hat{\eta}_j = X_{rj} - \hat{\kappa}^r X_{r(j-1)}, \quad \tilde{\phi}_\eta(t) = \frac{1}{n} \sum_{j=1}^n \exp(it\hat{\eta}_j),$$

$$\tilde{\phi}_\varepsilon(t) = \prod_{l=0}^{N_n-1} \frac{\tilde{\phi}_\eta(\hat{\kappa}^{rl} t)}{\tilde{\phi}_\eta(\hat{\kappa}^{rl+1} t)} \frac{\hat{\phi}_X(\hat{\kappa}^{rN_n} t)}{\hat{\phi}_X(\hat{\kappa}^{rN_n+1} t)}$$

und

$$\tilde{f}^{(l)}(x) = \frac{1}{2\pi} \int_{|t|\leq 1/h_n} e^{-itx} (-it)^l \tilde{\phi}_\varepsilon(t) \mathbf{1}_{|\tilde{\phi}_\varepsilon|(t)<M} dt$$

der V-Schätzer, dann folgt

$$\begin{aligned} E|\hat{f}^{(l)} - f^{(l)}|(x) &\leq_* 1/(n b_n^2 d_n^2 h_n^{l+3}) + 1/(n^{1/2} d_n) + \exp(-h_n^\beta \gamma_2) \\ &\leq_* n^{a\gamma_1(1-|\alpha|^{r\beta})/(1-|\alpha|^\beta)-1/2} + n^{-a\gamma_2}. \end{aligned}$$

Beweis. Der Fehler zerfällt in die drei Terme

$$\begin{aligned} E|\tilde{f}^{(l)} - f^{(l)}| \leq & E|\tilde{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \\ & + E\int_{|t|\leq 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon| \mathbf{1}_{D_n}(t)dt \\ & + \left| \int_{|t|\geq 1/h_n} e^{-itx}(-it)^l \phi_\varepsilon(t)dt \right|. \end{aligned}$$

Der Schätzer $\tilde{f}^{(l)}$ ist beschränkt durch h_n^{-l-1}. Mit Lemma 2.13, 2.14 und 2.15 können wir die Wahrscheinlichkeit $P(D^c)$ abschätzen und erhalten

$$E|\hat{f}^{(l)} - f^{(l)}|\mathbf{1}_{D_n^c} \leq_* 1/(nb_n^2 d_n^2 h_n^{2+l+1}).$$

Der zweite Term ist nach Lemma 2.16 abschätzbar mit

$$E\int_{|t|\leq 1/h_n} |t|^l |\tilde{\phi}_\varepsilon - \phi_\varepsilon| \mathbf{1}_{D_n}(t)dt \leq 1/(n^{1/2} d_n).$$

Für den letzten Term benutzen wir die Voraussetzung $|\phi_\varepsilon|(t) \leq d\exp(-|t|^\beta \gamma_2)$ und erhalten

$$\left| \int_{|t|\geq 1/h_n} e^{-itx}(-it)^l \phi_\varepsilon(t)dt \right| \leq_* \exp(-h_n^\beta \gamma_2) = n^{-a\gamma_2}.$$

□

2.3 Das verallgemeinerte Dekonvolutionsproblem

In einem verallgemeinerten Dekonvolutionsproblem ist man mit der Situation konfrontiert, für eine stationäre Zeitreihe $< (X_i, Y_i) >$ aus den Realisationen

$$Z_i = \phi(X_i, Y_i) \text{ für } i = 1, 2, \ldots$$

auf die Verteilung von X bzw. von Y zu schließen. Wir nehmen der Einfachheit halber an, dass die Familie $\{X_i, Y_j\}$ unabhängig ist. Die Aufgabe ist aber weiterhin unter diesen Bedingungen nicht eindeutig lösbar, da die Verteilung von (X, Y) nicht eindeutig durch die Verteilung von Z bestimmt ist. Eine zusätzliche Nebenbedingung sichert dann i. Allg. die Lösbarkeit der Aufgabe. Wir unterscheiden dabei die vier Nebenbedingungen:

(i) Die Verteilung $\mathcal{L}(X, Y)$ ist bekannt, die Funktion ϕ ist unbekannt. (NB I)

(ii) Entweder $\mathcal{L}(X)$ oder $\mathcal{L}(Y)$ ist unbekannt, die Funktion ϕ ist bekannt. (NB II)

(iii) Es gilt $\mathcal{L}(X) = \mathcal{L}(Y)$, und die Funktion ϕ ist bekannt. (NB III)

(iv) Man beobachtet zusätzlich noch Realisationen

$$\tilde{Z}_i = \zeta(\tilde{X}_i, \tilde{Y}_i) \text{ für } i = 1, 2, \ldots. \tag{NB IV}$$

Dabei sind die Familien $\{\tilde{X}_i\}$ und $\{\tilde{Y}_j\}$ unabhängig und identisch verteilt mit $\mathcal{L}(X) = \mathcal{L}(\tilde{X})$ und $\mathcal{L}(Y) = \mathcal{L}(\tilde{Y})$. Ferner sind die Funktionen ϕ und ζ bekannt.

Wir geben im Weiteren Kommentare für die vier Fälle an.

Die Nebenbedingung (NB I) sichert noch keine Eindeutigkeit der Funktion ϕ. Dazu seien N_1 und N_2 zwei unabhängige und standard-normal verteilte Zufallsvariablen. Es gilt dann:

$$aN_1 + bN_2 \stackrel{\mathcal{D}}{=} aN_1 - bN_2.$$

Das Beispiel zeigt, dass wir i. Allg. weitere Restriktionen verlangen müssen, damit die Aufgabe lösbar ist. Ein weiterer Punkt besteht darin, dass die Funktion ϕ von der Menge $\mathbb{R}^2$ nach $\mathbb{R}$ abbildet. Sie ist i. Allg. nicht injektiv, und das macht die Aufgabe fast nicht lösbar.

Unter der Nebenbedingung (NB II) nehmen wir an, dass die Verteilung $\mathcal{L}(Y)$ nicht bekannt ist, und sowohl die Verteilung $\mathcal{L}(Y)$ als auch die Verteilung $\mathcal{L}(X)$ eine Dichte f_X bzw. f_Y besitzt. Weiter setzen wir voraus, dass eine Funktion ζ existiert mit $\phi(x, \zeta(x, y)) = y$, und mit $\phi'(x, y) = \partial_y \phi(x, y)$ bezeichnen wir die erste Ableitung der zweiten Komponente. Dann hat die Verteilung $\mathcal{L}(Z)$ eine Dichte h_Z mit

$$(Tf_Y)(z) = h_Z(z) = \int f_Y(\zeta(x, z)) f_X(x) \zeta'(x, z) dx.$$

Wir setzen diese Abbildung auf den linearen Abschluss der Dichte fort zu einem linearen Operator T

$$Tu = \int u(\zeta(x, \cdot)) f_X(x) \zeta'(x, \cdot) dx.$$

Setzen wir Id als den Identischen Operator und sei $\| \mathrm{Id} - T \| < 1$, dann ist T invertierbar mit

$$T^{-1} = \sum_{k=0}^{\infty} (-1)^k (\mathrm{Id} - T)^k \text{ und } T^{-1} h_Z = f_Y.$$

Es gilt i. Allg. nicht, dass $\|T - \mathrm{Id}\| < 1$. Dies würde aber auch bedeuten, dass der Operator T auf der ganzen Definitionsmenge injektiv ist. Wir brauchen die Injektivität nur für das eine Element f_Y bzw. h_Z. Wissen wir, dass der Operator T im Punkt f_Y injektiv ist, dann können wir mittels $T^{-1}\hat{h}_Z$ einen Schätzer für f_Y definieren. Dieser zerfällt dann in die zwei Terme

$$T^{-1}\hat{h}_Z = T^{-1}(\hat{h}_Z - h_Z) + f_Y.$$

Der Schätzer $T^{-1}\hat{h}_Z$ ist also konsistent, falls T im Punkt f_Y injektiv ist, und $T^{-1}(\hat{h}_Z - h_Z) = o_P(1)$. Als Beispiel betrachten wir $\phi(x, y) = x + y$. Dafür gilt

$$(Tu - u)(z) = \int_0^1 \int u'(z - tx) x f_X(x) dx \text{ und } \|Tu - u\|_1 \leq \|u'\|_1 E|X|,$$

wobei $\|\cdot\|_1$ die L_1-Norm ist. Da der Differentialoperator auch linear ist, liefert diese

$$\left\| \sum_{k=0}^{\infty} (-1)^k (\mathrm{Id} - T)^k h_Z \right\|_1 \leq \sum_{k=0}^{\infty} (E|X|)^k \|h_Z^{(k)}\|_1.$$

Falls die rechte Seite endlich ist, dann haben wir $T^{-1}h_Z = f_Y$. Für den anderen Term haben wir

$$E\|T^{-1}(\hat{h}_Z - h_Z)\|_1 \leq \sum_{k=0}^{\infty} (E|X|)^k E\|\hat{h}_Z^{(k)} - h_Z^{(k)}\|_1.$$

Die rechte Seite ist aber in den meisten Fällen unbeschränkt. Um das Problem zu lösen, nehmen wir nun als Schätzer

$$\hat{f}_Y = \sum_{k=0}^{N_n} (-1)^k (\mathrm{Id} - T)^k \hat{h}_Z$$

für eine Folge $N_n \to \infty$. Der Schätzer, den wir hier vorstellen, ist zwar unter bestimmten Voraussetzungen konsistent, aber nicht optimal. Der Vorteil dieser Methode ist die allgemeine Anwendbarkeit auf andere Funktionen ϕ. Dabei müssen wir untersuchen, wann der Ausdruck

$$\sum_{k=0}^{\infty} (-1)^k (\mathrm{Id} - T)^k$$

für eine Norm endlich ist. Im vorigen Beispiel haben wir die L_1-Norm genommen. Dies ist aber nicht nötig. Wir können auch andere angenehme Normen benutzen.

Unter der Nebenbedingung (NB III) und unter gewissen Voraussetzungen an die Funktion ϕ erhalten wir

$$h_Z(z) = \int f_Y(\zeta(x, z)) f_Y(x) \zeta'(x, z) dx$$

und

$$Tu = \int u(\zeta(x,\cdot))u(x)\zeta'(x,\cdot)dx.$$

Hier ist der Operator T nicht linear, dennoch kann man die Inverse

$$T^{-1} = \sum_{k=0}^{\infty}(-1)^k(\mathrm{Id} - T)^k$$

definieren. Diese braucht auch nur für die Dichte h_Z endlich zu sein, um die Konsistenz des Schätzers

$$\hat{f}_Y = \sum_{k=0}^{N_n}(-1)^k(\mathrm{Id} - T)^k\hat{h}_Z$$

zu erhalten. Somit stellt die Neumann-Reihe wieder einen brauchbaren Lösungsansatz dar.

Unter der Nebenbedingung (NB IV) beobachten wir Paare von Realisationen $(Z_i, \hat{Z}_i)$ für $i = 1, 2, \ldots, n$. Wieder unter günstigen Voraussetzungen sind wir in der Lage, Dichten

$$h(\cdot) = \int f_X(\tilde{\phi}(x,\cdot))f_Y(x)\tilde{\phi}'(x,\cdot)dx$$

und

$$g(\cdot) = \int f_X(\tilde{\zeta}(x,\cdot))f_Y(x)\tilde{\zeta}'(x,\cdot)dx$$

mit dem Kernschätzer zu schätzen. Es ist dann möglich den Operator T durch

$$T(u,v) = \left(\int u(\tilde{\phi}(x,\cdot))v(x)\tilde{\phi}'(x,\cdot)dx, \int u(\tilde{\zeta}(x,\cdot))v(x)\tilde{\zeta}'(x,\cdot)dx\right)$$

und die Inverse durch

$$T^{-1} = \sum_{k=0}^{\infty}(-1)^k(\mathrm{Id} - T)^k$$

zu definieren. Ein brauchbarer Schätzer ist

$$(\hat{f}_X, \hat{f}_Y) = T^{-1}(\hat{h}, \hat{g}).$$

Dieser ist unter bestimmten Voraussetzungen sogar konsistent. Die obige Methode zeigt, dass man nicht in der Lage ist, nur die interessierte Dichte f_X ohne die Dichte f_Y zu schätzen.

Wir geben nun ein Beispiel an, wo man nur die interessierte Dichte schätzen muss. Seien

$$Z_i = \alpha X_i + \beta Y_i \text{ und } \hat{Z}_i = \gamma \hat{X}_i + \delta \hat{Y}_i.$$

Wir sind an die Dichte f_X interessiert. Ohne Beschränkung der Allgemeinheit nehmen wir an, dass $\alpha \cdot \beta \cdot \gamma \cdot \delta \neq 0$. Ferner soll gelten, dass $|(\delta\beta)/(\gamma\alpha)| \neq 1$. Reskalieren wir nun die Realisationen mit

$$\frac{1}{\alpha} Z_i = X_i + \frac{\beta}{\alpha} Y_i \text{ und } \frac{\beta}{\alpha\delta}\hat{Z}_i = \frac{\delta\beta}{\alpha\delta}\hat{X}_i + \frac{\beta}{\alpha}\hat{Y}_i,$$

dann sind wir in der Lage die charakteristische Funktion

$$\frac{\phi_X(t)}{\phi_X((\delta\beta)/(\gamma\alpha)t)} \text{ mit } \frac{\hat{\phi}_{Z/\alpha}(t)}{\hat{\phi}_{(\beta\hat{Z})/(\alpha\delta)}(t)}$$

zu schätzen. Ab hier können wir dann weiter mit dem ϕ_ε-Schätzer verfahren. Das Beispiel soll zeigen, dass man immer versuchen soll, ϕ-spezifisch zu dekonvolieren, um eine optimale Leistung des Schätzers zu erhalten. Ist dies nicht möglich, wie im Fall von $\phi(x, y) = r(x) + \sigma(x)y$, dann stellt die Neumann-Reihe einen guten Ansatz dar.

2.4 Schätzer der Innovationsdichte in periodisch beobachteten heteroskedastischen nichtlinearen Regressionsmodellen

Wir wollen nun aus den Beobachtungen

$$Z_i = Y_i\sigma(X_i) + r(X_i) \text{ für } i = 1, \dots,$$

die Dichte f_Y der Verteilung $\mathcal{L}(Y)$ schätzen. Dabei nehmen wir an, dass die beiden Verteilungen $\mathcal{L}(Y)$ und $\mathcal{L}(X)$ jeweils die Dichte f_Y bzw. f_X besitzen. Ferner sind die Zufallsvariablen $< X_i >$ und $< Y_i >$ unabhängig und identisch verteilt, und $< X_i, Y_j >$ sind unabhängig. Hier kennen wir die Funktion $\phi(x, y) = y\sigma(x) + r(x)$ und die Dichte f_X. Falls $\inf_{t\in\mathbb{R}} |\sigma(t)| > 0$, dann hat die Verteilung von Z eine Dichte

$$f_Z(z) = \int_{\mathbb{R}} f_Y((z - r(x)/\sigma(x))f_X(x)\frac{dx}{\sigma(x)}.$$

Der dazugehörige Operator T ist dann definiert durch

$$Tu = \int_{\mathbb{R}} u((z - r(x)/\sigma(x))f_X(x)\frac{dx}{\sigma(x)}.$$

Nun müssen wir den Ausdruck

$$\sum_{k=0}^{\infty}(-1)^k(\operatorname{Id}-T)^k f_Z$$

geeignet abschätzen. Dazu schreiben wir abkürzend $r(X) = r$ und $\sigma(X) = \sigma$ und präsentieren das folgende Lemma.

Lemma 2.20. Sei u eine differenzierbare Funktion, dann erhalten wir

$$\|Tu - u\| \leq a_1\|u\| + a_2\|u'\| + a_3\|u' \operatorname{Id}\|,$$

wobei $a_1 = |E1 - 1/\sigma|$, $a_2 = E|r/\sigma|$ und $a_3 = E|1 - 1/\sigma|$. Hier ist $\|u\| = \int |u|(z)dz$ die L^1-Norm.

Gilt zusätzlich $\partial_z Eu((z-r)/\sigma)/\sigma = Eu'((z-r)/\sigma)/\sigma^2$, dann folgt

$$\|(Tu - u)'\| \leq a_3\|u'\| + \|Tu' - u'\|$$

und

$$\|(Tu - u)' \cdot \operatorname{Id}\| \leq a_2\|u'\| + \|T(u' \operatorname{Id}) - (u' \operatorname{Id})\|.$$

Beweis. Wir benutzen die Gleichung

$$\begin{aligned} Tu(z) - u(z) =& E\left(u((z-r)/\sigma) - u(z/\sigma)\right)\frac{1}{\sigma} \\ &+ E(u(z/\sigma) - u(z))(1/\sigma) + u(z)E(1/\sigma - 1), \end{aligned}$$

dann liefert diese

$$\|Tu - u\| \leq |E1/\sigma - 1| \cdot \|u\| + E\int |u(z) - u(z + r/\sigma)|dz + E\int |u(z) - u(z\sigma)|dz.$$

Da u differenzierbar ist, kann man die letzten beiden Terme geeignet abschätzen, und wir haben

$$\begin{aligned} u(z) - u(z - r/\sigma) &= \int_0^1 u'(z + tr/\sigma)dt\frac{r}{\sigma} \text{ und} \\ u(z) - u(z\sigma) &= \int u'(tz + z\sigma(1-t))dt \cdot z \cdot (1 - \sigma). \end{aligned}$$

Da $|t + \sigma(1-t)| > 0$, können wir eine Transformation mit

$$z \longrightarrow \frac{z}{t + \sigma(1-t)} \text{ und } dz = \frac{1}{t + \sigma(1-t)}$$

durchführen. Dies ergibt

$$
\begin{aligned}
E\int |u(z)-u(z\sigma)|dz &\le \int |u'(z)z|dz\cdot E\int_0^1 \left|\frac{1-\sigma}{(t+\sigma(1-t))^2}\right| dt\\
&= \int |u'(z)z|dz\cdot E|1-1/\sigma|.
\end{aligned}
$$

Für die letzten beiden Gleichungen bemerken wir, dass

$$
\begin{aligned}
(Tu)'(z) =& Eu'((z-r)/\sigma)/\sigma^2\\
=& Eu'((z-r)/\sigma)/\sigma + Eu'((z-r)/\sigma)/\sigma(1/\sigma-1)
\end{aligned}
$$

und

$$
(Tu)'(z)\cdot z = T(u'\,\mathrm{Id})(z) + Eu'((z-r)/\sigma)(r/\sigma^2).
$$

□

Das Lemma 2.20 hat gezeigt, dass wir i. Allg. verlangen müssen, dass die Dichte f_Z beliebig oft differenzierbar ist, damit der Ausdruck

$$
\sum_{k=0}^{\infty}(-1)^k(\mathrm{Id}-T)^k f_Z
$$

wohldefiniert ist. Diese Voraussetzung kann man in der Praxis nicht überprüfen. Verlangt man stattdessen, dass die Verteilung $\mathcal{L}(Z)$ beschränkten Träger hat, dann ist man in der Lage EZ^k durch den empirischen Schätzer zu schätzen. Diese liefert $EZ = Er + E\sigma EY$ und

$$
EZ^k = E(r+\sigma Y)^k = \sum_{j=0}^{k}\binom{k}{j} Er^k\sigma^{k-j}EY^j.
$$

Also ist es auch möglich, die Momente EY^k und darum auch die charakteristische Funktion Ee^{itY} zu schätzen. Diese liefert dann einen Zugang, um die Dichte f_Y zu schätzen.

3 Lokale asymptotische Normalität

In diesem Kapitel legen wir den Grundstein für den Begriff der "Effizienz" fest. Dabei stellen wir die uniforme und die punktweise Version der lokalen asymptotischen

Normalität unseres Prozesses vor und wiederholen auch die uniforme, lokale asymptotische Normalität des klassischen AR(1)-Modells. Diese dient dem Vergleich und findet auch in der Konstruktion des effizienten Schätzers seine Anwendung.

Um den effizienten Schätzer zu bestimmen, braucht man nur die punktweise Version der asymptotischen Normalität. Diese legt den lokalen Parameterraum fest, der aus den Tangenten besteht. Aber um den effizienten Schätzer zu konstruieren, braucht man dafür die uniforme Version der lokalen asymptotischen Normalität. Einen Überblick über lokale asymptotische Normalität für Markov-Ketten bieten die Artikel von Wefelmeyer (1999), Höpfner *et al.* (1990), und speziell für lineare Prozesse Schick & Wefelmeyer (2002).

3.1 Einführung und Notation

Für ein Maß μ und eine messbare Funktion v bezeichnen wir mit μv das Integral von v bezüglich μ.

Wir fixieren nun ein Element (α, f) aus Θ, und bezeichnen mit P_α das Maß, daß eindeutig von (α, f) induziert wird.

Wir kürzen $Y = Y_1$, $Z = Z_1$, $\eta = \eta_1$ und $Z_i = Y_{i+1}$ ab, da wir häufig diese Zufallsvariablen brauchen. Für eine Funktion v setzen wir zudem

$$S_n(v) = \frac{1}{n} \sum_{j=1}^{n} v(Y_i, Z_i).$$

Eine Folge $< \alpha_n >$ heißt lokal beschränkt (gegen α), falls $\sqrt{n}(\alpha_n - \alpha)$ beschränkt ist. Häufig werden wir auch eine lokal beschränkte Folge $< \alpha_n >$ fixieren, und setzen dann $\eta_{n,i} = Z_i - \alpha_n^r Y_i$.

3.2 LAN für autoregressive Prozesse

Im klassischen autoregressiven Modell haben wir

$$Y_{n+1} = \alpha^r Y_n + \eta_{n+1}$$

mit $|\alpha^r| < 1$, und $< \eta_n >$ ist eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Dichte g. Hier ist der Parameterraum $\Theta = \mathcal{R} \times \mathcal{D}$. Der Raum $\mathcal{D}$ ist die Menge der Dichten mit endlicher Varianz und Fisher-Information.

Sei nun T die Menge der beschränkten und stetig differenzierbaren Funktionen u mit

$$\int g(x)u(x)dx = 0 \quad \text{und} \quad \sup_x |u'(x)| < \infty.$$

Wir fixieren weiterhin eine Funktion $u \in T$ und definieren für $a = (a_1, a_2) \in \mathbb{R}^2$ $\alpha_a = \alpha + a_1$,

$$g_a = g(1 + a_2 u) \quad \text{ und } \quad q_a(y, z) = g_a(z - \alpha_a^r y).$$

Für kleine a_1 und a_2 ist $|\alpha_a^r| < 1$. Da u beschränkt und bezüglich g zentriert ist, ist g_a für genügend kleine a_2 eine Dichte. Ferner hat die Zufallsvariable $\sum_{j=0}^{\infty} \alpha_a^{rj} \eta_{-j}^a$ eine Dichte h_a, wobei $< \eta_j^a >$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen ist mit Dichte g_a.

Für festes u aus T definieren wir den Likelihood-Quotient durch

$$\Lambda_n(a_1, a_2) = \sum_{k=1}^{n} \log \frac{q_a}{q}(Y_k, Z_k) + \log \frac{h_a}{h}(Y_0)$$

und den Operator D auf T durch

$$(Du)(y, z) = \left(-r\alpha^{r-1} y \ell_g(z - \alpha^r y), u(z - \alpha^r y)\right)^\top,$$

wobei $\ell_g = g'/g$. Wir sind nun bereit, LAN für das klassische autoregressive Modell zu wiederholen.

Theorem 3.1 (LAN für klassisches autoregressives Modell)**.** Hat g endliche Fisher-Information und endliches zweites Moment, dann gilt:

$$\Lambda_n(t_n n^{-1/2}, n^{-1/2}) = (t_n, 1) n^{1/2} S_n(Du) - 1/2 (t_n, 1) \operatorname{Var}(Du)(t_n, 1)^\top + o_P(1).$$

Zusätzlich haben wir $\mathcal{L}(n^{1/2} S_n(Du) \mid P) \Rightarrow \mathcal{N}(0, \operatorname{Var}(Du))$. Dabei ist

$$\operatorname{Var}(Du) = P_{Y,Z}(Du)(Du)^\top$$

die Kovarianz-Matrix, und $< t_n >$ ist eine beschränkte Folge in $\mathbb{R}$.

Beweis. Siehe zum Beispiel Koul & Schick (1997) und Schick & Wefelmeyer (2002). □

Insbesondere besteht der lokale Tagentenraum $\mathcal{H}$ aus Funktionen der Form

$$(t, 1)(Du)$$

mit (t, u) aus $\mathbb{R} \times T$. Der Raum $\mathcal{H}$ ist ein Teilraum aus $L^2(P_{Y,Z})$. Er ist aber nicht abgeschlossen.

Sei nun $< \alpha_n >$ eine beschränkte Folge gegen α. Bezeichnen wir mit $P_{\alpha_n}^{(n)}$ das n-dimensionale Wahrscheinlichkeitsmaß, dass eindeutig durch die Übergangsdichte $q_{\alpha_n}(y, z) = g(z - \alpha_n y)$ und invariante Startverteilung induziert wird, dann gilt

$$\log \frac{dP_{\alpha_n}^n}{dP^n}(Y_0, \ldots, Y_n) = \Lambda_n(\alpha - \alpha_n, 0).$$

Bemerkung 3.2. Nach dem ersten Lemma von LeCams siehe zum Beispiel van der Vaart (1998)[Lemma 6.4], folgt, dass die beiden Folgen $< P_{\alpha_n} >$ und $< P >$ benachbart sind. Diese Eigenschaft findet später Anwendung in der Konstruktion des effizienten Schätzers.

3.3 LAN für periodisch beobachtete autoregressive Prozesse

In diesem Abschnitt nehmen wir an, dass

$$Y_{n+1} = \alpha^r Y_n + \eta_{n+1} \quad \text{und} \quad \eta_n = \sum_{j=0}^{r-1} \varepsilon_{rn-j}\alpha^j.$$

Dabei ist $|\alpha| < 1$, und $< \varepsilon_j >$ ist eine Folge von unabhängig und identisch verteilten Zufallsvariablen mit Dichte f. Dieser Prozess kommt vor, falls wir von einem klassischen autoregressiven Modell nur jede r-te Realisationen sehen. Wir haben also an die Innovationen $< \eta_n >$ noch eine strukturelle Annahme, denn die Dichte von η ist

$$L(\alpha, f)(x) = \int f\left(x - \sum_{j=1}^{r-1} \alpha^{r-j} y_j\right) \prod_{k=1}^{r-1} f(y_k) dy_k.$$

Hier ist der Innovationsraum

$$\mathcal{I} = \{L(\alpha, f) \,|\, (\alpha, f) \in \mathcal{R} \times \mathcal{D}\}.$$

Der Parameterraum Θ ist aber nicht das Produkt von $\mathcal{R}$ und $\mathcal{I}$, da die Innovationen vom Regressionsparameter abhängig sind.

Sei nun, wie im klassischen Fall, T die Menge der beschränkten und stetig differenzierbaren Funktionen u mit

$$\int f(x)u(x)dx = 0 \quad \text{und} \quad \sup_x |u'(x)| < \infty.$$

Wir fixieren weiterhin eine Funktion $u \in T$, und definieren für $a = (a_1, a_2) \in \mathbb{R}^2$, $\alpha_a = \alpha + a_1$,

$$f_a = f(1 + a_2 u) \quad \text{und} \quad q_a(y_0, y_r) = \int f_a\left(y_r - \sum_{j=0}^{r-1} \alpha_a^{r-j} y_j\right) \prod_{k=1}^{r-1} f_a(y_k) dy_k.$$

Für kleine a_1 und a_2 ist $|\alpha_a^r| < 1$. Da u beschränkt und bezüglich f zentriert ist, ist f_a eine Dichte. Ferner hat die Zufallsvariable $\sum_{j=0}^{\infty} \alpha_a^j \varepsilon_{-j}^a$ eine Dichte h_a, wobei $< \varepsilon_j^a >$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen ist mit Dichte f_a.

Für festes u aus T definieren wir den Likelihood-Quotienten durch

$$\Lambda_n(a_1, a_2) = \sum_{k=1}^{n} \log \frac{q_a}{q}(Y_k, Z_k) + \log \frac{h_a}{h}(Y_0)$$

und den Operator D auf T durch

$$(Du)(y_0, y_r) = \begin{pmatrix} -ry_0\alpha^{r-1}E[\ell_r \,|\, \eta = \zeta] - \sum_{j=1}^{r-1}(r-j)\alpha^{r-1-j}E[\ell_r\varepsilon_j \,|\, \eta = \zeta] \\ \sum_{j=1}^{r} E[u_j \,|\, \eta = \zeta] \end{pmatrix}$$

mit $\eta = \sum_{j=0}^{r-1} \alpha^j \varepsilon_{r-j}$, $\zeta = y_r - \alpha^r y_0$, $\ell_r = \ell_f(\varepsilon_r)$ und $u_j = u(\varepsilon_j)$. Der Likelihood-Quotient $\Lambda_n(a)$ ist von der Funktion u abhängig. Diese haben wir aber wegen der Übersichtlichkeit weggelassen.

Wie im klassischen Fall haben wir das folgende Theorem.

Theorem 3.3 (Uniforme lokale asymptotische Normalität)**.** Für alle u aus T gilt

$$n^{1/2}\Lambda_n(t_n n^{-1/2}, n^{-1/2}) = \nu_n^\top n^{1/2} S_n(Du) - \frac{1}{2}\nu_n^\top \operatorname{Var}(Du)\nu_n + o_P(1),$$

wobei $\nu_n^\top = (t_n, 1)$, und $< t_n >$ eine beschränkte Folge in $\mathbb{R}$ ist.

Zusätzlich gilt

$$\mathcal{L}(n^{1/2} S_n(Du) \,|\, P) \Rightarrow \mathcal{N}(0, \operatorname{Var}(Du)).$$

Speziell folgt die lokale asymptotische Normalität.

Theorem 3.4 (Lokale asymptotische Normalität)**.** Für alle t aus $\mathbb{R}$ und u aus T haben wir

$$\Lambda_n(tn^{-1/2}, n^{-1/2})(t,1)n^{1/2}S_n(Du) - \frac{1}{2}(t,1)\operatorname{Var}(Du)(t,1)^\top + o_P(1).$$

Insbesondere können wir den Operator D auf $\mathbb{R} \times T$ erweitern:

$$\begin{aligned} D_{t,u}(y,z) = (t,1)(Du)(y,z) = &- try\alpha^{r-1}E[\ell_r \,|\, \eta = \zeta] \\ &- t\sum_{j=1}^{r-1}(r-j)\alpha^{r-1-j}E[\ell_r\varepsilon_j \,|\, \eta = \zeta] \\ &+ \sum_{j=1}^{r} E[u_j \,|\, \eta = \zeta] \end{aligned}$$

mit $\zeta = z - \alpha^r y$.

Man beachte, dass $E[\ell_r \,|\, \eta = \zeta] = \ell_g(\zeta)$ gilt. Führen wir die Abkürzungen

$$\aleph(y,z) = \sum_{j=1}^{r-1} (r-j)\alpha^{r-1-j} E[\ell_r \varepsilon_j \,|\, \eta = \zeta] \text{ und}$$
$$(\daleth u)(y,z) = \sum_{j=1}^{r} E[u_j \,|\, \eta = \zeta]$$

ein, dann können wir $D_{t,u}$ kompakter schreiben als

$$D_{t,u}(y,z) = -t(r\alpha^{r-1} y \ell_g(\zeta) + \aleph(y,z)) + (\daleth u)(y,z).$$

Der lokale Tangentenraum $\mathcal{H}$ besteht also aus Funktionen der Form $-t(yr\alpha^{r-1}\ell_g + \aleph) + (\daleth u)$ mit $(t,u) \in \mathbb{R} \times T$. Für das Theorem 3.3 ist hinreichend, dass der Pfad $a \to q_a^{1/2}$ Hellinger differenzierbar ist.

Lemma 3.5 (Hellinger-Differenzierbarkeit für die Übergangsdichte)**.** Sei $a = (a_1, a_2)^\top$ und $|\cdot|$ die euklidische Norm für Vektoren, dann gilt λ-f-s für alle y_0

$$\int \left(\sqrt{q_a} - \sqrt{q} - a^\top 1/2 (Du)\sqrt{q}\right)^2 (y_0, y_r) dy_r = (1 + y_0^2) o(|a|^2).$$

Hier ist der o-Term nicht von y_0 abhängig.

Beweis. Wir kürzen ab

$$K_a(\bar{y}_{0,r}) = f_a(y_r - \sum_{j=1}^{r} \alpha_a^j y_{r-j}) \prod_{k=1}^{r-1} f_a(y_k)$$

mit $\bar{y}_{0,r} = (y_0, \ldots, y_r)$. Die Abbildung $a \to K_a$ ist stetig differenzierbar mit Gradient

$$\partial_{a_1} K_a = -\sum_{k=1}^{r} k\alpha_a^{k-1} y_{r-k} \frac{K_a}{f_a(y_r - \sum_{j=1}^{r} \alpha_a^j y_{r-j})} f_a'(y_r - \sum_{j=1}^{r} \alpha_a^j y_{r-j}) \quad \text{und}$$
$$\partial_{a_2} K_a = \sum_{j=1}^{r-1} \frac{K_a}{f_a(y_j)} (fu)(y_j) + \frac{K_a}{f_a(y_r - \sum_{j=1}^{r} \alpha_a^j y_{r-j})} (fu)(y_r - \sum_{j=1}^{r} \alpha_a^j y_{r-j}).$$

Kürzen wir weiter ab $\nabla K_a = (\partial_{a_1} K_a, \partial_{a_2} K_a)^\top$, dann gilt

$$\int K_a(\bar{y}_{0,r}) d\bar{y}_{1,r-1} - \int K_b(\bar{y}_{0,r}) d\bar{y}_{1,r-1} - (b-a)^\top \int (\nabla K_a)(\bar{y}_{0,r}) d\bar{y}_{1,r-1}$$
$$= (b-a)^\top \int \int_0^1 (\nabla K_{a+(s-1)(b-a)} - \nabla K_a)(\bar{y}_{0,r}) ds d\bar{y}_{1,r-1}.$$

Zeigen wir nun, dass die rechte Seite mit der Rate $o(|b-a|)$ gegen Null konvergiert, dann ist die Abbildung $a \to q_a$ stetig differenzierbar mit Gradient $\nabla q_a = \int \nabla K_a$. Da die Variable s über ein Kompaktum integriert wird, reicht zu zeigen, dass

$$\int |(\nabla K_{a+s} - \nabla K_a)| \, (\bar{y}_{0,r}) d\bar{y}_{1,r-1} \to 0, \text{ für } s \to 0.$$

Wir integrieren über die Variable y_0 und y_r und zeigen nun

$$\int |(\nabla K_{a+s} - \nabla K_a)| \, (\bar{y}_{0,r}) d\bar{y}_{0,r} \to 0, \text{ für } s \to 0.$$

Da $c|f| < |f_a| < C|f|$, u und u' beschränkt sind, die Abbildung $a \to \nabla K_a$ stetig ist und man eine Transformation $y_r \to y_r + \sum_{j=1}^r \alpha_a^j y_{r-j}$ machen kann, folgt die Aussage mit den Sätzen aus Elstrodt (2004)[§5 Satz 5.3 und §5 Satz 5.4].

Es folgt nun, dass die Abbildung $a \to q_a$ λ^2-f-s stetig differenzierbar ist, und dies impliziert λ^2-f-s

$$\sqrt{q_a} - \sqrt{q} = a^\top 1/2 \int_0^1 \frac{\nabla q_{as}}{\sqrt{q_a}} ds.$$

Wir zeigen nun, dass das Integral $\int (\frac{\nabla q_a}{\sqrt{q_a}})^2 (y_0, y_r) dy_r$ endlich und stetig in a ist. Denn das würde mit dem Satz aus Elstrodt (2004)[§5 Satz 5.4] implizieren, dass

$$\int \left| \frac{\nabla q_a}{\sqrt{q_a}} - (Du)\sqrt{q} \right|^2 (y_0, y_r) dy_r = o(1).$$

Seien $\varepsilon_1^a, \varepsilon_2^a, \ldots, \varepsilon_r^a$ unabhängige und identisch verteilte Zufallsvariablen mit Verteilungsdichte f_a. Da $|f_a| <_* f$, konvergiert die Funktionsfolge $< f_a >$ in L^1 gegen f und auch $xf_a(x)$ in L^1 gegen $xf(x)$. Setzen wir weiterhin $\eta_a = \sum_{j=1}^r \varepsilon_{j-r}^a \alpha_a^j$ und $\zeta_a = y_r - \alpha_a^r y_0$, dann gilt

$$\frac{\partial_{a_1} q_a}{q_a}(y_0, y_r) = -r y_0 \alpha_a^{r-1} E[\ell^a(\varepsilon_r^a) \,|\, \eta_a = \zeta_a] - \sum_{j=1}^{r-1} (r-j) \alpha_a^{r-1-j} E[\ell^a(\varepsilon_r^a) \varepsilon_j^a \,|\, \eta_a = \zeta_a],$$

$$\frac{\partial_{a_2} q_a}{q_a}(y_0, y_r) = \sum_{j=1}^r E[\frac{u}{1+a_2 u}(\varepsilon_j^a) \,|\, \eta_a = \zeta_a],$$

wobei $\ell^a = \frac{f_a'}{f_a} = \frac{f'}{f} + a_2 \frac{u'}{1+a_2 u}$. Die Jensensche Ungleichung für bedingte Erwartungswerte impliziert

$$\left(\frac{\partial_{a_1} q_a}{q_a}\right)^2 (y_0, y_r) \leq_* E[(\ell^a(\varepsilon_r^a))^2 (y_0 + \sum_{j=1}^{r-1} |\varepsilon_j^a|)^2 \,|\, \eta_a = \zeta_a] \quad \text{und}$$

$$\left(\frac{\partial_{a_2} q_a}{q_a}\right)^2 (y_0, y_r) \leq_* \|u\|_\infty,$$

wobei $\|\cdot\|_\infty$ die Supremumsnorm bezeichnet. Da $q_a(y_0,\cdot)$ in L^1 gegen $q(y_0,\cdot)$ konvergiert, folgt, dass $\left(\frac{\partial_{a_2} q_a}{q_a}\right)^2 q_a$ uniform integrabel ist. Für den ersten Term beachte man, dass $(\ell^a(\varepsilon_r^a))^2(y_0+\sum_{j=1}^{r-1}|\varepsilon_j^a|)^2$ in L^1 konvergiert, also uniform integrabel ist. Das zeigt, dass das Integral
$\int(\frac{\nabla q_a}{\sqrt{q_a}})^2(y_0,y_r)dy_r$ endlich und stetig in a ist.

Da die Ungleichung

$$\int\left(\frac{\partial_{a_1} q_a}{q_a}\right)^2 q_a(y_0,y_r)dy_r \leq_* \left(|y_0|+\sum_{j=1}^{r-1}\sqrt{E|\varepsilon_j^a|^2)}\right)^2 E[(\ell^a(\varepsilon_r^a))^2]$$

gilt, folgt

$$\int\left|\frac{\nabla q_a}{\sqrt{q_a}}-(Du)\sqrt{q}\right|^2(y_0,y_r)dy_r=(1+y_0^2)o(1).$$

□

Das folgende Lemma ist die Voraussetzung von Koul & Schick (1997)[Theorem 2.4].

Lemma 3.6 (Hellinger-Differenzierbarkeit für die Dichte)**.** Sei $a=(a_1,a_2)$, dann gilt

$$\int\left(\sqrt{g_a}-\sqrt{g}-1/2a^\top\begin{pmatrix}-\sum_{j=1}^{r-1}(r-j)\alpha^{r-1-j}E[\ell_r\varepsilon_j\,|\,\eta=\cdot]\\ \sum_{j=1}^{r}E[u_j\,|\,\eta=\cdot]\end{pmatrix}\right)^2 d\lambda=o(\|a\|^2),$$

wobei $g(x)=\int f(x-\sum_{j=1}^{r-1}\alpha^j y_j)\prod_{k=1}^{r-1}f(y_k)dy_k$ die Dichte von $\eta=\sum_{j=0}^{r-1}\alpha^j\varepsilon_{r-j}$ und $g_a(x)=\int f_a(x-\sum_{j=1}^{r-1}\alpha_a^j y_j)\prod_{k=1}^{r-1}f_a(y_k)dy_k$.

Beweis. Setzen wir $y_0=0$ in Lemma 3.5 ein, dann erhalten wir das Resultat. □

Lemma 3.5 wurde am Anfang formuliert und bewiesen. Dieses soll seine Anwendung in Roussas (1972) finden, da Roussas (1972) Hellinger-Differenzierbarkeit des Übergangskern voraussetzt. Später haben wir gesehen, dass man auch Koul & Schick (1997) benutzen kann. Dort wird verwendet, dass die lineare Autoregression eine Markovkette ist mit der Gleichung

$$X_{n+1}=\phi_\alpha(X_n,\varepsilon_{n+1})=\psi_\alpha(X_n)+\varepsilon_{n+1}.$$

Aus diesem Grund kann man nun die Pfade für α und für f separat betrachten. Das vereinfacht den Beweis vom Theorem 3.3 erheblich. Aber allgemein für den Beweis der lokalen asymptotischen Normalität ist Roussas (1972) besser geeignet.

Wir sind nun bereit, das Theorem 3.3 zu beweisen.

Beweis zu Theorem 3.3. Wir zeigen zuerst, dass $\int |h_a - h|$ gegen Null konvergiert. Denn daraus folgt, dass

$$\frac{h_a}{h}(Y_0) - 1 = o_P(1).$$

Seien die beiden Folgen von Zufallsvariablen $< \varepsilon_j >$ und $< \varepsilon_j^a >$ so gewählt, dass sie unabhängig sind. Ferner sei $\pi(Z) = \frac{dP_Z}{d\lambda}$ der Operator, der jeder Zufallsvariablen Z deren Dichte zuordnet. Dann gilt

$$\begin{aligned}
\int |h_a - h| &= \lambda \left| \pi\left(\sum_{j=0}^{\infty} \alpha_a^j \varepsilon_j^a\right) - \pi\left(\sum_{j=0}^{\infty} \alpha^j \varepsilon_j\right) \right| \\
&\le \lambda \left| \pi\left(\sum_{j=0}^{\infty} \alpha_a^j \varepsilon_j^a\right) - \pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right) \right| \\
&\quad + \lambda \left| \pi\left(\sum_{j=0}^{\infty} \alpha^j \varepsilon_j\right) - \pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right) \right|.
\end{aligned}$$

Wir zeigen nun, dass die beiden Terme beliebig klein werden. Für den ersten Term betrachte man, dass $\pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right) = *_{j=0}^{N} \pi\left(\alpha^j \varepsilon_j\right) * \pi\left(\sum_{j>N} \alpha_a^j \varepsilon_j^a\right)$, und dass gilt

$$\lambda \left| \pi\left(\sum_{j=0}^{\infty} \alpha_a^j \varepsilon_j^a\right) - \pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right) \right| \le \sum_{j=0}^{N} \lambda \left| \pi\left(\alpha_a^j \varepsilon_j^a\right) - \pi(\alpha^j \varepsilon_j) \right|$$

denn $\prod_{i=1}^{N} a_i - \prod_{i=1}^{N} b_i = \sum_{i=1}^{N} \prod_{k=1}^{i-1} a_k (a_i - b_i) \prod_{k=i+1}^{N} b_k$ und $\int |f * g| \le \int |f| \int |g|$.

Da $\lambda \left|\pi(\alpha_a^j \varepsilon_j^a) - \pi(\alpha^j \varepsilon_j)\right|$ gegen Null konvergiert, konvergiert der obere Term für festes N gegen Null. Für den zweiten Term setzen wir $X = \sum_{j=0}^{\infty} \alpha^j \varepsilon_j$ und $X_a = \sum_{j=0}^{\infty} \alpha_a^j \varepsilon_j^a$ ab. Es gilt

$$\begin{aligned}
\pi(X)(x) - \pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right)(x) &= E g_N(x - \alpha^{N+1} X) - E g_N(x - \alpha_a^{N+1} X_a) \\
&= \int_0^1 E g_N'(x - \alpha^{N+1} X + t(\alpha^{N+1} X - \alpha_a^{N+1} X_a))(\alpha^{N+1} X - \alpha_a^{N+1} X_a) dt
\end{aligned}$$

mit $g_N = \pi(\sum_{j=0}^{N} \alpha^j \varepsilon_j)$. Daraus folgt:

$$\begin{aligned}
\lambda \left| \pi(X)(x) - \pi\left(\sum_{j=0}^{N} \alpha^j \varepsilon_j + \sum_{j>N} \alpha_a^j \varepsilon_j^a\right)(x) \right| &\le \int |g_N'| \cdot E|\alpha^{N+1} X - \alpha_a^{N+1} X_a)| \\
&\le E\ell_f^2(\varepsilon) \cdot (\alpha^{N+1} E|X| + \alpha_a^{N+1} E|X_a|) \\
&\to 0,
\end{aligned}$$

da $(\int |g_N'|)^2 \leq \int (g_N'/g_N)^2 g_N \leq E\ell_f^2(\varepsilon)$ nach Hájek *et al.* (1999)[Theorem 1].

Da der Prozess strikt stationär und ergodisch ist, folgt die punktweise lokale asymptotische Normalität nach Roussas (1972)[Theorem 4.1] und Lemma 3.5. Allerdings hat Roussas das Theorem nicht auf die uniforme Version erweitert, was man analog nachrechnen kann. Der Prozess ist andererseits auch eine Markov-Kette, also kann man genauso gut Höpfner *et al.* (1990)[Theorem 1.24 und Theorem 1.27] und Lemma 3.5 benutzen, um die lokale asymptotische Normalität zu zeigen. Wie Roussas hat Höpfner das Ergebnis nicht auf die uniforme Version erweitert. Dies führt dazu, dass wir Koul & Schick (1997)[Theorem 2.4] benutzen werden und deren Voraussetzungen nachrechnen.

Wir führen wie bei Koul & Schick (1997) die Funktionen $H_j(\alpha_n) = \alpha_n^r Y_j$ und $\dot{H}_j(\alpha_n) = r\alpha_n^{r-1} Y_j$ ein, und zeigen für lokal beschränkte Folgen $< \alpha_n >$ und $< \theta_n >$

(i) $\sum_{j=1}^n |H_j(\alpha_n) - H_j(\theta_n) - (\alpha_n - \theta_n)\dot{H}_j(\alpha_n)|^2 = o_{\alpha_n}(1)$,

(ii) $\max_{1\leq j\leq n} \frac{1}{\sqrt{n}} |\dot{H}_j(\alpha_n)| = o_{\alpha_n}(1)$,

(iii) $\frac{1}{n}\sum_{j=1}^n \dot{H}_j(\alpha_n) = r\alpha^{r-1} E_\alpha Y + o_{\alpha_n}(1)$,

(iv) $\frac{1}{n}\sum_{j=1}^n \dot{H}_j^2(\alpha_n) = r^2\alpha^{2r-2} E_\alpha Y^2 + o_{\alpha_n}(1)$.

Die Aussage (i) folgt, da $(\alpha_n^r - \theta_n^r) - (\alpha_n - \theta_n) r\alpha_n^{r-1} = o(1/n)$ und $\frac{1}{n}\sum_{j=1}^n Y_j^2 = O_{\alpha_n}(1)$.

Für (ii) verwende man, dass

$$\begin{aligned}\max_{1\leq k\leq n} |Y_k| &\leq \max_{1\leq k\leq n} |\alpha_n|^{rk} |Y_0| + \sum_{j=0}^{rk-1} |\alpha_n^j||\varepsilon_{rk-j}| \\ &\leq |Y_0| + \frac{1}{1-|\alpha|^r} \max_{0\leq k\leq rn} |\varepsilon_k| \text{ unter } P_{\alpha_n}.\end{aligned}$$

Da $< \varepsilon_j >$ endliche Varianz hat, folgt $\max_{1\leq k\leq n} |\varepsilon_k| = o_{\alpha_n}(\sqrt{n})$.

Für (iii) zeigen wir nun, dass

$$\frac{1}{n}\sum_{j=1}^n Y_j - E_\alpha Y = o_{\alpha_n}(1).$$

Da $Y_{j+1} = \alpha_n^r Y_j + \eta_{j+1}$, haben wir

$$\frac{1}{n}\sum_{j=1}^{n-1} Y_j(1-\alpha_n^r) + \frac{1}{n} Y_n - \alpha_n^r \frac{1}{n} Y_0 = \frac{1}{n}\sum_{j=1}^n \eta_j - E_{\alpha_n}\eta + E_{\alpha_n}\eta \quad \text{ unter } P_{\alpha_n}.$$

Weil zudem $\mathrm{Var}_{\alpha_n}(\frac{1}{n}\sum_{j=1}^n \eta_j - E_{\alpha_n}\eta)$ gegen Null konvergiert, folgt die Aussage.

Für (iv) benutzen wir $Y_{j+1}^2 = \eta_{j+1}^2 + 2\alpha_n^r \eta_{j+1} Y_j + \alpha_n^{2r} Y_j^2$ und diese liefert

$$
\begin{aligned}
\frac{1}{n}\sum_{j=1}^{n-1} Y_j^2(1-\alpha_n^{2r}) + \frac{1}{n}Y_n^2 - \alpha_n^{2r}\frac{1}{n}Y_0^2 =& \frac{1}{n}\sum_{j=1}^{n}\eta_j^2 - E_{\alpha_n}\eta^2 \\
&+ E_{\alpha_n}\eta^2 + 2\alpha_n^r \frac{1}{n}\sum_{j=1}^{n}(\eta_j - E_{\alpha_n}\eta)Y_{j-1} \\
&+ 2\alpha_n^r E_{\alpha_n}\eta \frac{1}{n}\sum_{j=1}^{n} Y_{j-1} \quad \text{unter } P_{\alpha_n}.
\end{aligned}
$$

Der erste Term konvergiert gegen Null, denn $E_{\alpha_n}\exp(it\frac{1}{n}\sum_{j=1}^n \eta_j^2 - E_{\alpha_n}\eta^2) \to 1$ für alle t aus $\mathbb{R}$. Da $\mathrm{Var}(\frac{1}{n}\sum_{j=1}^n(\eta_j - E_{\alpha_n}\eta)Y_{j-1}) = \frac{1}{n}\mathrm{Var}_{\alpha_n}(\eta)E_{\alpha_n}Y^2$, erhalten wir zusammen mit (iii)

$$
\frac{1}{n}\sum_{j=1}^{n} Y_j^2 - E_\alpha Y^2 = o_{\alpha_n}(1).
$$

Das Theorem folgt nun mit Lemma 3.6. □

Wir listen nun in der folgenden Tabelle die Unterschiede zwischen dem klassischen autoregressiven Modell mit und ohne periodische Beobachtungen auf. Tabelle 2 gibt die Unterschiede im semiparametrischen Fall an, während Tabelle 3 die Unterschiede im parametrischen Fall angibt.

Semiparametrischer Fall

	vollbeobachtete klassische $AR(1)$-Zeitreihe	periodisch beobachtete $AR(1)$-Zeitreihe
Innovationen	η_n	$\eta_n = \sum_{j=0}^{r-1}\varepsilon_{rn-j}\alpha^j$
Dichte der Innovation	g	$g(x) = Ef(x - \sum_{j=1}^{r-1}\alpha^{r-j}\varepsilon_j)$
$u \in T$	$Eu(\eta) = 0$	$Eu(\varepsilon) = 0$
der lokale Tangentenraum $\mathcal{H}$	$-tr\alpha^{r-1}y\ell_g + u$	$-t(yr\alpha^{r-1}\ell_g + \aleph) + (\daleth u)$

Tabelle 2: Der lokale Tangentenraum der beiden Modelle

Parametrischer Fall

	K_r-Modell	T_r-Modell
der lokale Tangentenraum $\mathcal{H}$	$-tr\alpha^{r-1}y\ell_g$	$-t(yr\alpha^{r-1}\ell_g + \aleph)$
die Varianz des kanonischen Gradients	$1/\mathrm{Var}(r\alpha^{r-1}Y\ell_g(\eta))$	$1/\mathrm{Var}(Yr\alpha^{r-1}\ell_g(\eta) + \aleph)$

Tabelle 3: Die Varianz des kanonischen Gradienten zwischen den beiden Modellen

Wie wir in Tabelle 3 sehen können, trägt die struktuelle Annahme tatsächlich dazu bei, den Regressionsparameter besser zu schätzen, falls die Innovationen zentriert sind. Es gilt nämlich die folgende Ungleichung

$$\begin{aligned}\operatorname{Var}(Yr\alpha^{r-1}\ell_g(\eta) + \aleph(\eta)) &= \operatorname{Var}(Yr\alpha^{r-1}\ell_g(\eta)) + \operatorname{Var}(\aleph(\eta)) \\ &> \operatorname{Var}(Yr\alpha^{r-1}\ell_g(\eta)).\end{aligned}$$

4 Faltungssatz und Effizienz

In diesem Abschnitt wollen wir effiziente Schätzer charakterisieren. Dabei bezeichnen wir mit $\lfloor\cdot,\cdot\rfloor$ das Skalarprodukt auf dem Hilbertraum $L^2(Y,Z)$. Für $\Bbbk \in \mathbb{R}\times T$ sei $< a_{n,\Bbbk} >$ eine Folge in Θ mit

$$\log \frac{dP^{(n)}_{a_{n,\Bbbk}}}{dP^{(n)}_{\alpha,f}}(Y_0, Y_1, \ldots, Y_n) = n^{1/2} S_n(D_\Bbbk) - 1/2 \operatorname{Var}(D_\Bbbk) + o_P(1).$$

Hier ist mit $P^{(n)}_{a_{n,\Bbbk}}$ das n-dimensionale Maß vom Parameter $a_{n,\Bbbk}$ gemeint. Sei nun κ ein Funktional von Θ nach $\mathbb{R}$.

Definition 4.1. (i) Wir sagen, κ ist differenzierbar in (α, f) mit Gradient g, wenn $g \in L^2(Y,Z)$ und

$$n^{1/2}\left(\kappa(a_{n,\Bbbk}) - \kappa(\alpha, f)\right) \to \lfloor D_\Bbbk, g\rfloor.$$

Der kanonische Gradient g_0 ist die Projektion eines beliebigen Gradienten auf den Abschluss von $\mathcal{H}$.

(ii) Ein Schätzer $\hat\kappa$ heißt regulär für κ in (α, f) mit Limes V, wenn

$$n^{1/2}\left(\hat\kappa - \kappa(a_{n,\Bbbk})\right) \Rightarrow V \text{ unter } P_{a_{n,\Bbbk}}$$

für alle $\Bbbk$ aus $\mathbb{R}\times T$.

(iii) Ferner heißt der Schätzer $\hat\kappa$ asymptotisch linear für κ in (α, f) mit Einflussfunktion v, wenn v aus $L^2(Y,Z)$ und

$$n^{1/2}\left(\hat\kappa - \kappa(\alpha, f)\right) = n^{1/2}(S_n(v) - Ev(Y,Z)) + o_P(1).$$

Theorem 4.2 (Faltungssatz)**.** Ist κ ein differenzierbares Funktional mit kanonischem Gradienten g_0 und der Schätzer $\hat{\kappa}$ regulär in (α, f), so gilt für eine von $\mathcal{N}$ unabhängige Zufallsvariable U:

$$n^{1/2}(S_n(g_0), \hat{\kappa} - \kappa(\alpha, f)) \Rightarrow (\sqrt{\operatorname{Var}(g_0)}\mathcal{N}, U) \text{ unter } P^{(n)}.$$

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

Definition 4.3. Ein Schätzer $\hat{\kappa}$ heißt effizient für $\kappa(\alpha, f)$, wenn

$$n^{1/2}\left(\hat{\kappa} - \kappa(\alpha, f)\right) \Rightarrow \sqrt{\operatorname{Var}(g_0)}\mathcal{N}.$$

Beweis zu Theorem 4.2. Siehe Wefelmeyer (2007)[Satz 5]. □

5 Effizientes Schätzen des autoregressiven Parameters

In diesem Abschnitt wollen wir nachweisen, dass das Funktional $\kappa(\alpha, f) = \alpha$ differenzierbar ist. Ferner wollen wir den kanonischen Gradienten schätzen, um dann im nächsten Abschnitt einen effizienten Schätzer für α zu konstruieren. Es gilt, falls

$$\log \frac{dP^{(n)}_{\alpha_n, f_n}}{dP^{(n)}_{\alpha, f}}(Y_0, Y_1, \dots, Y_n) = n^{1/2} S_n(D_{t,u}) - 1/2 \operatorname{Var}(D_{t,u}) + o_P(1),$$

dass $\alpha_n = \alpha + t/n^{1/2}$. Ferner haben wir $n^{1/2}(\kappa(\alpha_n, f_n) - \kappa(\alpha, f)) = t$ für alle $(t, u) \in \mathbb{R} \times T$. Wir müssen nun ein g_0 aus $\overline{\mathcal{H}}$ finden mit $t = \lfloor D_{t,u}, g_0 \rfloor$ für alle $(t, u) \in \mathbb{R} \times T$.

Es gilt $D_{t,u}(Y, Z) = -tr\alpha^{r-1} Y \ell_g(\eta) - t\aleph(\eta) + (\daleth u)(\eta)$ mit

$$\aleph(\eta) = \sum_{j=1}^{r-1} (r - j)\alpha^{r-1-j} E[\ell_r \varepsilon_j \,|\, \eta]$$

und

$$(\daleth u)(\eta) = \sum_{j=1}^{r} E[u_j \,|\, \eta].$$

Bezeichnen wir nun mit Proj die Projektion auf dem Raum $\overline{\{\daleth(u)\,;\, u \in T\}}$, und wählen wir als Funktion

$$w(Y, Z) = -r\alpha^{r-1} Y \ell_g(\eta) - r\aleph(\eta) - \operatorname{Proj}(-r\alpha^{r-1} Y \ell_g(\eta) - r\aleph(\eta)),$$

dann gilt:

$$\begin{aligned}\lfloor D_{t,u}, w\rfloor =& \lfloor -tr\alpha^{r-1} Y \ell_g(\eta) - t\aleph(\eta) + (\daleth u)(\eta), w\rfloor \\ =& t\lfloor w, w\rfloor + \lfloor(\daleth u)(\eta), w\rfloor.\end{aligned}$$

Hier ist der Abschluss von $\{\daleth(u)\,;\, u \in T\}$ im Raum $L^2(\eta)$ gemeint.

Da die Funktion w senkrecht auf $\overline{\{\daleth(u)\,;\, u \in T\}}$ steht, verschwindet $\lfloor(\daleth u)(\eta), w\rfloor$. Ein Gradient ist somit die Funktion $w/\|w\|^2$. Die Funktion $\mathrm{Proj}(-r\alpha^{r-1} Y \ell_g(\eta) - \aleph(\eta))$ liegt im Raum $\overline{\mathcal{H}}$, da die Funktionen $D_{0,u}$ alle im Raum $\overline{\mathcal{H}}$ liegen. Das zeigt, dass die Funktion

$$g_0 = w/\|w\|^2$$

der kanonische Gradient ist.

Wir werden sehen, dass $\mathrm{Proj}(Y\ell_g(\eta)) = EY \cdot \ell_g(\eta)$. Da die Projektion linear ist, besteht die ganze Schwierigkeit darin, die Funktion $\aleph(\eta)$ zu projizieren. Wir bemerken, dass

$$\mathrm{Proj}(\aleph(\eta)) = \arg\min_{w \in \mathcal{S}} \|\aleph(\eta) - w(\eta)\|.$$

Hier ist $\mathcal{S} = \overline{\{\daleth(u)\,;\, u \in T\}}$.

Wir unterscheiden die folgenden drei Fälle:

(F_1) $\min_{w\in\mathcal{S}} \|\aleph(\eta) - w(\eta)\| = 0$,

(F_2) $\min_{w\in\mathcal{S}} \|\aleph(\eta) - w(\eta)\| = \|\aleph(\eta)\|$,

(F_3) $0 < \min_{w\in\mathcal{S}} \|\aleph(\eta) - w(\eta)\| < \|\aleph(\eta)\|$.

Der Fall (F_1) ist der angenehmste Fall von allen. In diesem Fall haben wir

$$\begin{aligned}\|w\|^{-2} &= \|r\alpha^{r-1}(Y - EY)\ell_g(\eta)\|_2^2 \\ &= (r\alpha^{r-1})^2 \operatorname{Var}(Y) \operatorname{Var}(\ell_g(\eta)) \qquad \text{und} \\ g_0 &= -\|w\|^{-2} r\alpha^{r-1}(Y - EY)\ell_g(\eta) \\ &= -\frac{(Y - EY)\ell_g(\eta)}{r\alpha^{r-1} \operatorname{Var}(Y) \operatorname{Var}(\ell_g(\eta))}.\end{aligned}$$

Der kanonische Gradient stimmt in Fall (F_1) also mit dem kanonischen Gradienten im klassischen autoregressiven Modell ohne die strukturelle Annahme an die Innovationen überein. Das impliziert, dass die strukturelle Annahme an die Innovationen nicht zur Varianz-Reduktion beiträgt. Im nächsten Kapitel geben wir hinreichende Bedingungen für den Fall (F_1) an, und konstruieren auch einen effizienten Schätzer.

Der Fall (F_2) ist genau das Gegenteil von (F_1). Im Fall (F_2) haben wir

$$\begin{aligned}\|w\|^{-2} &= \|r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)\|_2^2 \quad \text{und}\\ g_0(Y,Z) &= -\frac{r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)}{\|r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)\|_2^2}.\end{aligned}$$

Sind die Innovationen zusätzlich zentriert, dann haben wir weiter

$$\mathrm{Proj}(r\alpha^{r-1}Y\ell_g(\eta)+\aleph(\eta)) = r\alpha^{r-1}EY\cdot\ell_g+\mathrm{Proj}(\aleph(\eta))=0.$$

Dies führt dazu, dass wir von einem adaptiven System reden können. Der kanonische Gradient stimmt mit dem kanonischen Gradient im klassischen autoregressiven Modell mit struktureller Annahme an die Innovationen und bekannter Dichte f überein. Dieser Fall kommt eher selten vor, da es notwendig ist, dass

$$\lfloor \aleph(\eta), (\daleth u)(\eta)\rfloor = 0 \qquad \text{für alle } u\in T.$$

Das ist eine sehr starke Einschränkung. Dieser Fall wird daher auch nicht mehr weiter behandelt.

Für den Fall (F_3) haben wir

$$\begin{aligned}\|w\|^{-2} &= \|r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)-\mathrm{Proj}(\aleph)(\eta)\|_2^2 \text{ und}\\ g_0(Y,Z) &= -\frac{r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)-\mathrm{Proj}(\aleph(\eta))}{\|r\alpha^{r-1}(Y-EY)\ell_g(\eta)+\aleph(\eta)-\mathrm{Proj}(\aleph)(\eta)\|_2^2}.\end{aligned}$$

Um den effizienten Schätzer im Fall (F_3) zu konstruieren, brauchen wir einen Schätzer für f. Diesen haben wir in Kapitel 2 angegeben. Später in Kapitel 6.3 werden wir diskutieren, wie wir die Funktionen $\aleph(\eta)$ und $\mathrm{Proj}(\aleph(\eta))$ schätzen können.

Wir zeigen nun, dass $\mathrm{Proj}(Y\ell_g(\eta)) = EY\cdot\ell_g(\eta)$ gilt.

Lemma 5.1. Es existiert ein u aus $L^2(\varepsilon)$ mit

$$(\daleth u)(\eta) = \ell_g(\eta).$$

Ferner haben wir

$$\mathrm{Proj}(Y\ell_g(\eta)) = EY\cdot\ell_g(\eta).$$

Beweis. Sei g die Dichte von η. Dann gilt mit Hájek *et al.* (1999)[Chapter2, Theorem 1], dass g differenzierbar ist mit

$$g'(x) = \int f'(x - \sum_{j=1}^{r-1} \alpha^j y_j) \prod_{k=1}^{r-1} f(y_k) dy_k.$$

Wieder mit demselben Satz gilt auch

$$g'(x) = \alpha^i \int f(x - \sum_{j=1}^{r-1} \alpha^j y_j) \prod_{k=1}^{r-1} f(y_k) \frac{f'}{f}(y_i) dy_k.$$

für alle $i = 1, \ldots, r$. Daraus folgt

$$\sum_{j=1}^{r} E[\ell_j \,|\, \eta] = \sum_{j=1}^{r-1} \alpha^j E[\ell_r \,|\, \eta] = \alpha \frac{1-\alpha^{r-1}}{1-\alpha} \ell_g(\eta).$$

Wählen wir nun $u = \alpha^{-1} \frac{1-\alpha}{1-\alpha^{r-1}} \ell_f$, dann folgt die erste Aussage.

Für die zweite Aussage betrachten wir

$$\begin{aligned} \lfloor \mathrm{Proj}(Y\ell_g(\eta)), (\urcorner v)(\eta) \rfloor &= \lfloor Y\ell_g(\eta), (\urcorner v)(\eta) \rfloor \\ &= EY \lfloor \ell_g(\eta), (\urcorner v)(\eta) \rfloor, \end{aligned}$$

da Y und η unabhängig sind. □

Die ganze Schwierigkeit besteht nun darin, die Funktion $E[\ell_r \varepsilon_j \,|\, \eta]$ zu projizieren. Dazu werden wir benutzen, dass

$$\mathrm{Proj}(E[\ell_r \varepsilon_j \,|\, \eta]) = \arg\min_{w \in \mathcal{S}} \| E[\ell_r \varepsilon_j \,|\, \eta] - w(\eta) \|.$$

5.1 Die Konstruktion des effizienten Schätzers unter einer zusätzlichen strukturellen Annahme

In diesem Kapitel werden wir hinreichende Bedingungen dafür angeben, so dass

$$\mathrm{Proj}(\aleph(\eta)) = \aleph(\eta).$$

Danach konstruieren wir einen Schätzer α_{eff} mit

$$\begin{aligned} \alpha_{eff} - \alpha &= S_n(g_0) + o_P(n^{-1/2}) \\ &= -\frac{1}{r\alpha^{r-1} \operatorname{Var}(Y) \operatorname{Var}(\ell_g(\eta))} \frac{1}{n} \sum_{j=1}^{n} \ell_g(\eta_j)(Y_{j-1} - EY) + o_P(n^{-1/2}). \end{aligned}$$

Da die Einflussfunktion genau der kanonische Gradient ist, ist der Schätzer α_{eff} der effiziente Schätzer für α.

Für eine integrierbare Funktion u sei

$$u^\wedge(t) = \int e^{itx} u(x) dx$$

die Fouriertransformation. Es gilt das folgende Lemma.

Lemma 5.2. (i) Ist die Funktion u aus $L^1(\lambda)$ differenzierbar, und ist u' integrierbar, dann gilt

$$(u')^\wedge(t) = -itu^\wedge(t).$$

(ii) Sei $< a_n >$ eine Folge von reellen Zahlen mit $\sum_{j=0}^{\infty} |a_j| < \infty$, und h sei eine integrierbare Funktion. Dann ist die Funktion $x \to \sum_{j=0}^{\infty} h(xa_j^{-1})$ λ-fast-überall wohldefiniert und integrierbar. Ferner haben wir

$$\left(\sum_{j=0}^{\infty} h(xa_j^{-1}) \right)^\wedge (t) = \sum_{j=0}^{\infty} a_j h^\wedge(a_j t).$$

(iii) Das Integral über u' ist zentriert, d. h. $\int u' d\lambda = 0$.

Beweis. Sei $< a_n >$ eine Nullfolge, dann gilt

$$\begin{aligned} \frac{1}{a_n} \int e^{itx}(u(x+a_n) - u(x)) dx &= \int e^{itx} \int_0^1 u'(x + sa_n) ds dx \\ &= \int_0^1 e^{-itsa_n} ds \cdot (u')^\wedge(t) \to (u')^\wedge(t), \end{aligned}$$

da u' integrierbar ist. Andererseits haben wir

$$\begin{aligned} \frac{1}{a_n} \int e^{itx}(u(x+a_n) - u(x)) dx =& \frac{1}{a_n}(e^{-ita_n} - 1) u^\wedge(t) \\ & \to -itu^\wedge(t). \end{aligned}$$

Für die zweite Aussage betrachten wir

$$\int \left| \sum_{j=M}^{N} h(xa_j^{-1}) \right| dx \leq \int \sum_{j=M}^{N} \left| h(xa_j^{-1}) \right| dx = \sum_{j=M}^{N} |a_j| \int |h(x)| dx \to 0$$

für $N, M \to \infty$. Also ist $\sum_{j=0}^{N} h(xa_j^{-1})$ eine Cauchy-Folge in $L^1(\lambda)$. Außerdem, da

$$|\sum_{j=0}^{\infty} h(xa_j^{-1})| \leq \sum_{j=0}^{\infty} |h(xa_j^{-1})|,$$

folgt die Aussage mit dem Satz über dominierte Konvergenz.

Für die letzte Aussage betrachten wir

$$\begin{aligned}\int |1/a|\,|u(x+a) - u(x)|dx \leq & \int \left| \int_0^1 u'(x+ta)dt \right| dx \\ \leq & \int_0^1 \int |u'(x)|dx < \infty.\end{aligned}$$

□

Definition 5.3. Sei f eine Dichte, deren charakteristische Funktion nirgends verschwindet. Wir sagen, die Dichte f aus $\mathcal{F}$ erfüllt die Annahme (TF), falls eine Funktion h aus $L^1(\lambda)$ und Zahlen $B_0, \ldots, B_m$ aus $\mathbb{R}$ existieren mit

$$\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(t) = h^\wedge(t) + \sum_{j=0}^{m-1} B_j(-it)^j. \qquad (TF)$$

Dabei ist $\mathrm{Id}(x) = x$ die identische Funktion. Weiterhin sei die Dichte f m-mal differenzierbar, $\int |f^{(j)}|(x)dx < \infty$ für alle $j = 0, \ldots, m$ und die beiden charakteristischen Funktionen $|h^\wedge|$ und $|f^\wedge(t)t^m|$ sind integrierbar.

Bemerkung 5.4. Erfüllt eine Dichte f die Annahme TF, dann gilt

$$h^\wedge(t) + \sum_{j=0}^{m-1} B_j(-it)^j = \frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(t) = i\frac{(f^\wedge)'}{f^\wedge}(t) \qquad (2)$$

$$= i\partial_t \log f^\wedge(t). \qquad (3)$$

Insbesondere erfüllt eine Dichte die Annahme (TF), falls der Term $i\frac{(f^\wedge)'}{f^\wedge}(t)$ wieder eine Fouriertransformierte ist. Ist nun f eine Dichte mit endlichem ersten Moment. Dann ist die Funktion $\exp(\rho(f^\wedge(t) - 1))$ nach Ushakov (1999)[Corollary 1.3.4] wieder eine charakteristische Funktion. Ferner haben wir

$$i\frac{\exp(\rho(f^\wedge(t) - 1))'}{\exp(\rho(f^\wedge(t) - 1))} = i(f^\wedge)'(t) = (f\,\mathrm{Id})^\wedge(t).$$

Dies zeigt, dass es eine große Klasse von charakteristischen Funktionen gibt, die die Gleichung (2) erfüllen.

Nun geben wir eine kleine Liste von Dichten an, die die Annahme (TF) erfüllt.

Name	Dichte	Charakteristische Funktion	$i(f^\wedge)'/f^\wedge$
Standard-Laplace-Verteilung	$\frac{1}{2}e^{\vert x\vert}$	$\frac{1}{1+t^2}$	$\frac{-2ti}{1+t^2}$
Exponential-Verteilung	$\lambda e^{-\lambda x}1_{x\geq 0}$	$\frac{\lambda}{\lambda - it}$	$\frac{-it}{\lambda - it}$
Normal-Verteilung	$\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$e^{i\mu t-\frac{\sigma^2 t^2}{2}}$	$-\mu - i\sigma^2 t$

Tabelle 4: Die Annahme TF

Unter der Annahme (TF) zeigen wir nun das folgende Lemma.

Lemma 5.5. Sei f aus $\mathcal{F}$, und f erfülle die Annahme (TF). Dann existiert für alle $k = 1, \ldots, r-1$ eine zentrierte Funktion $v \in L^1(\varepsilon)$ mit

$$\sum_{j=0}^{r-1} E[v(\varepsilon_j) \,|\, \eta] = E[\ell_f(\varepsilon_r)\varepsilon_k \,|\, \eta].$$

Beweis. Wir nehmen zuerst an, dass v existiert. Dann gilt

$$\int (fv)(x - \sum_{j=1}^{r-1} \alpha^j y_j) \prod_{d=1}^{r-1} f(y_d) d\bar{y}_{1,r-1} + \sum_{k=1}^{r} \int f(x - \sum_{j=1}^{r-1} \alpha^j y_j) \prod_{d=1}^{r-1} f(y_d) v(y_k) d\bar{y}_{1,r-1}$$
$$= \int f'(x - \sum_{j=1}^{r-1} \alpha^j y_j) \prod_{d=1}^{r-1} f(y_d) y_k d\bar{y}_{1,r-1}.$$

Nehmen wir nun auf beiden Seiten die Fouriertransformation und setzen

$$\phi(t) = \prod_{j=0}^{r-1} f^\wedge(\alpha^j t),$$

dann erhalten wir

$$\phi(t)\sum_{k=0}^{r-1}\frac{(fv)^\wedge}{f^\wedge}(\alpha^k t) = \phi(t)\left(\frac{(f')^\wedge}{f^\wedge}(t)\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^j t)\right)$$
$$= -it\phi(t)\left(\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^j t)\right).$$

Da $\phi(t)$ nirgends Null wird, formen wir weiter um

$$\begin{aligned}
\frac{(fv)^\wedge}{f^\wedge}(t) &= -it\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^j t) - \sum_{k=1}^{r}\frac{(fv)^\wedge}{f^\wedge}(\alpha^k t) + \frac{(fv)^\wedge}{f^\wedge}(\alpha^r t)\\
&= -it\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^j t) + \alpha it\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^{j+1}) + \frac{(fv)^\wedge}{f^\wedge}(\alpha^r t)\\
&= -it\sum_{k=0}^{N}\left(\alpha^{kr}\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^{j+kr}t) - \alpha^{kr+1}\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^{j+kr+1}t)\right)\\
&\quad + \frac{(fv)^\wedge}{f^\wedge}(\alpha^{(N-1)r}t).
\end{aligned}$$

Der Term $(fv)^\wedge(t)$ konvergiert gegen Null für $t \to 0$. Mit der Annahme (TF) haben wir

$$\begin{aligned}
\sum_{k=0}^{N}\left(\alpha^{kr}\frac{(f\,\mathrm{Id})^\wedge}{f^\wedge}(\alpha^{j+kr}t)\right) &= \sum_{k=0}^{N}\alpha^{kr}\left(h^\wedge(\alpha^{j+kr}t) + \sum_{d=0}^{m-1}B_d(-it\alpha^{j+kr})^d\right)\\
&= \sum_{k=0}^{N}\alpha^{kr}h^\wedge(\alpha^{j+kr}t) + \alpha^{kr}\sum_{d=0}^{m-1}B_d(-it\alpha^{j+kr})^d.
\end{aligned}$$

Die erste Reihe ist nach Lemma 5.2 summierbar, und die zweite Reihe ist eine geometrische Reihe. Also können wir zum Grenzwert übergehen und erhalten

$$\frac{(fv)^\wedge}{f^\wedge}(t) = -it\sum_{k=0}^{\infty}\left(\alpha^{kr}h^\wedge(\alpha^{j+kr}t) - \alpha^{kr+1}h^\wedge(\alpha^{j+kr+1}t)\right) + \sum_{d=1}^{m}C_d(-it)^d$$

für geeignete Zahlen $C_1, \ldots, C_m$. Schreiben wir nun

$$L(x) = \sum_{k=0}^{\infty}h(x/\alpha^{j+kr}) - h(x/\alpha^{j+kr+1})$$

und definieren wir

$$v = (f' * L)/f + \sum_{d=1}^{m}C_d f^{(d)}/f,$$

dann ist v aus $L^1(\varepsilon)$. Ferner ist v auch zentriert und erfüllt damit die Behauptung. □

Mit Lemma 5.5 sehen wir, dass der Fall F_1 eintritt, falls die Dichte f die Annahme (TF) erfüllt. Nun geben wir die Konstruktion eines effizienten Schätzers α_{eff} an.

5.2 Die q sample splitting Methode

Wie wir gesehen haben, ist der kanonische Gradient im Fall F_1 die Funktion

$$(y,z) \to -\frac{1}{r\alpha^{r-1}\operatorname{Var}(Y)\operatorname{Var}(\ell_g(\eta))}\ell_g(z-\alpha^r y)(y-EY).$$

Um nun den effizienten Schätzer zu konstruieren, müssen wir Schätzer für

$$\begin{aligned} 1/W(\alpha) &= r\alpha^{r-1}\operatorname{Var}_\alpha(Y)\operatorname{Var}_\alpha(\ell_g(\eta)) \text{ und} \\ \phi(y,z;\alpha) &= \ell_g(z-\alpha^r y)(y-E_\alpha Y) \end{aligned}$$

angeben. Im allgemeinen sieht der Schätzer für die Einflussfunktion folgendermaßen aus

$$-\widetilde{W}(\alpha, Y_0, \ldots, Y_n)\frac{1}{n}\sum_{j=1}^{n}\widetilde{\phi}(Y_{j-1}, Y_j; \alpha; Y_0, \ldots, Y_n).$$

Das hat aber den technischen Nachteil, dass die beiden Schätzer $\widetilde{W}$ und $\widetilde{\phi}$ nicht unabhängig von den Beobachtungen $(Y_0, \ldots, Y_n)$ sind. Ist $n = qm_n + (q-1)t_n + r_n$ mit $t_n \to \infty$ und $(n - qm_n)/\sqrt{n} \to 0$, und definieren wir

$$\mathbf{Y}_{n,i} = (Y_{(i-1)(m_n+t_n)}, \ldots, X_{im_n+(i-1)t_n}),$$

dann sind die beiden Sequenzen

$$< \mathcal{L}(\mathbf{Y}_{n,1}, \ldots, \mathbf{Y}_{n,q} \mid P_\alpha) > \text{ und } < \bigotimes_{j=1}^{q} \mathcal{L}(\mathbf{Y}_{n,j} \mid P_\alpha) >$$

nach Schick (2001)[Lemma 6.1] benachbart. Demnach können wir benutzen, dass die q Folgen $\mathbf{Y}_{n,1}, \ldots, \mathbf{Y}_{n,q}$ unabhängig sind. Nun definieren wir Schätzer für die Dichte g und deren Ableitung g', wobei g die Dichte der Innovation $< \eta_j >$ ist. Wir folgen nun einer Vorgehensweise von Schick (1987)[3.7. Example 1]. Für eine Nullfolge $< a_n >$ mit $na_n^6 \to \infty$ definieren wir

$$\widetilde{g}_m(x; y_1, \ldots, y_m) = a_m + \frac{1}{ma_m}\sum_{j=1}^{m}\rho\left(\frac{x-y_j}{a_m}\right)$$

und für eine lokal beschränkte Folge $< \alpha_n >$

$$\widetilde{\ell}_g(x; \alpha_n, y_1, \ldots, y_n) = \frac{(\widetilde{g}_{n-1})'}{\widetilde{g}_{n-1}}(x; y_2 - \alpha_n^r y_1, \ldots, y_n - \alpha_n^r y_{n-1}),$$

wobei $\rho(x) = e^{-x}(1+e^{-x})^{-2}$ die logistische Verteilungsdichte ist. Wir definieren weiter

$$\widetilde{W}_n^{-1}(\alpha_n; \mathbf{Y}_{n,1}; \ldots; \mathbf{Y}_{n,q}) = r\alpha_n^{r-1}\left(\frac{1}{m_n}\sum_{j=1}^{m_n}(\mathbf{Y}_{n,1})_j^2 - \left(\frac{1}{m_n}\sum_{j=1}^{m_n}(\mathbf{Y}_{n,2})_j\right)^2\right) \cdot \frac{1}{m_n}\sum_{j=1}^{m_n}\widetilde{\ell}_g^2\left((\mathbf{Y}_{n,3})_{j+1} - \alpha_n^r(\mathbf{Y}_{n,3})_j; \alpha_n; \mathbf{Y}_{n,4}\right),$$

$$\widetilde{\phi}(\alpha_n; \mathbf{Y}_{n,1}; \ldots; \mathbf{Y}_{n,q}) = \frac{1}{m_n}\sum_{j=1}^{m_n}\widetilde{\ell}_g^2\left((\mathbf{Y}_{n,1})_{j+1} - \alpha_n^r(\mathbf{Y}_{n,1})_j; \alpha_n; \mathbf{Y}_{n,2}\right) \cdot \left((\mathbf{Y}_{n,1})_j - \frac{1}{m_n}\sum_{j=1}^{m_n}(\mathbf{Y}_{n,3})_j\right)$$

und den effizienten Schätzer

$$\alpha_{eff} = \widetilde{\alpha} + \frac{1}{q!}\sum_{j_1,\ldots,j_q}\widetilde{W}_{m_n}(\widetilde{\alpha}; \mathbf{Y}_{n,j_1}; \ldots; \mathbf{Y}_{n,j_q})\widetilde{\phi}(\widetilde{\alpha}; \mathbf{Y}_{n,j_1}; \ldots; \mathbf{Y}_{n,j_q})$$

für einen diskretisierten $n^{1/2}$-konsistenten Schätzer $\widetilde{\alpha}$. Der folgende Satz beweist, dass der Schätzer auch effizient ist.

Theorem 5.6. Ist α_{eff} der oben definierte Schätzer für α, dann gilt

$$\alpha_{eff} - \alpha = \frac{1}{n}\sum_{j=1}^{n} W(\alpha)\phi(Y_{i-1}, Y_i; \alpha) + o_P(n^{-1/2}).$$

Sei

$$\widetilde{\alpha}^{\star} = \frac{\sum_{j=1}^{n}(Y_{j+1} - \bar{Y}_n)(Y_j - \bar{Y}_n)}{\sum_{j=1}^{n}(Y_j - \bar{Y}_n)^2}$$

der Kleinste-Quadrate-Schätzer. Dann ist

$$\widetilde{\alpha} = n^{-1/2}\arg\min_{z\in\mathbb{Z}}(|z - n^{1/2}\widetilde{\alpha}^{\star}|)$$

ein diskreter und $n^{1/2}$-konsistenter Schätzer für α.

Beweis. Wir benutzen Schick (2001)[Theorem 7.1] und rechnen dessen Voraussetzungen nach. Wir fixieren nun eine lokal beschränkte Folge $< \alpha_n >$ und zeigen

$$\alpha_n - \alpha + \sum_{j=1}^{n} W(\alpha_n)\phi(Y_{j-1}, Y_j; \alpha_n) - \sum_{j=1}^{n} W(\alpha)\phi(Y_{j-1}, Y_j; \alpha) = o_P(n^{-1/2}).$$

Das ist die Voraussetzung $(A2)$ von Schick (2001). Zusätzlich zum Maß P_α betrachten wir noch das Maß P_{α_n}, dass eindeutig durch den Übergangskern $(y,z) \to g(z-\alpha_n^r y)$ induziert wird. Unter P_{α_n} gilt

$$Y_{n+1} = \alpha_n^r Y_n + \eta_{n+1} \quad \text{mit} \quad \eta_{n+1} = \sum_{j=0}^{r-1} \alpha^j \varepsilon_{r(n+1)-j}.$$

Ferner sind die beiden Sequenzen $< P_\alpha >$ und $< P_{\alpha_n} >$ nach Bemerkung 3.2 benachbart. Also können wir zwischen die beiden Maßen wechseln. Da

$$\begin{aligned} W^{-1}(\alpha+a) =& r(\alpha+a)^{r-1} \operatorname{Var}_{\alpha+a}(Y) \operatorname{Var}_{\alpha+a}(\ell_g(\eta)) \\ =& r(\alpha+a)^{r-1} \frac{\operatorname{Var}_\alpha(\eta)}{1-(\alpha+a)^r} \operatorname{Var}_\alpha(\ell_g(\eta)) \text{ und} \\ E_{\alpha+a} Y =& \frac{E_\alpha \eta}{1-(\alpha+a)^r}, \end{aligned}$$

sind die beiden Abbildungen $a \to W(\alpha+a)$ und $a \to E_{\alpha+a}Y$ stetig, und es reicht zu zeigen, dass

$$\frac{W(\alpha)}{n} \sum_{j=1}^n (\phi(Y_{j+1}, Y_j; \alpha_n) - \phi(Y_{j+1}, Y_j; \alpha)) + (\alpha_n - \alpha) = o_P(n^{-1/2}).$$

Die Abbildung $a \to \phi(y,z;\alpha+a)$ ist nicht differenzierbar. Also müssen wir zuerst die Funktion $a \to \phi(y,z;\alpha+a)$ durch eine Version von differenzierbaren Funktionen annähern. Für beliebig kleine a^{-1} und $b = o(a^{-2})$ definieren wir nun

$$\ell_a = \ell_g \mathbf{1}_{|\ell_g|<a} \quad \text{und} \quad \ell_{a,b}(x) = \int_{\mathbb{R}} \ell_a(x-by)\rho(y)dy$$

mit der logistischen Verteilungsdichte ρ. Unter P_{α_n} ist $(\ell_g - \ell_{a,b})(Y_{j+1} - \alpha_n^r Y_j)(Y_j - E_{\alpha_n} Y)$ eine Martingal-Differenz. Da $E_{\alpha_n}(\ell_g - \ell_{a,b})^2 (Y_1 - \alpha_n^r Y_0)$ gegen Null konvergiert, gilt nach Hall & Heyde (1980)[Theorem 3.2]

$$\frac{1}{n} \sum_{j=1}^n (\ell_g - \ell_{a,b})(Y_{j+1} - \alpha_n^r Y_j)(Y_j - E_{\alpha_n} Y) = o_{P_{\alpha_n}}(n^{-1/2}).$$

Mit demselben Argument haben wir auch

$$\frac{1}{n} \sum_{j=1}^n (\ell_g - \ell_{a,b})(Y_{j+1} - \alpha^r Y_j)(Y_j - E_\alpha Y) = o_{P_\alpha}(n^{-1/2}).$$

Also reduziert sich das Problem auf

$$o_P(n^{-1/2}) = \frac{W(\alpha)}{n}\sum_{j=1}^{n}\left(\ell_{a,b}(Y_{j+1}-\alpha_n^r Y_j)-\ell_{a,b}(Y_{j+1}-\alpha^r Y_j)\right)(Y_j - E_\alpha Y) \qquad (4)$$
$$+ (\alpha_n - \alpha)$$

Ab hier können wir benutzen, dass $\ell_{a,b}$ differenzierbar ist, und erhalten

$$\ell_{a,b}(Y_{j+1}-\alpha_n^r Y_j)-\ell_{a,b}(Y_{j+1}-\alpha^r Y_j) = \int_0^1 \ell'_{a,b}(Y_{j+1}-\alpha^r Y_j + t(\alpha^r-\alpha_n^r)Y_j)dt \cdot (\alpha^r - \alpha_n^r)Y_j.$$

Wir setzen nun abkürzend $c_n = \alpha^r - \alpha_n^r$, $\zeta_n = \max\{|Y_j|\,;\, j = 1,\ldots,n\}$ und $\eta_j = Y_j - \alpha^r Y_{j-1}$, und zeigen

$$\frac{1}{n}\sum_{j=1}^{n}\int_0^1 \left(\ell'_{a,b}(\eta_{j+1}+tc_nY_j)-\ell'_{a,b}(\eta_{j+1})\right)dt\cdot Y_j(Y_j - E_\alpha Y) = o_{P_\alpha}(1).$$

Es gilt

$$\left|\frac{1}{n}\sum_{j=1}^{n}\int_0^1 \left(\ell'_{a,b}(\eta_{j+1}+tc_nY_j)-\ell'_{a,b}(\eta_{j+1})\right)dt\cdot Y_j(Y_j - E_\alpha Y)\right| \le$$
$$\left|\frac{c_n}{n}\sum_{j=1}^{n}\int_0^1\int_0^1 t\ell''_{a,b}(\eta_{j+1}+stc_nY_j)dtds\cdot Y_j^2(Y_j - E_\alpha Y)\right| \le$$
$$\left|c_n\frac{a}{b}\right|\|\rho''\|_{L^1}\left|\frac{1}{n}\sum_{j=1}^{n}Y_j^2(Y_j - E_\alpha Y)\right| \le \left|\zeta_n c_n\frac{a}{b}\right|\|\rho''\|_{L^1}\frac{1}{n}\sum_{j=1}^{n}|Y_j(Y_j - E_\alpha Y)|,$$

da $|\ell_{a,b}(x)| \le |a/b|\|\rho''\|_{L^1}$. Die Zufallsvariablen $<\zeta_n>_n$ erfüllen die Eigenschaft, dass $n^{-1/2}\zeta_n = o_P(1)$. Also existiert eine Folge $d_n^{-1}\to\infty$ mit $d_n^{-1}n^{-1/2}\zeta_n = O_P(1)$. Wählen wir nun a so, dass $d_n a/b = o(1)$, dann ist $\left|\zeta_n c_n\frac{a}{b}\right|\frac{1}{n}\sum_{j=1}^{n}|Y_j(Y_j - E_\alpha Y)| = o_P(1)$. Setzen wir das Ergebnis in (4) ein, dann erhalten wir

$$\frac{W(\alpha)}{n}\sum_{j=1}^{n}\left(\ell_{a,b}(Y_{j+1}-\alpha_n^r Y_j)-\ell_{a,b}(Y_{j+1}-\alpha^r Y_j)\right)(Y_j - E_\alpha Y) =$$
$$(\alpha^r - \alpha_n^r)\frac{W(\alpha)}{n}\sum_{j=1}^{n}\ell'_{a,b}(\eta_{j+1})Y_j(Y_j - E_\alpha Y).$$

Für a gegen Null haben wir

$$E_\alpha\ell'_{a,b}(\eta) = \int \ell_{a^{-1}}(x)\rho(y)g'(x+yb)dxdy \to E\ell_g^2(\eta).$$

Mit Hall & Heyde (1980)[Theorem 2.19] erhalten wir dann

$$\begin{aligned}(\alpha^r - \alpha_n^r)\frac{W(\alpha)}{n}\sum_{j=1}^{n} \ell'_{a,b}(\eta_{j+1})Y_j(Y_j - E_\alpha Y) =&(\alpha^r - \alpha_n^r)W(\alpha)E\ell_g^2(\eta)\operatorname{Var}_\alpha Y \\ &+ o_{P_\alpha}(n^{-1/2}) \\ =&(\alpha^r - \alpha_n^r)/(r\alpha^{r-1}) + o_{P_\alpha}(n^{-1/2}),\end{aligned}$$

und somit die Behauptung

$$\alpha_n - \alpha + \sum_{j=1}^{n} W(\alpha_n)\phi(Y_{j-1}, Y_j; \alpha_n) - \sum_{j=1}^{n} W(\alpha)\phi(Y_{j-1}, Y_j; \alpha) = o_P(n^{-1/2}).$$

Sei nun $\tilde{Y}_1, \ldots, \tilde{Y}_n$ auch eine Markov-Kette mit derselben Übergangsverteilung wie $Y_1, \ldots, Y_n$, die unabhängig von $Y_1, \ldots, Y_n$ ist. Unter dem Maß P_{α_n} sind die Zufallsvariablen

$$< Y_{i+1} - \alpha_n^r Y_i >$$

unabhängig und identisch verteilt mit Dichte g, die unabhängig von α_n ist. Schreiben wir $\hat{\eta}_{i+1} = Y_{i+1} - \alpha_n^r Y_i$, dann erhalten wir mit Schick (1987)[Example 3.7], dass

$$E_{\alpha_n}\left(\widetilde{\ell}_g(\hat{\eta}_1; \alpha_n; \tilde{Y}_1, \ldots, \tilde{Y}_n) - \ell_g(\hat{\eta}_1)\right)^2. \tag{5}$$

Ferner haben wir

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n} \widetilde{\ell}_g^2(\hat{\eta}_{i+1}; \alpha_n; \tilde{Y}_1, \ldots, \tilde{Y}_n) =&\frac{1}{n}\sum_{j=1}^{n}\left(\widetilde{\ell}_g(\hat{\eta}_{i+1}; \alpha_n; \tilde{Y}_1, \ldots, \tilde{Y}_n) - \ell_g(\hat{\eta}_{i+1})\right)^2 \\ &+ 2\frac{1}{n}\sum_{j=1}^{n}\left(\widetilde{\ell}_g(\hat{\eta}_{i+1}; \alpha_n; \tilde{Y}_1, \ldots, \tilde{Y}_n) - \ell_g(\hat{\eta}_{i+1})\right)\ell_g(\hat{\eta}_{i+1}) \\ &+ \frac{1}{n}\sum_{j=1}^{n} \widetilde{\ell}_g^2(\hat{\eta}_{i+1}).\end{aligned}$$

Wegen Gleichung (5), haben wir dann

$$\frac{1}{n}\sum_{j=1}^{n} \widetilde{\ell}_g^2(\hat{\eta}_{i+1}; \alpha_n; \tilde{Y}_1, \ldots, \tilde{Y}_n) = E\ell_g^2(\eta) + o_{P_{\alpha_n}}(1).$$

Die Voraussetzungen von Schick (2001)[Theorem 7.1] kann man nun leicht nachrechnen. □

6 Ausblick

In diesem Kapitel diskutieren wir nun, wie wir die erbrachten Ergebnisse auf andere Zeitreihen übertragen können. Dazu betrachten wir die ARMA(p,q)-Zeitreihe. Dann untersuchen wir, ob für den Faltungsschätzer $\hat{f} * \hat{f}$ in einem T_2-Modell die Rate $n^{1/2}$ erreicht werden kann.

6.1 Erweiterung auf die ARMA(p,q)-Reihe

Die Zeitreihe $< X_n >$ heißt eine ARMA(p,q)-Zeitreihe, falls für alle n

$$X_n - \sum_{j=1}^{p} \tau_j X_{n-j} = \varepsilon_n + \sum_{j=1}^{q} \phi_j \varepsilon_{n-j}$$

gilt, wobei wir hier annehmen, dass $< \varepsilon_n >$ unabhängig und identisch verteilt ist. Ferner nehmen wir an, dass die Reihe $< X_n >$ kausal und invertierbar ist. Ist nun B der Lagoperator, dann können wir die obige Beziehung kompakter schreiben als

$$\tau(B)X_n = \phi(B)\varepsilon_n,$$

wobei natürlich $\tau(z) = 1 - \sum_{j=1}^{p} \tau_j z^j$ und $\phi(z) = 1 + \sum_{j=1}^{q} \phi_j z^j$. Ist F_ε die Verteilungsfunktion von ε, dann ist der Prozess eindeutig durch die Angabe von $(\tau, \phi, F_\varepsilon)$ bestimmt. Für eine feste Zahl r aus $\mathbb{N}$ behaupten wir nun:

Lemma 6.1. Die Zeitreihe $< X_{rn} >_n$ ist weiterhin eine ARMA-Zeitreihe.

Beweis. Das Polynom τ zerfällt in ein Produkt $\tau(z) = \prod_{k=0}^{p}(1 - a_k z)$ mit $a_i \in \mathbb{C}$. Multiplizieren wir mit $\hat{\tau}(z) = \prod_{k=0}^{p} \sum_{j=0}^{r-1} (a_k z)^j$, dann erhalten wir

$$\hat{\tau}(z)\tau(z) = \prod_{k=0}^{p}(1 - a_k^r z^r).$$

Man bemerke, dass $\hat{\tau}$ ein reelles Polynom ist. Die rechte Seite ist ein Polynom mit reellen Koeffizienten, dass die Darstellung $\hat{\tau}(z)\tau(z) = 1 - \sum_{k=1}^{p} b_k z^{rk}$ hat. Es folgt

$$\hat{\tau}(B)\tau(B)X_{rn} = \hat{\tau}(B)\phi(B)\varepsilon_{rn}$$

und

$$\hat{\tau}(B)\tau(B)X_{rn} = X_{rn} - \sum_{k=1}^{p} b_k X_{r(n-k)}.$$

$\square$

Die Zeitreihe $< X_{rn} >_n$ ist nach der obigen Darstellung nicht mehr kausal und invertierbar. Definieren wir die Innovationen $\eta_j = \hat{\tau}(B)\varepsilon_{rj}$, dann ist die Zeitreihe $< X_{rn} >$ bezüglich $< \eta_j >$ kausal und invertierbar, da

$$\hat{\tau}(B)\tau(B)X_{rn} = \phi(B)\eta_n.$$

Ein Nachteil haben die Innovationen $< \eta_j >$ trotzdem, sie sind i. Allg. nicht unabhängig verteilt.

Da $< X_n >_n$ eine ARMA-Zeitreihe ist, hat die Autokovarianzfunktion nach Brockwell & Davis (1987)[Theorem 4.4.2] eine rationale Spektraldichte. Nach Kolmogorov & Rozanov (1960) ist dann $< X_n >_n$ ρ-mischend mit $\rho(n) \leq e^{-cn}$ für ein $c > 0$. Nach Definition von ist $< X_{rn} >_n$ auch ρ-mischend. Mit dieser Erkenntnis, können wir Schätzer für die Koeffizienten von $\hat{\tau}\tau$ und ϕ definieren und zeigen, dass sie konsistent sind. Trotzdem ist es schwierig, die Koeffizienten von τ zu schätzen, da wir zuerst die Nullstellen von $\hat{\tau}\tau$ herausfinden müssen und erst dann die Koeffizienten von τ ausrechnen können.

Da wir in der Lage sind, die Polynome $\hat{\tau}\tau$ und ϕ zu schätzen, können wir auch die Residuen von η definieren und die Dichte von η schätzen, falls η eine Dichte hat. Danach können wir die Neumann-Reihe aus Kapitel 2.3 benutzen, um die Parameterdichte von ε zu schätzen. Diese benutzt die Koeffizienten von $\hat{\tau}$, und deshalb muss das die Rate verschlechtern.

6.2 Faltung von Dekonvolutionsschätzern

Seien $< Z_n >_n$ die Realisationen einer unabhängig und identisch verteilten Zufallsvariable mit Dichte h_Z. Dann ist

$$\hat{h}_Z(z) = \frac{1}{n}\sum_{j=1}^{n} K_b(z - Z_j)$$

mit Bandbreite $b \to 0$, Kern K und $K_b(\cdot) = K(\cdot/b)/b$ ein Schätzer für die Dichte h_Z im Punkt z. Dieser ist i. Allg. nicht $n^{1/2}$-konsistent. Integriert man den Kernschätzer mit einer bekannten Funktion τ, dann ist $\int \hat{h}_Z(z)\tau(z)dz$ $n^{1/2}$-konsistent. Integriert man einen Kernschätzer mit einer bekannten Funktion, dann verbessert das die Rate. Nimmt man dagegen einen Dekonvolutionsschätzer, dann untersuchen wir nun, ob mit dem integrierten Dekonvolutionsschätzer die optimale Rate erreicht oder nicht. Dabei demonstrieren wir diese Tatsache anhand des Dekonvolutionsschätzers aus Fan (1991).

Lemma 6.2. Sei $\hat{f}_X$ der Dekonvolutionsschätzer aus Fan (1991) mit beobachtbaren Realisationen $< Y_n >$ und

$$\hat{f}_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \exp(-itx) \phi_K(th_n) \frac{\hat{\phi}_Y}{\phi_\varepsilon}(t) dt.$$

Ist die Funktion $t \to (f_X * \tau(-\cdot))(t)$ im Punkt $t = 0$ Lipschitz-stetig, und die Funktion $|\phi_\tau / \phi_\varepsilon|^2$ integrierbar, dann haben wir

$$\sup_n \operatorname{Var}(n^{1/2} \int (\hat{f}_X - f_X)(t)\tau(t)dt) < \infty,$$

bei geeigneter Wahl des Kerns K.

Beweis. Da $|\phi_\varepsilon| \leq 1$, ist $|\phi_\tau|^2$ auch integrierbar, also haben wir

$$\int f_X(z+t)\tau(t)dt = \frac{1}{2\pi} \int \exp(-itz)\phi_X(t)\phi_\tau(-t)dt.$$

Setzen wir $z = 0$ ein, dann erhalten wir

$$\begin{aligned} n^{1/2} \int (\hat{f}_X - f_X)(t)\tau(t)dt =& \frac{1}{2\pi} \int \phi_K(th_n) \frac{\hat{\phi}_Y - \phi_Y}{\phi_\varepsilon}(t)\phi_\tau(-t)dt \\ &+ \frac{1}{2\pi} \int (\phi_K(th_n) - 1)\phi_X(t)\phi_\tau(-t)dt. \end{aligned}$$

Nach der bekannten Gleichung erhalten wir

$$\begin{aligned} \operatorname{Var}(n^{1/2} \int (\hat{f}_X - f_X)(t)\tau(t)dt) =& n \operatorname{Var}(\int \hat{f}_X(t)\tau(t)dt)) \\ &+ n \left(\int_{\mathbb{R}} g(-h_n t)K(t)dt - g(0) \right)^2. \end{aligned}$$

Hier ist $g(x) = (f_X * \tau(-\cdot))(x)$. Da die Realisationen $< Y_i >$ unabhängig und identisch verteilt sind, kann man die Varianz vereinfachen, und es gilt

$$n \operatorname{Var}(\int \hat{f}_X(t)\tau(t)dt)) = \int \frac{1 - |\phi_Y|^2}{|\phi_\varepsilon|^2} |\phi_\tau|^2(t) |\phi_K|^2(th_n) dt.$$

Mit der Voraussetzung erhalten wir das Resultat. □

Das obige Lemma soll zeigen, dass die Verbesserung der Rate durch das Integrieren einer bekannten Funktion davon abhängt, welche Güte die beiden bekannten Funktionen haben. In diesem Fall muss gelten, dass

$$\left| \frac{\phi_\tau}{\phi_\varepsilon} \right| (t) = O(t^{-1/2-\delta})$$

für ein $\delta > 0$.

In einem integrierten Dekonvolutionsschätzer haben wir sowohl eine Funktion, die die Rate verschlechtert, als auch eine Funktion, die die Rate verbessert. Hier konkurrieren zwei Gegenspieler um die Rate. Überwiegt die Verbesserung, dann kann man die optimale Rate erreichen. Beim Dekonvolutionsschätzer kann man nur von einer Verschlechterung der Rate reden, da wir keinen Gegenspieler haben, der die Rate wieder verbessern kann.

Nun wollen wir eine Aussage über den Faltungsschätzer $\hat{f} * \hat{f}$ machen, wobei $\hat{f}$ der Dekonvolutionsschätzer aus Kapitel 2 ist. Dieser zerfällt in die zwei Terme

$$\hat{f} * \hat{f} - f * f = (\hat{f} - f) * (\hat{f} - f) + 2(\hat{f} - f) * f.$$

Während der erste Term mit der optimalen Rate gegen Null konvergieren kann, hängt die Rate des letzten Terms davon ab, in welcher Funktionsklasse die Dichte f liegt. Existieren positive Konstanten ν, μ und G, so dass für die Fouriertransformation ϕ_f der Dichte f die Abschätzung

$$d_0|t|^{-\nu} \le |\phi_f|(t) \le d_1|t|^{-\mu} \text{ für alle } |t| > G$$

gilt, dann folgt

$$\left| \frac{\phi_f(t)}{\prod_{k=1}^{r} \phi_f(\alpha^k t)} \right| \le_* |t|^{r\nu - \mu},$$

was nicht integrierbar ist. Darum können wir nicht erwarten, dass der Term $(\hat{f} - f) * f$ die optimale Rate hat.

Gilt dagegen die Abschätzung $\exp(-|t|^{\beta}\gamma_1) \le |\phi_f|(t) \le d \exp(-|t|^{\beta}\gamma_2)$ für $|t| > G$ für die Fouriertransformation ϕ_f der Dichte f, dann kann man die optimale Rate erreichen, da

$$\left| \frac{\phi_f(t)}{\prod_{k=1}^{r} \phi_f(\alpha^k t)} \right| \le_* \exp\left(-|t|^{\beta}\left(\gamma_2 - \gamma_1|\alpha|^r/(1 - |\alpha|^r)\right)\right).$$

Die rechte Seite ist quadratintegrierbar, falls $\gamma_2 - \gamma_1|\alpha|^r/(1 - |\alpha|^r) > 0$.

6.3 Konstruktion eines effizienten Schätzers ohne strukturelle Annahme

Im Fall F_3 ist der kanonische Gradient die Funktion

$$g_0(Y, Z) = -\frac{r\alpha^{r-1}(Y - EY)\ell_g(\eta) + \aleph(\eta) - \mathrm{Proj}(\aleph(\eta))}{\|r\alpha^{r-1}(Y - EY)\ell_g(\eta) + \aleph(\eta) - \mathrm{Proj}(\aleph)(\eta)\|_2^2}.$$

Die Fisher-Information ℓ_g ist durch die Residuen $\eta_{i,n} = Z_i - \hat{\alpha}^r Y_i$ schätzbar. Da die beiden Familien $< P_\alpha >$ und $< P_{\alpha_n} >$ für alle beschränkten Folge $< \alpha_n >$ benachbart sind, können wir so tun, als ob $\hat{\alpha}$ eine beschränkte Folge $< \alpha_n >$ wäre. Dann können wir von P_α zu P_{α_n} übergehen.

Wir diskutieren nun, wie wir die Funktion $\aleph(\eta)$ schätzen können. Die Funktion $E[\ell_r \varepsilon_j \,|\, \eta = x]$ hat die Darstellung

$$E[\ell_r \varepsilon_j \,|\, \eta = x] = g^{-1}(x) \int f'(x - \sum_{k=1}^{r-1} \alpha^k y_{r-k}) \prod_{k=1}^{r-1} f(y_k) y_j dy_k.$$

Hier ist g die Dichte von η. Mit dem Dekonvolutionsschätzer aus Kapitel 2 sind wir in der Lage, die Funktion $E[\ell_r \varepsilon_j \,|\, \eta = x]$ zu schätzen.

Die ganze Schwierigkeit besteht darin, die Funktion $\mathrm{Proj}\,(\aleph)$ zu schätzen. Da wir keine explizite Form der Funktion $\mathrm{Proj}\,(\aleph)$ kennen, versuchen wir sie durch eine Reihe zu approximieren. Wir wählen dazu als Basis von $L^2(\lambda)$ die Hermite-Funktionen

$$h_j(x) = \frac{H_j(x) e^{-x^2/2}}{(2^j j! \pi^{1/2})^{1/2}},$$

wobei $H_j(x)$ die Hermite-Polynome sind (Siehe Szegő (1975)). Dann bilden die Funktionen

$$x_j = \frac{h_j}{\sqrt{f}}(x)$$

eine Basis in $L^2(\varepsilon)$. Also spannt $\Gamma(x_1), \Gamma(x_2), \ldots$ den Raum $\overline{\{\Gamma(u)\,;\, u \in L^2(\varepsilon)\}}$ auf. Definieren wir nun eine Orthonormalbasis $\{e_j\}_j$ durch $e_1 = \Gamma(x_1)/\|\Gamma(x_1)\|$,

$$\tilde{e}_n := \Gamma(x_n) - \sum_{j=1}^{n-1} (\Gamma(x_n), e_j) e_j\,,$$

$$e_j = \begin{cases} 0 & \text{falls } \tilde{e}_j = 0 \\ \tilde{e}_j / \|\tilde{e}_j\| & \text{sonst,} \end{cases}$$

dann gilt

$$\mathrm{Proj}(E[\ell_r \varepsilon_j \,|\, \eta]) = \sum_{k=0}^{\infty} \lfloor E[\ell_r \varepsilon_j \,|\, \eta], e_k \rfloor e_k,$$

falls die Funktion $\mathrm{Proj}(E[\ell_r \varepsilon_j \,|\, \eta])$ im Raum $\overline{\{\Gamma(u)\,;\, u \in L^2(\varepsilon)\}}$ liegen. Die Koeffizienten

$\lfloor E[\ell_r \varepsilon_j \,|\, \eta], e_k \rfloor$ vereinfachen sich zu $\lfloor \ell_r \varepsilon_j, e_k \rfloor$ und sind demnach schätzbar.

Literatur

Belomestny, D. 2003. Rates of convergence for constrained deconvolution problem. *ArXiv Mathematics e-prints*, June.

Bradley, Richard C. 1999. On the growth of variances in a central limit theorem for strongly mixing sequences. *Bernoulli*, **5**(1), 67–80.

Brockwell, Peter J., & Davis, Richard A. 1987. *Time series: theory and methods*. Springer Series in Statistics. New York: Springer-Verlag.

Devroye, Luc. 1994. On the non-consistency of an estimate of Chiu. *Statist. Probab. Lett.*, **20**(3), 183–188.

Doukhan, Paul, León, José, & Portal, Frédéric. 1984. Vitesse de convergence dans le théorème central limite pour des variables aléatoires mélangeantes à valeurs dans un espace de Hilbert. *C. R. Acad. Sci. Paris Sér. I Math.*, **298**(13), 305–308.

Elstrodt, Jürgen. 2004. *Maß- und Integrationstheorie*. Springer, Berlin.

Fan, Jianqing. 1991. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**(3), 1257–1272.

Greenwood, P. E., & Wefelmeyer, W. 1995. Efficiency of empirical estimators for Markov chains. *Ann. Statist.*, **23**(1), 132–143.

Hájek, Jaroslav, Šidák, Zbyněk, & Sen, Pranab K. 1999. *Theory of rank tests*. Second edn. Probability and Mathematical Statistics. San Diego, CA: Academic Press Inc.

Hall, P., & Heyde, C. C. 1980. *Martingale limit theory and its application*. New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics.

Höpfner, Reinhard, Jacod, Jean, & Ladelli, Lucia. 1990. Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields*, **86**(1), 105–129.

Kolmogorov, A. N., & Rozanov, Ju. A. 1960. On a strong mixing condition for stationary Gaussian processes. *Teor. Verojatnost. i Primenen.*, **5**, 222–227.

Koul, Hira L., & Schick, Anton. 1997. Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli*, **3**(3), 247–277.

Lin, Zhengyan, & Lu, Chuanrong. 1996. *Limit theory for mixing dependent random variables*. Mathematics and its Applications, vol. 378. Dordrecht: Kluwer Academic Publishers.

Lukacs, Eugene. 1970. *Characteristic functions*. Hafner Publishing Co., New York. Second edition, revised and enlarged.

Meyn, S. P., & Tweedie, R. L. 1993. *Markov chains and stochastic stability*. Communications and Control Engineering Series. London: Springer-Verlag London Ltd.

Müller, Ursula U., Schick, Anton, & Wefelmeyer, Wolfgang. 2007. *Statistical Models and Methods for Biomedical and Technical Systems (F. Vonta, M. Nikulin, N. Limnios and C. Huber, eds.)*. Birkhäuser, Boston. Chap. Estimators for partially observed Markov chains.

Prakasa Rao, B. L. S. 1983. *Nonparametric functional estimation*. Probability and Mathematical Statistics. New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers].

Roussas, George G. 1972. *Contiguity of probability measures: some applications in statistics*. London: Cambridge University Press. Cambridge Tracts in Mathematics and Mathematical Physics, No. 63.

Schick, A., & Wefelmeyer, W. 2002. Efficient estimation in invertible linear processes. *Math. Methods Statist.*, **11**(3), 358–379 (2003).

Schick, Anton. 1987. A note on the construction of asymptotically linear estimators. *J. Statist. Plann. Inference*, **16**(1), 89–105.

Schick, Anton. 2001. Sample splitting with Markov chains. *Bernoulli*, **7**(1), 33–61.

Szegő, Gábor. 1975. *Orthogonal polynomials*. Fourth edn. Providence, R.I.: American Mathematical Society. American Mathematical Society, Colloquium Publications, Vol. XXIII.

Ushakov, Nikolai G. 1999. *Selected topics in characteristic functions*. Modern Probability and Statistics. Utrecht: VSP.

van der Vaart, A. W. 1998. *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, vol. 3. Cambridge: Cambridge University Press.

Wefelmeyer, Wolfgang. 1999. Efficient estimation in Markov chain models: an introduction. *Pages 427–459 of: Asymptotics, nonparametrics, and time series*. Statist. Textbooks Monogr., vol. 158. New York: Dekker.

Wefelmeyer, Wolfgang. 2007. *Vorlesung über Statistik für Zeitreihen, Wintersemester 2006/2007*. http://www.mi.uni-koeln.de/ wefelm/.

*Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Herrn Prof. Dr. Wolfgang Wefelmeyer betreut worden.

Teilpublikation: keine

*Lebenslauf

Kim Kuen Tang

geboren am 04.07.1979 in Rheinbach

Persönliche Daten

Name:	Tang
Vorname:	Kim Kuen
Geburtsdatum:	04.07.1979
Geburtsort:	Rheinbach
Familienstand:	verheiratet, ein Sohn(4) und eine Tochter(3)
Staatsangehörigkeit:	deutsch

Hochschulausbildung

2007 Promotion in Mathematik, Universität zu Köln. (Erwarteter Abschluss im Dezember 2007)

2004 Diplom-Mathematik, Universität zu Köln.

2001 Vordiplom in Mathematik, Universität zu Köln.

1999 Aufnahme des Studiums in Mathematik, Universität zu Köln.

Diplomarbeit

Titel:	Rekurrenz von katalytischen Verzweigungsprozessen, Universität zu Köln, Wintersemester 2003/2004.
Betreuer:	Prof. Dr. Achim Klenke (Universität Mainz)

Schulbildung

1999 allgemeine Hochschulreife, Deutzer Gymnasium Schaurtestraße.

1996 Fachoberschulreife, Otto-Hahn-Realschule Bensberg.

1990 Grundschule, Katholische Grundschule Frankenforst in Refrath.

1986 Einschulung

Berufserfahrung

2004–2007 wissenschaftlicher Mitarbeiter an der Universität zu Köln, Fachbereich: Mathematische Statistik bei Herrn Prof. Dr. Wolfgang Wefelmeyer.

2001–2004 studentische Hilfskraft an der Universität zu Köln, eigenständige Leitung von Übungen in Analysis I und II, Algorithmische Mathematik, Stochastik 0 und I und Funktionentheorie.