DEUTSCHE GEODÄTISCHE KOMMISSION

bei der Bayerischen Akademie der Wissenschaften

Reihe A | Theoretische Geodäsie | Heft Nr. 120

Annette Teusch

Einführung in die Spektral- und Zeitreihenanalyse mit Beispielen aus der Geodäsie

München 2006

Verlag der Bayerischen Akademie der Wissenschaften

in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0938-2836 | ISBN 3 7696 8200 9

DEUTSCHE GEODÄTISCHE KOMMISSION

bei der Bayerischen Akademie der Wissenschaften

Reihe A | Theoretische Geodäsie | Heft Nr. 120

Annette Teusch

Einführung in die Spektral- und Zeitreihenanalyse mit Beispielen aus der Geodäsie

München 2006

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission bei der C. H. Beck'schen Verlagsbuchhandlung München

ISSN 0938-2836 ISBN 3 7696 8200 9

Adresse des Herausgebers /
address of the publisher

Deutsche Geodätische Kommission

Alfons-Goppel-Straße 11
D – 80 539 München

Telefon +49 - (0)89 - 23 031 - 1113
Telefax +49 - (0)89 - 23 031 - 1283/ - 1100
E-mail hornik@dgfi.badw.de
http://dgk.badw.de

Diese Publikation ist als pdf-Dokument veröffentlicht im Internet unter der Adresse /
This volume is published in the internet

http://dgk.badw.de

ISSN 0938-2836

ISBN 3 7696 8200 9

Vorwort

Geodätische Messdaten können vielfach als Zeitreihen aufgefasst werden, die mit den Methoden der Spektral- und Zeitreihenanalyse bearbeitbar sind. Im Besonderen gilt dies beispielsweise für die Auswertung von GPS-Beobachtungen, die auf Permanentstationen oder im Rahmen von Messkampagnen im stationären ("statischen") Modus durchgeführt werden. Die aus der Ausgleichung solcher Daten hervorgehenden Residuen der Phasenmessungen sind oft nicht normalverteilt, sondern enthalten systematische und zeitlich variierende stochastische Komponenten, deren Ursachen in einer unzureichenden mathematischen Modellbildung zu suchen sind. Mittels der Residuenzeitreihen ist es andererseits möglich, die mathematischen Modellansätze zu überprüfen und die deterministischen und stochastischen Modellkomponenten weiter zu analysieren und ggf. zu verbessern.

Nachdem in der Vergangenheit dem funkionalen Modell der GPS-Beobachtungen die meiste Aufmerksamkeit geschenkt wurde, rückt in den letzten Jahren die Verbesserung des stochastischen Modells von GPS-Beobachtungen stärker in den Mittelpunkt. Das stochastische Modell von GPS-Beobachtungen, das im Wesentlichen durch ihre Kovarianzmatrix repräsentiert wird, ist ein vielschichtiges und derzeit kontrovers diskutiertes Thema. Die Auswertung mit der üblichen Standardsoftware beruht i.Allg. auf der Approximation der Kovarianzmatrix der ursprünglichen GPS-Phasenmessungen durch eine skalierte Einheitsmatrix. Dieser Fall entspricht - stochastisch gesehen - der Annahme zeitlich und räumlich unkorrelierter Beobachtungen mit gleicher Varianz, sodass die Daten eines jeden Satelliten (bzw. eines jeden Satellitenpaares nach Bildung der Doppeldifferenzen) als unkorrelierte Zeitreihe mit konstanter Varianz betrachtet werden. Neuere Untersuchungen weisen darauf hin, dass dieses vereinfachte Modell homoskedastischer Zeitreihen - besonders vor dem Hintergrund hochpräziser geodätischer Anwendungen - zu unrealistischen Genauigkeitsaussagen sowie zu systematischen Fehlern in den geschätzten Parametern führt.

Das vom Unterzeichner zusammen mit Herrn Prof. Dr. Wolfgang Bischoff vom Institut für Mathematische Stochastik der Universität Karlsruhe (jetzt: Fakultät für Mathematik und Geographie der Katholischen Universität Eichstätt-Ingolstadt) initiierte Forschungsprojekt "Geokinematische Analysen in einem regionalen GPS-Netz mit stochastischen Modellansätzen" hatte primär das Ziel das herkömmliche stochastische Modell der GPS-Trägerphasenbeobachtungen weiterzuentwickeln. Eine erste Verbesserung gelingt mittels einer elevationsabhängigen Gewichtung der Phasenmessungen, welche die Diagonalstruktur der Kovarianzmatrix beibehält. In diesem Modell sind die Beobachtungen noch immer unkorreliert, haben nun aber unterschiedliche Varianzen. Noch realitätsnäher ist die Vorgabe einer vollbesetzten Kovarianzmatrix, deren Elemente aus einer einheitlichen Kovarianzfunktion abgeleitet sind, welche die - beispielsweise durch die Erdatmosphäre erzeugten - physikalischen Korrelationen zwischen den Beobachtungen beschreibt. Beim Übergang vom einfacheren zum jeweils komplexeren stochastischen Modell sind zwei Probleme zu lösen: Einerseits sind gewisse Modellparameter (Varianz- bzw. Kovarianzfunktionen) anzupassen und andererseits ist die Wirksamkeit der Modellverbesserung anhand statistischer Tests zu beurteilen.

Mit beiden Problemkreisen befasste sich das o.g. interdisziplinäre Forschungsvorhaben, wobei die enge Kooperation zwischen Geodäten und Mathematikern zu weitreichenden Ergebnissen führte. Im Laufe des Projekts wurde von der Mitarbeiterin Frau Dipl.-Math. Annette Teusch ein reicher Methodenschatz zur Zeitreihenanalyse und Spektralanalyse sowie zu Schätz- und Testverfahren zusammengetragen und in Form eines Berichts aufbereitet. Da diese Methoden nicht nur auf GPS-Messungen sondern auch für andere geodätische Zeitreihen anwendbar sind, entstand der Gedanke dieses Manuskript zu publizieren und damit einem weiteren Kreis von Interessenten zugänglich zu machen.

Die vorliegende Monographie hat das Ziel, den Leser in die Theorie der Zeitreihenanalyse einzuführen und ihm einen Einblick in die Methodenvielfalt dieses Gebietes zu geben. Die moderne Theorie der Zeitreihenanalyse kann grundsätzlich in drei Bereiche eingeteilt werden. Während (1) die klassische Zeitreihenanalyse die Zeitreihe auf der Zeitachse betrachtet und die Autokorrelationsfunktion im Zentrum steht, befasst sich (2) die Spektralanalyse mit der Betrachtung der Zeitreihe im Frequenzbereich, wobei die Fourier-Transformierte eine zentrale Rolle spielt. Eine Synthese dieser beiden Bereiche wird durch (3) die Wavelet-Analyse hergestellt, welche die Zeitreihen- und Spektralanalyse miteinander verknüpft, indem die "Frequenz in Abhängigkeit von der Zeit" betrachtet wird. Der Schwerpunkt der Monographie liegt auf den ersten beiden Bereichen, während die Wavelet-Analyse nur beispielhaft in den Abschnitten 2 und 3.1.4 angerissen wird. Da die Fourier- bzw. Waveletanalyse für Zeitreihen auf der Fourier- bzw. Wavelet-Theorie für deterministische Signale aufbaut, wird zunächst in die entsprechenden Theorien für deterministische Funktionen eingeführt. Eine zentrale Stellung nehmen statistische Tests - insbesondere zur Beurteilung der Homoskedastizität von Zeitreihen - ein.

Für das Verständnis wird mathematisches Wissen vorausgesetzt, wie es in etwa in den Vorlesungen zur Höheren Mathematik für Ingenieure präsentiert wird; darüber hinaus wird von der Kenntnis der grundlegenden Begriffe

aus der Wahrscheinlichkeitstheorie ausgegangen. Weiterführende Begriffe aus der Funktionalanalysis, Fourier-Theorie und Stochastik werden im Anhang erläutert. Die Art und Weise der Präsentation stellt einen Kompromiss zwischen mathematischer Strenge und Anwendungsbezug dar; Beweise werden nur gebracht, wenn sie zum Verständnis der Sache beitragen und kurz sind. Beispiele zur Zeitreihenanalyse von GPS-Daten dienen der Verdeutlichung des behandelten Stoffes. Die Auswahl der Themen aus dem schier unübersehbaren Bereich der Zeitreihenanalyse orientiert sich an dem Themengebiet des o.g. Forschungsprojekts, greift aber auch teilweise auf benachbarte Gebiete über.

Dank gebührt den Mitgliedern der Arbeitsgruppe: Frau Dipl.-Math. Annette Teusch und Herrn Prof. Dr. Wolfgang Bischoff vom Institut für Mathematische Stochastik sowie Herrn Dr.-Ing. Jochen Howind vom Geodätischen Institut der Universität Karlsruhe und Herrn Prof. Dr.-Ing. Hansjörg Kutterer vom Geodätischen Institut der Universität Hannover. Die Deutsche Forschungsgemeinschaft (DFG) hat das Forschungsvorhaben von Juni 2000 bis August 2004 finanziell gefördert (HE 1433/11). Schließlich sei der DGK sehr herzlich für die Publikation in der Veröffentlichungsreihe A gedankt.

Karlsruhe, im Juli 2006 Bernhard Heck

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Kapitel 1. Fourierentwicklung

1.1 Fourier-Reihen periodischer Funktionen

Die Entwicklung einer periodischen Funktion φ in eine Fourier-Reihe entspricht ihrer Darstellung als (unendliche) Summe von Sinus- und Cosinus-Funktionen unterschiedlicher Frequenz und Amplitude. Zur Klärung der Frage, unter welchen Bedingungen bzw. in welchem Sinne diese Reihe (als Folge der Teilsummen) gegen die Funktion φ konvergiert, soll zunächst der Begriff der *beschränkten Variation* eingeführt werden.

1.1.1 Definition: Eine Funktion φ heißt *von beschränkter Variation auf* $[a,b]$, falls eine Konstante M existiert, so dass für jede Zerlegung $a = x_0 \leq x_1 \leq \cdots \leq x_n = b$ des Intervalls $[a,b]$ stets gilt

$$\sum_{k=1}^{n} |\varphi(x_k) - \varphi(x_{k-1})| \leq M.$$

Anschaulich gesprochen ist eine Funktion genau dann von beschränkter Variation auf einem bestimmten Intervall, wenn ihre Werte dort nur begrenzt schwanken. Insbesondere ist eine Funktion φ auf einem Intervall $[a,b]$ von beschränkter Variation, falls $[a,b]$ in endlich viele Teilintervalle zerlegt werden kann, auf denen φ monoton und beschränkt ist (s. Heuser (1986[1])).

Ist eine Funktion φ von beschränkter Variation auf $[a,b]$, so ist sie absolut Riemann-integrierbar auf $[a,b]$, d.h.

$$\int_a^b |\varphi(t)|dt < \infty.$$

1.1.2 Bemerkung und Definition: Es sei φ eine 2π-periodische Funktion und absolut Riemann-integrierbar auf $[-\pi,\pi]$. Dann existieren die Integrale

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \varphi(t)\cos(nt)dt,\ n \in \mathbb{N}\cup\{0\}, \tag{1.1}$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \varphi(t)\sin(nt)dt,\ n \in \mathbb{N}. \tag{1.2}$$

Die Werte a_n $(n \in \mathbb{N}\cup\{0\})$ und b_n $(n \in \mathbb{N})$ heißen *Fourier-Koeffizienten* von φ.

1.1.3 Satz und Definition: *Ist eine 2π-periodische Funktion φ*

(i) stetig und

(ii) von beschränkter Variation auf $[-\pi,\pi]$,

dann konvergiert die Folge $(\varphi_m(t))$ mit

$$\varphi_m(t) := \frac{a_0}{2} + \sum_{n=1}^{m}(a_n\cos(nt) + b_n\sin(nt)) \tag{1.3}$$

für $m \to \infty$ in jedem Punkt $t \in [-\pi,\pi]$ gegen $\varphi(t)$. Man sagt in diesem Zusammenhang, φ lasse sich in eine Fourier-Reihe

$$\varphi(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n\cos(nt) + b_n\sin(nt)) \tag{1.4}$$

entwickeln.

Beweis: Heuser (1986[2]).

1.1.4 Bemerkung: Fordert man für die Funktion φ nur stückweise Stetigkeit, dann konvergiert φ_m an den Stetigkeitsstellen punktweise gegen φ und an den Unstetigkeitsstellen t_0 gegen $\frac{1}{2}(\varphi(t_0-0)+\varphi(t_0+0))$. Dabei bezeichnet $\varphi(t_0-0)$ den links-, $\varphi(t_0+0)$ den rechtseitigen Grenzwert an der Stelle t_0.

Alternativ zur Darstellung (1.4) kann die Fourier-Reihe auf die Form

$$\varphi(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} A_n \cdot \sin(nt + \phi_n)$$

gebracht werden (spektrale Darstellung). Dabei gilt $A_n = \sqrt{a_n^2 + b_n^2}$ und $\tan \phi_n = \frac{a_n}{b_n}$. A_n heißt die *Amplitude* und ϕ_n die *Phasenverschiebung.*

Physikalisch betrachtet lässt sich die spektrale Darstellung einer deterministischen periodischen Funktion φ wie folgt beschreiben: Betrachtet man φ als Welle, die einen physikalischen, periodischen Vorgang in Abhängigkeit von der Zeit t beschreibt, wie z.B. den Zustand eines elektrischen Stromes, so besagt die spektrale Darstellung von φ, dass φ als Überlagerung (u.U. unendlich vieler) phasenverschobener Sinusfunktionen darstellbar ist. A_n gibt dabei die Amplitude der Sinusfunktion $\sin(nt + \phi_n)$ an, $\frac{1}{2}A_n^2$ die Leistung (=Energie pro Zeiteinheit), die diese Komponente zu φ beiträgt. Die Anzahl der Zyklen pro Zeiteinheit wird durch n bestimmt. Wird t in Sekunden gemessen, dann hat der Term $A_n \cdot \sin(nt + \phi_n)$ eine *Frequenz* von $\frac{n}{2\pi}$ Zyklen pro Sekunde (=$\frac{n}{2\pi}$ Hertz).

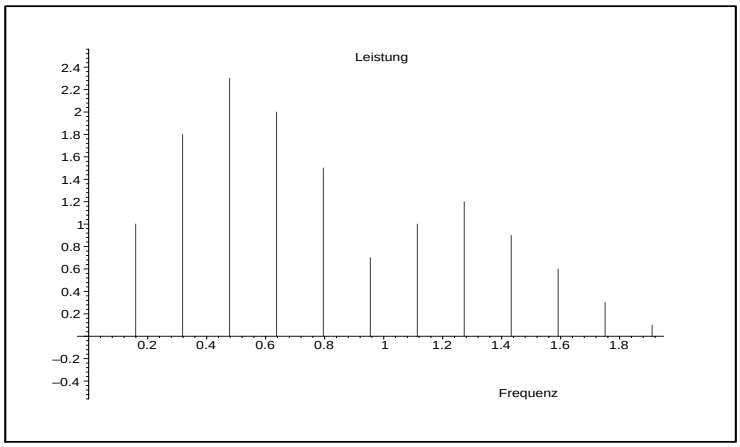

Abbildung 1.1: Diskretes Leistungsspektrum

Eine graphische Darstellung der Amplituden A_n bzw. der Phasenverschiebungen ϕ_n in Abhängigkeit von der Frequenz heißt *Amplituden-* bzw. *Phasenspektrum.* Eine graphische Darstellung der Punkte $(\frac{n}{2\pi}, \frac{1}{2}A_n^2),\ n \in \mathbb{N}$, heißt *diskretes Leistungsspektrum.* Abbildung 1.1 zeigt ein diskretes Leistungsspektrum.

1.1.5 Bemerkung: Die Existenz der Riemann-Integrale (1.1) und (1.2) wird durch die absolute Integrierbarkeit von φ sichergestellt. Für die Konvergenz der Summe in (1.3) genügt es i.A. jedoch **nicht**, die absolute Integrierbarkeit von φ zu fordern. Vielmehr muss φ auch von beschränkter Variation sein. Diese Aussage wird in Katznelson (1968) formal bewiesen, wird jedoch intuitiv klar, wenn man die physikalische Interpretation der Fourier-Theorie beachtet: Hochfrequente Schwankungen in φ zu führen zu positiven Werten a_n, b_n bzw. A_n^2 bei hohen Frequenzen, also für große n. Dies wiederum kann zur Folge haben, dass die Summe in (1.3) nicht konvergiert. Die Voraussetzung, dass φ von beschränkter Variation sein soll, gewährleistet hingegen die Existenz der Reihe in (1.4).

1.1.6 Bemerkung: Periodische Funktionen mit beliebiger Periode $2d$ lassen sich unter entsprechenden Voraussetzungen wie oben in eine Fourier-Reihe der Form

$$\varphi(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos\left(\frac{n\pi t}{d}\right) + b_n \sin\left(\frac{n\pi t}{d}\right)\right) \tag{1.5}$$

entwickeln. Dabei ist

$$a_n = \frac{1}{d}\int_{-d}^{d} \varphi(t) \cos\left(\frac{n\pi t}{d}\right) dt,\ n \in \mathbb{N} \cup \{0\}, \tag{1.6}$$

$$\text{und } b_n = \frac{1}{d}\int_{-d}^{d} \varphi(t) \sin\left(\frac{n\pi t}{d}\right) dt,\ n \in \mathbb{N}. \tag{1.7}$$

1.1.1 Komplexe Form der Fourier-Reihe

Ist i die imaginäre Einheit aus der Theorie der komplexen Zahlen ($i^2 = -1$), so gilt für $\theta \in \mathbb{R}$

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

Man setzt nun in (1.5)

$$\cos\left(\frac{n\pi t}{d}\right) = \frac{1}{2}\left(e^{\frac{n\pi it}{d}} + e^{-\frac{n\pi it}{d}}\right) \tag{1.8}$$

$$\sin\left(\frac{n\pi t}{d}\right) = -i\cdot\frac{1}{2}\left(e^{\frac{n\pi it}{d}} - e^{-\frac{n\pi it}{d}}\right) \tag{1.9}$$

und erhält die komplexe Form der Fourier-Reihe $2d$-periodischer Funktionen:

$$\begin{aligned}\varphi(t) &= \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(\frac{a_n}{2}\cdot\left[e^{\frac{n\pi it}{d}} + e^{-\frac{n\pi it}{d}}\right] - i\cdot\frac{b_n}{2}\cdot\left[e^{\frac{n\pi it}{d}} - e^{-\frac{n\pi it}{d}}\right]\right)\\ &= \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(\frac{a_n}{2} - i\cdot\frac{b_n}{2}\right)e^{\frac{n\pi it}{d}} + \sum_{n=1}^{\infty}\left(\frac{a_n}{2} + i\cdot\frac{b_n}{2}\right)e^{-\frac{n\pi it}{d}}\\ &= \sum_{n=-\infty}^{\infty} c_n\cdot e^{\frac{n\pi it}{d}}\end{aligned} \tag{1.10}$$

mit

$$c_n = \begin{cases}\frac{1}{2}(a_n - ib_n) & ,\ n \geq 1,\\ \frac{1}{2}a_0 & ,\ n = 0,\\ \frac{1}{2}(a_{-n} + ib_{-n}) & ,\ n \leq -1\end{cases} \overset{(1.6),(1.7)}{=} \frac{1}{2d}\int_{-d}^{d}\varphi(t)e^{-\frac{n\pi it}{d}}dt.$$

Dabei sind a_n und b_n wie in (1.6) und (1.7) definiert.

1.1.2 Fourier-Theorie im Raum $L^2(-\pi,\pi)$

Für eine größere Funktionenklasse als jene, welche die Voraussetzungen (i) und (ii) erfüllen, konvergiert φ_m in einem anderen Sinne, nämlich im Raum $L^2(-\pi,\pi)$ aller auf dem Intervall $[-\pi,\pi]$ *Lebesgue-integrierbaren* Funktionen (s.u.). Da die Definition des *Lebesgue-Integrals* und die hierfür ebenfalls notwendige Definition der *Borel-Messbarkeit* den hier zur Verfügung stehenden Rahmen sprengen würde, sei diesbezüglich auf die Standardliteratur zur Maßtheorie verwiesen. Eine gute Einführung in dieses Themengebiet bietet z.B. Hesse (2003). An dieser Stelle soll lediglich angemerkt werden, dass das Lebesgue-Integral nur für Borel-messbare Funktionen definiert ist, die Borel-Messbarkeit für eine breite Funktionenklasse, z.B. stetige Funktionen, monotone Funktionen, etc., erfüllt ist und eine reellwertige, Borel-messbare Funktion, die auf einem kompakten Intervall Riemann-integrierbar ist, dort auch Lebesgue-integrierbar ist. Es gibt jedoch Lebesgue-integrierbare Funktionen, die nicht Riemann-integrierbar sind. Existieren für eine Funktion φ beide Integrale, so stimmen die Integralwerte miteinander überein. Es ist daher üblich, für Lebesgue-Integrale dieselbe Schreibweise wie im Falle eines Riemann-Integrals zu verwenden. Der Leser sollte sich jedoch bewusst sein, dass in den unten definierten Räumen $L^2(a,b)$ und L^2 stets von Lebesgue-Integralen die Rede ist (die im Falle der Existenz der entsprechenden Riemann-Integrale mit diesen übereinstimmen). Die Borel-Messbarkeit der betreffenden Funktionen wird dabei stillschweigend vorausgesetzt.

In diesem Abschnitt werden einige grundlegende Begriffe aus der Funktionalanalysis, wie z.B. Innenprodukt, Norm, Vollständigkeit, etc., verwendet. Es sei bereits an dieser Stelle darauf hingewiesen, dass der Anhang A.1 einen kurzen Überblick über diese Grundlagen der Funktionalanalysis bietet.

1.1.7 Definition: Es sei φ eine Borel-messbare (reell- oder komplexwertige) Funktion mit existierendem Lebesgue-Integral

$$\int_a^b \varphi^2(t)dt < \infty.$$

Dann heißt φ *auf dem Intervall* $[a,b]$ *quadratisch (Lebesgue-)integrierbar.*

Die Menge aller quadratisch (Lebesgue-)integrierbaren Funktionen ist ein Vektorraum (über $\mathbb{R}$ bzw. $\mathbb{C}$), denn

$$\text{aus } \int_a^b \varphi^2(t)dt < \infty \text{ und } \int_a^b \psi^2(t)dt < \infty \text{ folgt } \int_a^b (\varphi\psi)(t)dt < \infty$$

und somit sind mit φ und ψ auch $\alpha\varphi$ ($\alpha \in \mathbb{R}$ bzw. $\alpha \in \mathbb{C}$) und $\varphi + \psi$ quadratisch Lebesgue-integrierbar.

Man beachte, dass wegen $\mathbb{R} \subset \mathbb{C}$ der Raum der komplexwertigen, auf $[a, b]$ quadratisch (Lebesgue-)integrierbaren Funktionen den Raum der reellwertigen, auf $[a, b]$ quadratisch (Lebesgue-)integrierbaren Funktionen enthält.

Auf dem Vektorraum aller komplexwertigen, auf $[a, b]$ quadratisch Lebesgue-integrierbaren Funktionen wird durch

$$\langle \varphi, \psi \rangle := \int_a^b \varphi(t)\overline{\psi(t)}dt \tag{1.11}$$

($\overline{z}$ ist die zu $z \in \mathbb{C}$ konjugiert komplexe Zahl) ein Innenprodukt eingeführt. Den so erhaltenen Innenproduktraum bezeichnet man mit $L^2(a, b)$.

Das Innenprodukt (1.11) induziert auf $L^2(a, b)$ eine Norm

$$\|\varphi\|^2 = \langle \varphi, \varphi \rangle = \int_a^b \varphi(t)\overline{\varphi(t)}dt. \tag{1.12}$$

Aus der Normeigenschaft $\|\varphi\| = 0 \Leftrightarrow \varphi = 0$ ergibt sich für $\varphi, \psi \in L^2(a, b)$

$$\varphi = \psi \iff \|\varphi(t) - \psi(t)\|^2 = 0. \tag{1.13}$$

Zwei Funktionen φ und ψ werden also auf $L^2[a, b]$ genau dann miteinander identifiziert, wenn ihr durch

$$\|\varphi(t) - \psi(t)\|^2 = \int_a^b |\varphi(t) - \psi(t)|^2 dt$$

definierter *Abstand* verschwindet.

1.1.8 Bemerkung: Unterscheiden sich zwei Funktionen auf dem Intervall $[a, b]$ an z.B. höchstens abzählbar vielen Punkten, also auf einer diskreten Menge, so sind sie in $L^2(a, b)$ identisch.

Im Folgenden ist die Gleichheit zweier quadratisch integrierbarer Funktionen, falls nicht anders erwähnt, immer im Sinne von (1.13) zu verstehen.

1.1.9 Definition: Es seien φ_n, $n \in \mathbb{N}$, und φ aus $L^2(a, b)$ und es gelte

$$\int_a^b |\varphi(t) - \varphi_n(t)|^2 dt \overset{n\to\infty}{\longrightarrow} 0.$$

Dann sagt man, die Folge (φ_n) *konvergiert in* $L^2(a, b)$ oder *im quadratischen Mittel* gegen φ.

1.1.10 Bemerkung: Der Innenproduktraum $L^2(a, b)$ ist bzgl. seiner Norm (1.12) *vollständig* (s. Heuser 1992). $L^2(a, b)$ ist also ein Hilbertraum.

Wir betrachten nun speziell den Hilbertraum $L^2(-\pi, \pi)$, der die Funktionen $1, \cos(t), \sin(t), \cos(2t), \sin(2t), \cos(3t), \ldots,$ enthält. Aufgrund der Eigenschaften der Sinus- und Cosinusfunktionen

$$\begin{aligned}
\int_{-\pi}^{\pi} \cos(nt)\cos(mt)dt &= 0, \ m \neq n, \\
\int_{-\pi}^{\pi} \cos(nt)\sin(mt)dt &= 0, \ m \in \mathbb{N}, n \in \mathbb{N}_0, \\
\int_{-\pi}^{\pi} \sin(nt)\sin(mt)dt &= 0, \ m \neq n,
\end{aligned} \tag{1.14}$$

bilden die Funktionen $\cos(nt)$, $n \in \mathbb{N}_0$, $\sin(nt)$, $n \in \mathbb{N}$, eine Folge orthogonaler Elemente aus $L^2(-\pi, \pi)$. Tatsächlich bilden diese Funktionen sogar eine Orthogonalbasis des $L^2(-\pi, \pi)$ (vgl. Heuser (1986^2)). Man kann also die Sinus- und Cosinusfunktionen als Grundbausteine auffassen, aus denen jedes $\varphi \in L^2(-\pi, \pi)$ additiv zusammengesetzt werden kann.

1.1.11 Satz: *Ist φ eine 2π-periodische Funktion aus $L^2(-\pi,\pi)$, dann konvergiert*

$$\varphi_m(t) = \frac{a_0}{2} + \sum_{n=1}^{m}(a_n\cos(nt) + b_n\sin(nt))$$

mit a_n und b_n wie in (1.1) bzw. (1.2) im quadratischen Mittel gegen φ, d.h.

$$\int_{-\pi}^{\pi}(\varphi(t)-\varphi_m(t))^2dt \overset{m\to\infty}{\longrightarrow} 0. \tag{1.15}$$

Beweis: HEUSER (1986[2]).

Wir betrachten nun wieder die komplexe Darstellung (1.10) der Fourier-Reihe. Wegen $e^{it} = \cos(t) + i\sin(t)$ und (1.14) ist

$$\begin{aligned}\int_{-\pi}^{\pi}\exp(int)\exp(imt)dt &= \int_{-\pi}^{\pi}\cos(nt)\cos(mt)dt - \int_{-\pi}^{\pi}\sin(nt)\sin(mt)dt\\ &= 0,\ (m\neq n).\end{aligned}$$

Mit $|e^{it}| = 1\ \forall t\in\mathbb{R}$ ergibt sich, dass die Funktionen $e^{int},\ n\in\mathbb{Z}$, eine Orthonormalfolge in $L^2(-\pi,\pi)$ bilden. Da die Sinus- und Kosinusfunktionen eine Basis des $L^2(-\pi,\pi)$ darstellen, folgt mit (1.8) und (1.9) aus dem oben Erläuterten, dass die Menge $\{e^{int},\ n\in\mathbb{Z}\}$ sogar eine Orthonormalbasis des Hilbertraumes $L^2(-\pi,\pi)$ ist. Satz A.1.25 aus Anhang A.1 stellt somit sicher, dass jede Summe der Form

$$\sum_{n=-\infty}^{\infty} c_n\cdot e^{int}$$

mit

$$\sum_{n=-\infty}^{\infty}|c_n|^2 < \infty$$

ein Element des Hilbertraumes $L^2(-\pi,\pi)$, also eine 2π-periodische, auf $[-\pi,\pi]$ quadratisch Lebesgue-integrierbare Funktion ist.

1.2 Fourier-Integrale nicht-periodischer Funktionen

Für die Entwicklung einer Funktion in eine Fourier-Reihe ist deren Periodizität unabdingbar. Eine nicht-periodische Funktion φ kann man jedoch als Fourier-**Integral** darstellen. Die Herleitung des *Fourier-Integrals* reellwertiger Funktionen soll hier kurz erläutert werden.

Es sei φ stetig und von beschränkter Variation auf $\mathbb{R}$. Außerdem existiere das Riemann-Integral

$$\int_{-\infty}^{\infty}|\varphi(t)|dt.$$

Weiter sei φ_d die $2d$-periodische Funktion, die auf dem Intervall $[-d,d]$ mit φ übereinstimmt und außerhalb dieses Intervalles periodisch fortgesetzt, also gewissermaßen "kopiert" wurde (φ_d muss an den Stellen $(2k+1)d,\ k\in\mathbb{Z}$, nicht notwendigerweise stetig sein). Dann lässt sich φ_d in eine Fourier-Reihe entwickeln, wobei für jede Stetigkeitsstelle $t\in\mathbb{R}$ gilt

$$\varphi_d(t) = \sum_{n=-\infty}^{\infty} c_n\cdot e^{2\pi i p_n t}.$$

Dabei ist

$$p_n = \frac{n}{2d} \text{ und } c_n = \frac{1}{2d}\int_{-d}^{d}\varphi(t)e^{-2\pi i p_n t}dt.$$

Mit $p_n - p_{n-1} = \frac{1}{2d}$ ergibt sich

$$\varphi_d(t) = \lim_{m\to\infty} \sum_{n=-m}^{m} \left(\int_{-d}^{d} \varphi(s) e^{-2\pi i p_n s} ds \right) e^{2\pi i p_n t} (p_n - p_{n-1}).$$

Für $d \to \infty$ strebt $p_n - p_{n-1}$ gegen 0 und man erhält

$$\varphi(t) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \varphi(s) e^{-2\pi i p s} ds \right) e^{2\pi i p t} dp. \tag{1.16}$$

Die Existenz des Riemann-Integrals

$$g(p) := \int_{-\infty}^{\infty} \varphi(t) e^{-2\pi i p t} dt$$

ist gewährleistet, da φ nach Voraussetzung absolut Riemann-integrierbar ist. Für die Existenz des Riemann-Integrals in (1.16) genügt es **nicht**, die absolute Integrierbarkeit von φ zu fordern. (Man vergleiche hierzu die Ausführungen in Abschnitt 1.1.) Da wir φ jedoch als von beschränkter Variation vorausgesetzt haben, existiert das Riemann-Integral in (1.16). Schließlich gilt die Darstellung (1.16) für $\varphi(t)$, da φ nach Voraussetzung stetig ist (vgl. Abschnitt 1.1).

1.2.1 Fourier-Transformierte

Geht man in (1.16) zur "Winkelfrequenz" $\omega = 2\pi \cdot p$ über, so erhält man die gewohnte Form

$$\varphi(t) = \int_{-\infty}^{\infty} f(\omega) e^{i\omega t} d\omega \tag{1.17}$$

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) e^{-i\omega t} dt. \tag{1.18}$$

Die Darstellung (1.17) heißt *Fourier-Integral-Darstellung* von φ, die Funktion f heißt *Fourier-Transformierte* von φ. In der Gleichung (1.17) wird φ auch die *inverse Fourier-Transformierte* von f genannt. Die Fourier-Transformierte stellt gewissermaßen ein stetiges Amplitudenspektrum dar.

1.2.1 Bemerkung: Die Fourier-Transformierte ist eine komplexwertige Funktion. Wegen

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\varphi(t)\cos(\omega t) - i\varphi(t)\sin(\omega t)) dt$$

verschwindet jedoch der imaginäre Anteil, falls φ *gerade* ist, falls also gilt

$$\varphi(t) = \varphi(-t) \ \forall t \in \mathbb{R}.$$

1.2.2 Bemerkung: Gelegentlich findet man sowohl die Fourier-Transformierte als auch die Fourier-Integral-Darstellung mit dem Vorfaktor $\sqrt{\frac{1}{2\pi}}$ versehen. Und manchmal steht der Faktor $\frac{1}{2\pi}$ in der Fourier-Integral-Darstellung statt in (1.18). Dies ist Definitionssache und sollte nicht irritieren.

1.2.3 Bemerkung: Wie im Falle periodischer Funktionen lässt sich auch im Falle nicht-periodischer Funktionen die Forderung der Stetigkeit abschwächen. Es genügt, stückweise Stetigkeit zu fordern. An einer Unstetigkeitsstelle t_0 konvergiert dann das Fourier-Integral gegen $\frac{1}{2}(\varphi(t_0 - 0) + \varphi(t_0 + 0))$.

1.2.2 Ein Sampling-Theorem

Gelegentlich ist die Frage von Interesse, inwieweit die Kenntnis einiger Funktionswerte von φ Rückschlüsse auf die übrigen Funktionswerte zulässt. In der Theorie der Fourier-Entwicklungen spielt diesbezüglich ein Sampling-Theorem eine große Rolle, welches besagt, dass man unter gewissen Voraussetzungen in der Lage ist, aus den Funktionswerten an äquidistanten Stellen $t_1, t_2, \ldots$ die gesamte Funktion zu rekonstruieren.

1.2.4 Satz: (Sampling-Theorem von Whittaker / Abtast-Theorem) *Es sei φ eine reellwertige, stetige Funktion und für die Fourier-Transformierte f von φ gelte:*

- *Es existiert ein $\Delta > 0$ so, dass f außerhalb des Intervalles $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ verschwindet,*
- *f ist auf $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ stetig und von beschränkter Variation,*

dann kann φ an jeder Stelle $t \in \mathbb{R}$ aus den Funktionswerten an den Stellen $\ldots, -2\cdot\Delta, -\Delta, 0, \Delta, 2\cdot\Delta, 3\cdot\Delta, \ldots$ rekonstruiert werden und es gilt

$$\varphi(t) = \sum_{n=-\infty}^{\infty} \varphi(n\cdot\Delta) \cdot \frac{\sin\left(\frac{\pi}{\Delta}(t - n\cdot\Delta)\right)}{\left(\frac{\pi}{\Delta}(t - n\cdot\Delta)\right)}. \tag{1.19}$$

Beweis: Da f nach Voraussetzung außerhalb $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ verschwindet und innerhalb des Intervalles von beschränkter Variation ist, ist f absolut integrierbar, und wegen der Stetigkeit von φ gilt nach (1.17)

$$\varphi(t) = \frac{1}{2\pi}\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} f(\omega)e^{i\omega t}d\omega,\ t \in \mathbb{R}. \tag{1.20}$$

Desweiteren folgt aus der Stetigkeit von f (wegen $f(\omega) = 0$ für $|\omega| > \frac{\pi}{\Delta}$)

$$f(-\frac{\pi}{\Delta}) = f(\frac{\pi}{\Delta}) = 0. \tag{1.21}$$

Setzt man f außerhalb $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ periodisch fort, so lässt sich die periodisch fortgesetzte (und wegen (1.21) stetige) Funktion f_d in eine Fourier-Reihe entwickeln. Mit (1.10) erhält man

$$\begin{aligned} f_d(\omega) &= \sum_{n=-\infty}^{\infty} c_n e^{in\omega\cdot\Delta} \\ &= \sum_{n=-\infty}^{\infty} d_n e^{-in\omega\cdot\Delta},\ \omega \in \mathbb{R}, \end{aligned}$$

mit

$$d_n := c_{-n} = \frac{\Delta}{2\pi}\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} f(\omega)e^{in\omega\cdot\Delta}d\omega. \tag{1.22}$$

Insbesondere gilt also für $\omega \in [-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$

$$f(\omega) = f_d(\omega) = \sum_{n=-\infty}^{\infty} d_n e^{-in\omega\cdot\Delta}.$$

Andererseits ist nach (1.20)

$$\varphi(n\cdot\Delta) = \frac{1}{2\pi}\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} f(\omega)e^{i\omega n\cdot\Delta}d\omega$$

und es ergibt sich mit (1.22)

$$d_n = \Delta\cdot\varphi(n\cdot\Delta). \tag{1.23}$$

Also gilt nach (1.20)

$$\begin{aligned} \varphi(t) &= \frac{1}{2\pi}\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} f(\omega)e^{i\omega t}d\omega \\ &= \frac{1}{2\pi}\int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \sum_{n=-\infty}^{\infty} d_n e^{-in\omega\cdot\Delta}\cdot e^{i\omega t}d\omega. \end{aligned} \tag{1.24}$$

Da f_d stetig und auf $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ von beschränkter Variation ist, konvergiert die Fourier-Reihe von f_d gleichmäßig (s. Heuser (1986[2])). Somit lässt sich in (1.24) die Reihenfolge von Summation und Integration vertauschen und man erhält mit (1.23)

$$\begin{aligned}
\varphi(t) &= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \Delta \cdot \varphi(n \cdot \Delta) \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} e^{-in\omega \cdot \Delta} \cdot e^{i\omega t} d\omega \\
&= \frac{\Delta}{2\pi} \sum_{n=-\infty}^{\infty} \varphi(n \cdot \Delta) \int_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} e^{i\omega(t-n\cdot\Delta)} d\omega \\
&= \frac{\Delta}{2\pi} \sum_{n=-\infty}^{\infty} \varphi(n \cdot \Delta) \left. \frac{e^{i\omega(t-n\cdot\Delta)}}{i(t-n\cdot\Delta)} \right|_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \\
&= \frac{\Delta}{2\pi} \sum_{n=-\infty}^{\infty} \varphi(n \cdot \Delta) \left. \frac{\cos(\omega(t-n\cdot\Delta)) + i\sin(\omega(t-n\cdot\Delta))}{i(t-n\cdot\Delta)} \right|_{-\frac{\pi}{\Delta}}^{\frac{\pi}{\Delta}} \\
&= \frac{\Delta}{\pi} \sum_{n=-\infty}^{\infty} \varphi(n \cdot \Delta) \frac{\sin(\frac{\pi}{\Delta}(t-n\cdot\Delta))}{(t-n\cdot\Delta)}.
\end{aligned}$$

□

Im Falle einer Funktion φ, deren Fourier-Transformierte **nicht** außerhalb eines Intervalles verschwindet, muss in (1.19) ein "Fehlerterm" addiert werden. Ein entsprechendes Sampling-Theorem findet man in Higgins (1996), S. 95.

1.2.3 Fourier-Integrale in $L^2(\mathbb{R})$

Wie bei den periodischen Funktionen lässt sich auch hier der Begriff des Fourier-Integrals auf solche nicht-periodische Funktionen φ erweitern, die auf ganz $\mathbb{R}$ quadratisch Lebesgue-integrierbar sind (in Zeichen, $\varphi \in L^2 := L^2(\mathbb{R})$).

Für $\varphi \in L^2$ wird die Fourier-Transformierte f wie folgt definiert: Es sei

$$\mathbf{1}_{[-n,n]}(t) := \begin{cases} 1, & t \in [-n,n], \\ 0, & \text{sonst}, \end{cases}$$

und $\varphi_n = \varphi \cdot \mathbf{1}_{[-n,n]},\ n \in \mathbb{N}$. Da φ_n außerhalb des Intervalls $[-n,n]$ verschwindet, folgt aus der quadratischen (Lebesgue-)Integrierbarkeit die L^1-*Integrierbarkeit* auf dem Intervall $[-n,n]$, also die Existenz des Lebesgue-Integrals

$$\int_{-\infty}^{\infty} |\varphi_n(t)| dt = \int_{-n}^{n} |\varphi(t)| dt.$$

Für die Funktionen $\varphi_n,\ n \in \mathbb{N}$, existieren also wie oben definiert die Fourier-Transformierten f_n mit

$$f_n(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_n(t) e^{-i\omega t} dt,\ n \in \mathbb{N}.$$

Die Folge (f_n) konvergiert nun im quadratischen Mittel gegen eine Funktion $f \in L^2(\mathbb{R})$, d.h.

$$\int_{-\infty}^{\infty} (f_n(\omega) - f(\omega))^2 d\omega \overset{n\to\infty}{\longrightarrow} 0. \tag{1.25}$$

Die Funktion f ist eindeutig in dem Sinne, dass sie unabhängig von der Wahl der Folge φ_n ist. Wählt man eine andere Funktionenfolge (ψ_n) mit $\psi_n \to \varphi$ in L^2 und Fourier-Transformierten g_n von ψ_n, so strebt g_n ebenfalls gegen f (s. Bachman et al. (2000)).

Man definiert nun die so erhaltene quadratisch integrierbare Funktion f als die Fourier-Transformierte von φ. Um den Unterschied der Konvergenz im quadratischen Mittel zur punktweisen Konvergenz hervorzuheben, schreibt man für (1.25) auch

$$f = \text{l.i.m.} f_n.$$

In L^2 definiert man also

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-i\omega t}dt \;:=\; \text{l.i.m.}_{n\to\infty}\frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_n(t)e^{-i\omega t}dt. \tag{1.26}$$

1.2.5 Bemerkung: Existiert das Fourier-Integral von φ im Riemann'schen Sinne, so gilt

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi(t)e^{-i\omega t}dt \text{ (Riemann-Integral)},$$

d.h. das gemäß (1.26) definierte Integral stimmt in diesem Falle mit dem Riemann-Integral überein.

1.2.6 Satz: (Theorem von Plancherel) *Es sei $\varphi \in L^2$ und f_n die Fourier-Transformierte von $\varphi_n = \varphi \cdot \mathbf{1}_{[-n,n]}$, $n \in \mathbb{N}$. Dann gilt*

$$\varphi(t) = \text{l.i.m.}_{n\to\infty}\int_{-\infty}^{\infty} f_n(\omega)e^{i\omega t}d\omega. \tag{1.27}$$

Beweis: Bachman et al. (2000).

Die Betrachtung der Fourier-Theorie im Raum L^2 aller quadratisch integrierbaren Funktionen bietet den Vorteil, dass für alle $\varphi \in L^2$ die Fourier-Transformierte existiert und die Abbildung $\mathcal{F} : L^2 \to L^2$, die einer Funktion $\varphi \in L^2$ ihre Fourier-Transformierte f zuordnet, eine bijektive Abbildung ist. Es existiert also eine inverse Abbildung $\mathcal{F}^{-1} : L^2 \to L^2$, die der Fourier-Transformierten f ihre inverse Fourier-Transformierte zuordnet. Aus der Bijektivität von $\mathcal{F}$ folgt unter anderem, dass für die Fourier-Transformierte f einer Funktion φ die inverse Fourier-Transformierte stets wieder φ selbst ist. Man beachte dabei, dass eine Funktion $\tilde{\varphi}$, welche sich von φ lediglich an der Stelle t_0 durch $\tilde{\varphi}(t_0) := \frac{1}{2}(\varphi(t_0+0) + \varphi(t_0-0))$ unterscheidet, in L^2 mit φ identifiziert wird.

Kapitel 2. Wavelets

Ähnlich wie bei Fourier-Transformationen, mit welchen eine periodische Funktion φ als Summe von Sinus- und Cosinusfunktionen dargestellt werden kann, bietet die Theorie der Wavelets die Möglichkeit, φ als Summe kleiner "Wellenpakete", den *Wavelets,* darzustellen. Häufig werden diese so gewählt, dass sie außerhalb eines relativ kleinen Intervalles verschwinden. Verglichen mit der Theorie der Fourier-Transformationen hat dies den Vorteil, dass Informationen über Frequenz-Komponenten in Abhängigkeit vom Zeit-Parameter zur Verfügung stehen. Das Prinzip soll anhand einfacher Wavelets, den *Haar-Wavelets*, beschrieben werden.

2.1 Die Haar-Wavelet-Transformierte

Die *Haar-Wavelet-Transformation* approximiert eine Funktion φ durch eine Summe von Treppenfunktionen ξ_j $(j = 0, 1, 2, \ldots, n)$. Die einzelnen Schritte der Haar-Wavelet Transformation werden nun anhand einer einfachen Treppenfunktion illustriert. Das Beispiel ist Nievergelt (1999) entnommen. Es sei

$$\varphi(t) = \begin{cases} 5, & t \in [0, \frac{1}{4}), \\ 1, & t \in [\frac{1}{4}, \frac{1}{2}), \\ 2, \quad \text{falls} & t \in [\frac{1}{2}, \frac{3}{4}), \\ 8, & t \in [\frac{3}{4}, 1), \\ 0, & \text{sonst}, \end{cases}$$

die Funktion, die in Abbildung 2.1 gezeigt wird.

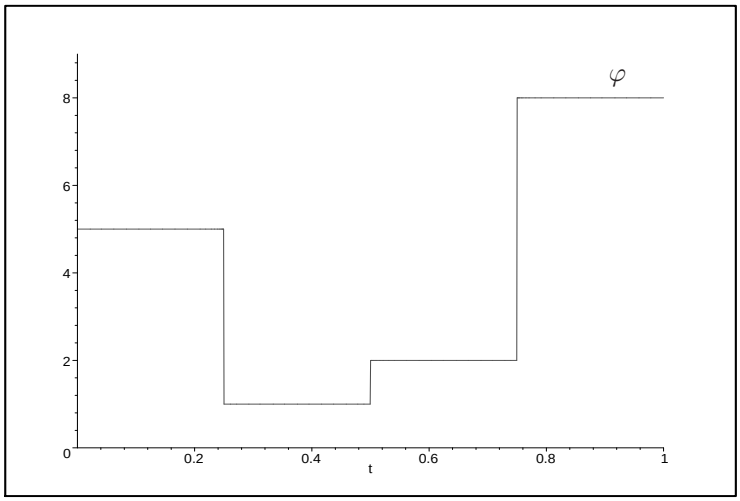

Abbildung 2.1: Beispiel: Die Treppenfunktion φ

Im ersten Schritt berechnet man das Mittel μ_0 aller Stufen und wählt als erstes Wavelet die Funktion, die auf dem Intervall $[0, 1)$ konstant diesen Wert annimmt, also

$$\xi_0(t) := \begin{cases} \mu_0 = 4, & t \in [0, 1), \\ 0, & \text{sonst}. \end{cases}$$

Im zweiten Schritt unterteilt man das Intervall $[0, 1)$ in seiner Mitte und berechnet für jedes Teilintervall gesondert die Mittelwerte $\mu_1^{(1)} = 3$ bzw. $\mu_1^{(2)} = 5$. Als zweites Wavelet wählt man jene Treppenfunktion, die auf dem jeweiligen Teilintervall konstant den Wert $\mu_1^{(j)} - \mu_0,\ j = 1, 2$, annimmt, d.h.

$$\xi_1(t) := \begin{cases} -1, & t \in [0, \frac{1}{2}), \\ 1, & t \in [\frac{1}{2}, 1), \\ 0, & \text{sonst}. \end{cases}$$

Im dritten Schritt unterteilt man die Intervalle $[0, \frac{1}{2})$ und $[\frac{1}{2}, 1)$ wiederum in ihrer Mitte und bildet die jeweiligen Mittelwerte $\mu_2^{(1)} = 5,\ \mu_2^{(2)} = 1,\ \mu_2^{(3)} = 2$ sowie $\mu_2^{(4)} = 8$. Wie oben erhält man zwei neue Wavelets, die auf jedem Teilintervall konstant die Werte

$$\begin{aligned} & \mu_2^{(i)} - \mu_1^{(1)} - \mu_0, \quad (i = 1,\ 2) \\ \text{bzw.} \quad & \mu_2^{(i)} - \mu_1^{(2)} - \mu_0, \quad (i = 3,\ 4) \end{aligned}$$

annehmen, also

$$\xi_2^{(1)}(t) = \begin{cases} 2, & t \in [0, \frac{1}{4}), \\ -2, & t \in [\frac{1}{4}, \frac{1}{2}), \\ 0, & \text{sonst}, \end{cases}$$

und

$$\xi_2^{(2)}(t) = \begin{cases} -3, & t \in [\frac{1}{2}, \frac{3}{4}), \\ 3, & t \in [\frac{3}{4}, 1), \\ 0, & \text{sonst}. \end{cases}$$

Man erhält so

$$\varphi = \xi_0 + \xi_1 + \xi_2^{(1)} + \xi_2^{(2)}.$$

Die Funktionen $\xi_0, \xi_1, \xi_2^{(1)}$ und $\xi_2^{(2)}$ sind in Abbildung 2.2 zu sehen.

Sofort wird einsichtig, dass $\xi_0, \xi_1, \xi_2^{(1)}$ und $\xi_2^{(2)}$ im Sinne des in (1.11) definierten Skalarproduktes paarweise orthogonal sind. Durch Normierung lässt sich nun leicht ein Orthonormalsystem bilden. Dazu sei

$$\phi(t) := \mathbf{1}\{0 \le t < 1\} = \begin{cases} 1, & t \in [0, 1), \\ 0, & \text{sonst}, \end{cases}$$

sowie

$$\begin{aligned} \psi(t) &:= \mathbf{1}\{0 \le t < \frac{1}{2}\} - \mathbf{1}\{\frac{1}{2} \le t < 1\}, \\ \psi_1^{(0)}(t) &:= \sqrt{2}\left(\mathbf{1}\{0 \le t < \frac{1}{4}\} - \mathbf{1}\{\frac{1}{4} \le t < \frac{1}{2}\}\right), \\ \psi_1^{(1)}(t) &:= \sqrt{2}\left(\mathbf{1}\{\frac{1}{2} \le t < \frac{3}{4}\} - \mathbf{1}\{\frac{3}{4} \le t < 1\}\right), \\ &\vdots \\ \psi_j^{(k)}(t) &:= 2^{\frac{j}{2}}\left(\mathbf{1}\{\frac{k}{2^j} \le t < \frac{k+\frac{1}{2}}{2^j}\} - \mathbf{1}\{\frac{k+\frac{1}{2}}{2^j} \le t < \frac{k+1}{2^j}\}\right), \\ & \quad j \in \mathbb{N},\ k = 0, \ldots, 2^j - 1. \end{aligned}$$

Eine leichte Rechnung ergibt

$$\psi_j^{(k)}(t) = 2^{\frac{j}{2}} \cdot \psi(2^j \cdot t - k),\ j \in \mathbb{N},\ k = 0, \ldots, 2^j - 1.$$

Die Funktionen $\psi_j^{(\cdot)},\ j \in \mathbb{N}$, sind gestauchte "Versionen" von ψ, die Funktionen $\psi_j^{(k)},\ k = 1, \ldots, 2^j - 1$, sind verschobene $\psi_j^{(0)}$.

Bevor nun Wavelets formal definiert werden ist anzumerken, dass die Definitionen in der Literatur nicht einheitlich sind. Diese Ausarbeitung übernimmt im Wesentlichen die Definitionen aus Vidakovic (1999) und Bachman et al. (2000).

2.1.1 Definition: Ein *Mother Wavelet* ist eine Funktion ψ auf $L^2(\mathbb{R})$ so, dass die Funktionen

$$\psi_j^{(k)}(t) := 2^{\frac{j}{2}} \cdot \psi(2^j \cdot t - k),\ j, k \in \mathbb{Z},$$

eine Orthonormalbasis des $L^2(\mathbb{R})$ bilden. Die Funktionen $\psi_j^{(k)}$ heißen *Wavelets*, j heißt *Level* und k ein *Shift* im Level j.

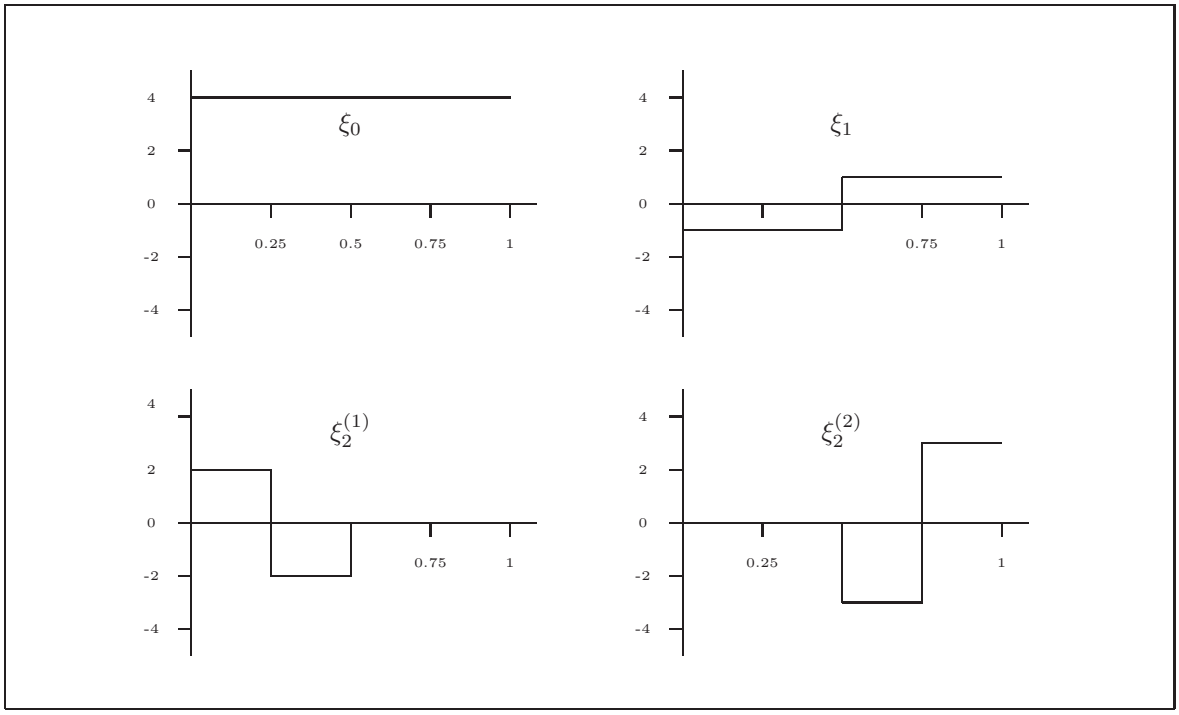

Abbildung 2.2: Die Funktionen $\xi_0, \xi_1, \xi_2^{(1)}$ und $\xi_2^{(2)}$

Betrachtet man zur Anschauung die Haar-Wavelets, so kann man sagen, dass alle Wavelets, die auf einem gleich langen Intervall von 0 verschieden sind, zum selben Level gehören. Die Länge dieses Intervalls nimmt mit zunehmendem Level ab. Entsprechend wird eine Funktion umso besser approximiert, je mehr Levels durchlaufen werden. Der Shift k bestimmt die Lage des Intervalls im Zeitparameter-Bereich, auf dem $\psi_j^{(k)}$ positive Werte annimmt. Entsprechend sind die Begriffe Shift und Level auch bei anderen Wavelets zu verstehen.

Anhand der Haar-Wavelets wird ein großer Vorteil der Wavelets gegenüber der Fourier-Transformation deutlich. Soll eine Funktion φ approximiert werden, deren Frequenz in der Zeit variiert (sagen wir, auf einem kurzen Intervall eine besonders hohe Frequenz aufweist), so wird jenes Wavelet mit dem (dieser Frequenz) entsprechenden Level und dem (dem Intervall) entsprechenden Shift stärker gewichtet als die anderen Komponenten desselben Levels. Auf diese Weise bietet die Theorie der Wavelets also die Möglichkeit, "Frequenz-Komponenten in Abhängigkeit von der Zeit" zu betrachten.

2.2 Die Konstruktion von Wavelets

2.2.1 Multiresolution Analysis

Das Ziel der Konstruktion von Wavelets ist die Konstruktion einer Orthonormalbasis (ONB) des $L^2 = L^2(\mathbb{R})$ der Form $\{\psi_{j,k} := 2^{j/2} \cdot \psi(2^j t - k),\ j,k \in \mathbb{Z}\}$ mit einer Funktion $\psi : \mathbb{R} \to \mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$). Dabei stellt man an die Funktion ψ die Bedingung

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{2.1}$$

Das Konstruktionsprinzip soll zunächst anhand der Haar-Wavelets veranschaulicht werden. Dazu sei $\phi(t) = \mathbf{1}\{0 \le t < 1\}$ wie in Abschnitt 2.1. Es ist leicht nachzuprüfen, dass ϕ die folgenden Eigenschaften besitzt:

(*i*) Die Menge $\{\phi(t-k), k \in \mathbb{Z}\}$ ist eine Orthogonalfolge in $L^2(\mathbb{R})$,

(*ii*) Für $t \in \mathbb{R}$ gilt $\phi(t) = \phi(2t) + \phi(2t-1)$ ("Scaling Identity").

Wir konstruieren nun einen linearen Untervektorraum (UVR) V_0 des L^2 so, dass die Menge $\{\phi(t-k),\ k\in\mathbb{Z}\}$ eine Orthogonal**basis** des V_0 bildet:

$$V_0 := \left\{ \varphi \in L^2 : \varphi(t) = \sum_{k=-\infty}^{\infty} a_k \phi(t-k),\ \text{wobei } (a_k) \subset \mathbb{R} \text{ mit } \sum_{k=-\infty}^{\infty} a_k^2 < \infty \right\}$$

ist der abgeschlossene lineare Unterraum stückweise konstanter Funktionen aus L^2 mit Sprüngen an den Stellen $k \in \mathbb{Z}$.

Es sei nun V_1 der abgeschlossene lineare UVR stückweise konstanter Funktionen aus L^2 mit Sprüngen an den Stellen $\frac{k}{2},\ k\in\mathbb{Z}$, d.h. V_1 sei so konstruiert, dass die Menge $\{\sqrt{2}\cdot\phi(2t-k),\ k\in\mathbb{Z}\}$ eine ONB des V_1 bildet. Für eine beliebige Funktion $\varphi \in L^2$ gilt

$$\varphi(t) \in V_1 \Longleftrightarrow \varphi\left(\frac{t}{2}\right) \in V_0.$$

Aus der "Scaling Identity"

$$\phi(t-k) = \phi(2t-2k) + \phi(2t-2k-1),\ k\in\mathbb{Z},\ \phi(t) = \mathbf{1}\{0 \le t < 1\},$$

folgt, dass jede der Funktionen $\phi(t-k)$, die die Basis für V_0 bilden, als Linearkombination von Funktionen $\phi(2t-k)$ darstellbar und somit im UVR V_1 enthalten ist. Somit gilt $V_0 \subseteq V_1$.

Allgemein bezeichne $V_j,\ j\in\mathbb{Z}$, den abgeschlossenen linearen UVR stückweise konstanter Funktionen aus L^2 mit Sprüngen an den Stellen $\frac{k}{2^j}$ $(k\in\mathbb{Z})$, d.h., V_j wird so konstruiert, dass die Menge

$$\left\{\phi_{j,k} := 2^{j/2}\cdot\phi(2^j t-k),\ k\in\mathbb{Z}\right\}$$

eine ONB des V_j bildet. Für $\varphi\in L^2$ ist

$$\varphi(t) \in V_j \Longleftrightarrow \varphi\left(\frac{t}{2^j}\right) \in V_0,\ j\in\mathbb{Z}.$$

Außerdem gilt

$$\cdots \subseteq V_{-1} \subseteq V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots \tag{2.2}$$

2.2.1 Definition: Eine aufsteigende Folge $\cdots \subseteq V_{-1} \subseteq V_0 \subseteq V_1 \subseteq V_2 \subseteq \cdots$ abgeschlossener Untervektorräume V_i des L^2 heißt *Multiresolution Analysis mit Scaling Function* ϕ, falls die V_i die folgenden Eigenschaften besitzen:

- (i) $\{\phi(t-k),\ k\in\mathbb{Z}\}$ ist eine Orthonormalbasis für V_0,
- (ii) $\varphi(t)\in V_j \Leftrightarrow \varphi\left(\frac{t}{2^j}\right)\in V_0$ $(j\in\mathbb{Z})$,
- (iii) $\bigcap_{j=-\infty}^{\infty} V_j = \{0\}$ (der Nullraum des L^2 entspricht der konstanten Nullfunktion),
- (iv) $\overline{\bigcup_{j=-\infty}^{\infty} V_j} = L^2$, wobei $\overline{\bigcup_{j=-\infty}^{\infty} V_j}$ den Abschluss der Menge $\bigcup_{j=-\infty}^{\infty} V_j$ bezeichnet. (Man sagt auch, $\bigcup_{j=-\infty}^{\infty} V_j$ liege *dicht* in L^2, d.h. jede Funktion $\varphi\in L^2$ lässt sich als L^2-Grenzwert einer konvergenten Folge von Funktionen aus $\bigcup_{j=-\infty}^{\infty} V_j$ darstellen.)

2.2.2 Bemerkung: Aus (i) und (ii) folgt, dass die Menge $\{2^{j/2}\cdot\phi(2^j t-k),\ k\in\mathbb{Z}\}$ eine Orthonormalbasis des V_j ist, $j\in\mathbb{Z}$ (s. Bachman et al. (2000)).

2.2.3 Bemerkung: Eine Multiresolution Analysis ist durch ihre Scaling Function eindeutig bestimmt. Eine umgekehrte Aussage gilt nicht, die Scaling Function ist nicht eindeutig durch die Multiresolution Analysis bestimmt (s. Bachman et al. (2000)).

2.2.4 Bemerkung: Die Folge (V_j) aus (2.2) bildet eine Multiresolution Analysis (vgl. Bachman et al. (2000)).

Nun zurück zum Spezialfall $\phi(t) = \mathbf{1}\{0 \leq t < 1\}$. Wäre die Vereinigung von Basen $\bigcup_{j=-\infty}^{\infty} \left\{2^{j/2} \cdot \phi(2^j t - k),\ k \in \mathbb{Z}\right\}$ selbst eine Orthonormalbasis des L^2, so hätten wir bereits unser Ziel erreicht. Doch die $\left\{2^{j/2} \cdot \phi(2^j t - k),\ k \in \mathbb{Z}\right\}$ sind zueinander nicht orthogonal ($j \in \mathbb{Z}$). Das nächste Ziel ist also, für V_{j+1} ($j \in \mathbb{Z}$) eine ONB zu bilden, die die ONB von V_j enthält. Wir beginnen mit $j = 0$.

2.2.5 Lemma: *Die Menge* $O_1 := \{\phi(t-k),\ k \in \mathbb{Z}\} \cup \{\psi(t-k),\ k \in \mathbb{Z}\}$ *mit*

$$\psi(t) := \phi(2t) - \phi(2t-1)$$

bildet eine ONB von V_1.

Beweis: Die Funktionen $\psi(t-k),\ k \in \mathbb{Z}$, sind in V_1 enthalten (denn $\psi(t-k) = \phi(2t-2k) - \phi(2t-2k-1)$) und sind orthogonal zu $\{\phi(t-k),\ k \in \mathbb{Z}\}$:

$$\begin{aligned}
\int \psi(t-k)\phi(t-k)dt &= \int \phi(2t-2k)\phi(t-k)dt - \int \phi(2t-2k-1)\phi(t-k)dt \\
&= \int \mathbf{1}_{[k,k+\frac{1}{2})}(t)\mathbf{1}_{[k,k+1)}(t)dt - \int \mathbf{1}_{[k+\frac{1}{2},k+1)}(t)\mathbf{1}_{[k,k+1)}(t)dt \\
&= \int \mathbf{1}_{[k,k+\frac{1}{2})}(t)dt - \int \mathbf{1}_{[k+\frac{1}{2},k+1)}(t)dt \\
&= \frac{1}{2} - \frac{1}{2} = 0.
\end{aligned}$$

Außerdem gilt

$$\int \psi(t-k)\psi(t-j) = 0 \ (k \neq j),$$

die Menge O_1 ist also ein Orthogonalsystem in V_1. Ist $\varphi \in V_1$, so existiert eine Darstellung

$$\begin{aligned}
\varphi(t) &= \sum_{k=-\infty}^{\infty} a_k \phi(2t-k) \\
&= \sum_{k=-\infty}^{\infty} \frac{a_{2k}}{2}\Big(\phi(2t-2k) + \phi(2t-2k-1) + \phi(2t-2k) - \phi(2t-2k-1)\Big) \\
&\quad + \sum_{k=-\infty}^{\infty} \frac{a_{2k+1}}{2}\Big(\phi(2t-2k) + \phi(2t-2k-1) - \phi(2t-2k) + \phi(2t-2k-1)\Big) \\
&= \sum_{k=-\infty}^{\infty} \left(\frac{a_{2k}}{2}\Big(\phi(t-k) + \psi(t-k)\Big) + \frac{a_{2k+1}}{2}\Big(\phi(t-k) - \psi(t-k)\Big)\right),
\end{aligned}$$

O_1 ist also ein *Erzeugendensystem* von V_1, d.h., jede Funktion $\varphi \in V_1$ lässt sich als gewichtete Summe von Elementen aus O_1 schreiben. Aus der Orthogonalität der Elemente von O_1 folgt deren lineare Unabhängigkeit, das Erzeugendensystem ist also minimal und somit eine Basis von V_1. Die Behauptung folgt nun aus

$$\|\psi(t-k)\|^2 = \int \phi^2(2t-2k)dt + \int \phi^2(2t-2k-1)dt = \|\phi(t-k)\|^2 = 1. \qquad \square$$

Wir bezeichnen den von den Funktionen $\psi(t-k),\ k \in \mathbb{Z}$, erzeugten linearen Unterraum mit W_0. Dann folgt aus Lemma 2.2.5, dass jede Funktion $\varphi \in V_1$ darstellbar ist in der Form

$$\varphi = \varphi_V + \varphi_W \text{ mit } \varphi_V \in V_0 \text{ und } \varphi_W \in W_0.$$

Man schreibt

$$V_1 = V_0 \oplus W_0$$

und sagt, V_1 ist die *direkte Summe* von V_0 und W_0. Da V_0 und W_0 zusätzlich zueinander orthogonal sind, nennt man W_0 das *orthogonale Komplement* von V_0 in V_1, d.h. $W_0 = V_0^{\perp} \cap V_1$.

Nun bildet man eine ONB für V_2, die die ONB von V_1 enthält. Hierzu sei W_1 das orthogonale Komplement von V_1 in V_2, also

$$W_1 \perp V_1 \text{ und } V_2 = V_1 \oplus W_1 = V_0 \oplus W_0 \oplus W_1.$$

Eine geeignete ONB für W_1 ist $\{\sqrt{2}\cdot\psi(2t-k),\ k\in\mathbb{Z}\}$, so dass durch die Menge $\{\phi(t-k),\ k\in\mathbb{Z}\}\cup\{\psi(t-k),\ k\in\mathbb{Z}\}\cup\{\sqrt{2}\cdot\psi(2t-k),\ k\in\mathbb{Z}\}$ eine ONB für V_2 gegeben ist.

Eine Weiterführung dieses Verfahrens ergibt für $j\in\mathbb{Z}$ Basen $\{2^{j/2}\cdot\psi(2^jt-k),\ k\in\mathbb{Z}\}$ für das orthogonale Komplement W_j von V_j in V_{j+1} und es gilt

$$W_j \perp V_j \text{ sowie } V_{j+1} = V_0 \oplus W_0 \oplus \cdots \oplus W_j. \tag{2.3}$$

Man beachte, dass

$$\begin{aligned} & V_0 \perp W_j,\ (j\in\mathbb{N}) \\ \text{und} \quad & W_i \perp W_j,\ (i\neq j). \end{aligned}$$

Außerdem sind die W_j als orthogonale Komplemente abgeschlossen (s. Bachman et al. (2000)).

2.2.6 Definition: Für eine Folge $(M_n)_{n\in\mathbb{N}}$ paarweise orthogonaler, abgeschlossener Unterräume eines Hilbertraumes ist

$$\bigoplus_{n=0}^{\infty} M_n := \overline{\left\langle \bigcup_{n=0}^{\infty} M_n \right\rangle},$$

wobei $\langle\ \cdot\ \rangle$ die lineare Hülle und $\overline{(\ \cdot\ )}$ den Abschluss bezeichnet.

Man betrachte nun die Menge $V_0 \oplus \bigoplus_{j=0}^{\infty} W_j$. Angenommen, es existiert ein $\varphi\in L^2$, das orthogonal zu $V_0\oplus\bigoplus_{j=0}^{\infty}W_j$ ist, also insbesondere $\varphi\perp V_0$ und $\varphi\perp W_j$ $(j\geq 0)$. Dann ist φ nach (2.3) auch orthogonal zu allen V_j $(j\geq 0)$, also $\varphi\perp\overline{\bigcup_{j=0}^{\infty}V_j} = L^2$. Die einzige Funktion, die diese Eigenschaft besitzt, ist das Nullelement in L^2. Also gilt

$$\left(V_0 \oplus \bigoplus_{j=0}^{\infty} W_j\right)^{\perp} = \{0\}. \tag{2.4}$$

2.2.7 Lemma: *Ist M eine Teilmenge des Hilbertraumes H, so gilt*

$$M^{\perp} = \{0\} \Longleftrightarrow M \text{ dicht in } H.$$

Beweis: Bachman et al. (2000).

Nach (2.4) und Lemma 2.2.7 ist $V_0\oplus\bigoplus_{j=0}^{\infty}W_j$ dicht in L^2. Da $V_0\oplus\bigoplus_{j=0}^{\infty}W_j$ nach Definition 2.2.6 abgeschlossen ist, folgt

$$V_0 \oplus \bigoplus_{j=0}^{\infty} W_j = L^2(\mathbb{R}). \tag{2.5}$$

Somit ist $\{\phi_{0,k}=\phi(t-k),\ k\in\mathbb{Z}\}\cup\{\psi_{j,k},\ j,k\in\mathbb{Z},\ j\geq 0\}$ eine ONB für L^2.

Für $j\in\mathbb{N}$ bezeichne nun W_{-j} das orthogonale Komplement von V_{-j} in V_{-j+1}. Es gilt

$$V_0 = V_{-1}\oplus W_{-1} = V_{-2}\oplus W_{-2}\oplus W_{-1} = V_{-n}\oplus W_{-n}\oplus\cdots\oplus W_{-1}\ (n\in\mathbb{N}).$$

2.2.8 Lemma: *Es gilt*

$$V_0 = \bigoplus_{j=1}^{\infty} W_{-j}\ .$$

Beweis: Es sei $\varphi \in V_0$ orthogonal zu $\bigoplus_{j=1}^{\infty} W_{-j}$. Insbesondere gilt dann $\varphi \perp W_{-j}\ \forall j \in \mathbb{N}$. Damit ist aber auch

$$\varphi \in \bigcap_{j=1}^{\infty} V_{-j} = \bigcap_{j=-\infty}^{\infty} V_j \overset{(iii)}{=} \{0\}.$$

Nach Lemma 2.2.7 ist $\bigoplus_{j=1}^{\infty} W_{-j}$ dicht in V_0. Da $\bigoplus_{j=1}^{\infty} W_{-j}$ abgeschlossen ist, folgt die Behauptung. □

Somit gilt

$$\bigoplus_{j=-\infty}^{\infty} W_j = L^2(\mathbb{R}).$$

Mit dieser Konstruktion erhält man also eine andere ONB $\{\psi_{j,k} := 2^{j/2} \cdot \psi(2^j t - k),\ j,k \in \mathbb{Z}\}$ aus *Wavelets* $\psi_{j,k}$ für den linearen Vektorraum L^2.

Das bis jetzt für den Spezialfall der Haar-Wavelets beschriebene Prinzip funktioniert für jede beliebige Multiresolution Analysis. Im Folgenden beschränken wir uns jedoch auf reellwertige Scaling Functions und Mother Wavelets.

2.2.2 Multiresolution Analysis ergibt Wavelet-Basis

Es sei $(V_j)_{j=-\infty}^{\infty}$ eine Multiresolution Analysis mit Scaling Function ϕ (aus technischen Gründen wird meist $\int \phi(t)dt \neq 0$ gefordert). Wegen $V_0 \subset V_1$ existiert eine Darstellung

$$\phi(t) = \sum_{k=-\infty}^{\infty} h_k \cdot \sqrt{2}\phi(2t-k) \text{ (allgemeine Form der Scaling Identity)}, \tag{2.6}$$

mit den Koeffizienten (vgl. Satz A.1.25)

$$h_k = \Big\langle \phi(t), \sqrt{2}\phi(2t-k) \Big\rangle = \sqrt{2} \int_{-\infty}^{\infty} \phi(t)\phi(2t-k)dt,\ k \in \mathbb{Z}.$$

2.2.9 Bemerkung: Die Folge (h_k) besitzt die Eigenschaften (vgl. Vidakovic (1999))

1.) $\sum_{k=-\infty}^{\infty} h_k = \sqrt{2}$ (Normiertheit),

2.) $\sum_{k=-\infty}^{\infty} h_{k-2h}h_{k-2l} = \begin{cases} 1, & h=l, \\ 0, & \text{sonst}, \end{cases}$ für $h,l \in \mathbb{Z}$ (Orthogonalität).

Auch hier ist anzumerken, dass die Bezeichnungen in der Literatur nicht eindeutig sind. Vidakovic (1999) sowie Percival / Walden (2000) bezeichnen die Folge $(h_k)_{k=-\infty}^{\infty}$ als *Wavelet Filter* und die Funktion

$$m_0(\omega) := \frac{1}{\sqrt{2}} \sum_{j=-\infty}^{\infty} h_j e^{-i\omega j}$$

als *Transfer-Funktion*. Bachman et al. (2000) nennen die Funktion m_0 *Wavelet Filter*.

Neben der allgemeinen Form (2.6) der Scaling Identity existiert eine Funktion $\psi \in W_0$, so dass

$$\{\psi_{j,k}(t) := 2^{j/2}\psi(2^j t - k),\ j,k \in \mathbb{Z}\}$$

eine Orthonormalbasis des $L^2(\mathbb{R})$ ist. Die Funktion ψ wird *Mother Wavelet* genannt, die Funktionen $\psi_{j,k}$ heißen *Wavelets*. Wegen $W_0 \subset V_1$ lässt sich ψ schreiben als

$$\psi(t) = \sum_{k=-\infty}^{\infty} g_k \cdot \sqrt{2}\phi(2t-k).$$

Für die Wahl der Koeffizienten g_k gibt es mehrere geeignete Möglichkeiten (entsprechend erhält man unterschiedliche Mother Wavelets). Der folgende Satz besagt, dass $g_k := (-1)^k h_{1-k}$ eine mögliche Wahl ist.

2.2.10 Satz: *Es sei $(V_j)_{j=-\infty}^{\infty}$ eine Multiresolution Analysis mit Scaling Function ϕ und die Funktion ψ sei gegeben durch*

$$\psi(t) = \sum_{k=-\infty}^{\infty} (-1)^k h_{1-k} \sqrt{2} \phi(2t-k)$$

mit (h_k) wie in (2.6). Dann ist

$$\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k),\ j,k \in \mathbb{Z}\}$$

eine Orthonormalbasis des $L^2(\mathbb{R})$.

Beweis: Vidakovic (1999).

2.2.11 Bemerkung: In der Regel stellt man an die Folge (g_k) die Bedingung

$$\sum_{k=-\infty}^{\infty} g_k = 0. \tag{2.7}$$

2.2.3 Konstruktion einer Multiresolution Analysis

Nach Satz 2.2.10 und Bemerkung 2.2.3 gilt es also, eine Scaling Function ϕ zu finden, die eine Multiresolution Analysis erzeugt. Ein besonderes Interesse gilt dabei der Konstruktion von Scaling Functions mit kompaktem Träger. Zu diesem Zweck werden zunächst die wichtigsten Eigenschaften einer solchen Scaling Function betrachtet:

2.2.12 Satz: *Es sei ϕ eine Scaling Function mit kompaktem Träger, $(h_k), k \in \mathbb{Z}$, die Koeffizientenfolge aus (2.6) und Φ die Fourier-Transformierte von ϕ mit $\Phi(0) \neq 0$. Dann ist Φ stetig und*

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{j=-\infty}^{\infty} h_j e^{-i\omega j}$$

ist ein trigonometrisches Polynom mit den Eigenschaften

(a) m_0 *ist stetig und 2π-periodisch,*

(b) $|m_0(\omega)|^2 + |m_0(\omega+\pi)|^2 = 1$ *für alle* $\omega \in \mathbb{R}$,

(c) $m_0(0) = 1$.

Weiter existiert ein $n \in \mathbb{N}$, so dass

$$m_0(\omega) = \frac{1}{\sqrt{2}} \sum_{j=-n}^{n} h_j e^{-i\omega j}.$$

Beweis: Bachman et al. (2000).

Hat man umgekehrt ein trigonometrisches Polynom gefunden, das die Bedingungen (a) - (c) erfüllt, wie findet man dann eine Funktion ϕ, die die Scaling Identity erfüllt? Eine Idee liefern die folgenden Lemmata:

2.2.13 Lemma: *Die Scaling Identity*

$$\phi(t) = \sum_{k=-\infty}^{\infty} h_k \cdot \sqrt{2}\phi(2t-k) \tag{2.8}$$

ist äquivalent zu

$$\Phi(\omega) = m_0\left(\frac{\omega}{2}\right) \cdot \Phi\left(\frac{\omega}{2}\right). \tag{2.9}$$

Beweis: Man bilde die Fourier-Transformierte von ϕ :

$$\begin{aligned}\Phi(\omega) &= \sqrt{2}\sum_{k=-\infty}^{\infty} h_k \cdot \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(2t-k)e^{-i\omega t}dt \\ &= \frac{1}{\sqrt{2}}\sum_{k=-\infty}^{\infty} h_k e^{-i\frac{\omega}{2}k}\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(2t-k)e^{-i\frac{\omega}{2}(2t-k)}d(2t-k) \\ &= m_0\left(\frac{\omega}{2}\right)\cdot\Phi\left(\frac{\omega}{2}\right).\end{aligned}$$

□

Eine wiederholte Anwendung von (2.9) ergibt die alternative Darstellung der Scaling Identity

$$\Phi(\omega) = \left(\prod_{j=1}^{n} m_0\left(\frac{\omega}{2^j}\right)\right)\Phi\left(\frac{\omega}{2^n}\right),\ n\in\mathbb{N}. \tag{2.10}$$

2.2.14 Lemma: *Sind ϕ, Φ und m_0 wie in Satz 2.2.12, so konvergiert das Produkt in (2.10) für $n\to\infty$ gleichmäßig auf jeder beschränkten Teilmenge von $\mathbb{R}$. Gilt außerdem $\Phi(0)=1$, so ist*

$$\Phi(\omega) = \prod_{j=1}^{\infty} m_0\left(\frac{\omega}{2^j}\right). \tag{2.11}$$

Beweis: Bachman et al. (2000).

Gegeben sei nun ein trigonometrisches Polynom m_0 mit (a) - (c). Man setze, motiviert durch (2.11),

$$\Phi(\omega) := \prod_{j=1}^{\infty} m_0\left(\frac{\omega}{2^j}\right). \tag{2.12}$$

Unter den Voraussetzungen (a) - (c) konvergiert das Produkt (2.12), und die so definierte Funktion Φ ist aus $L^2(\mathbb{R})$ und stetig (s. Bachman et al. (2000)). Die inverse Fourier-Transformierte ϕ von Φ ist ebenfalls ein Element aus $L^2(\mathbb{R})$ und erfüllt die Scaling Identity

$$\phi(t) = \sum_{k=-\infty}^{\infty} h_k\sqrt{2}\phi(2t-k) \tag{2.13}$$

(vgl. Bachman et al. (2000)). Hat man nun auf die beschriebene Weise eine Funktion $\phi\in L^2$ gefunden, die die Scaling Indentity (2.13) erfüllt, so ist das System $\{\phi(t-k),\ k\in\mathbb{Z}\}$ jedoch nicht notwendigerweise orthonormal, mit anderen Worten, die Bedingungen (a) - (c) an die Funktion m_0 sind zwar notwendig, aber nicht hinreichend für die Orthonormalität von $\{\phi(t-k),\ k\in\mathbb{Z}\}$. Die folgenden Sätze 2.2.15 und 2.2.17 besagen, dass die zusätzliche Bedingung

(d) $m_0(\omega)\neq 0$ für $\omega\in[-\frac{\pi}{2},\frac{\pi}{2}]$

die Orthonormalität von $\{\phi(t-k),\ k\in\mathbb{Z}\}$ gewährleistet, dass wir also mit der Funktion ϕ eine Scaling Funktion gefunden haben, die eine Multiresolution Analysis erzeugt.

2.2.15 Satz: *Eine Familie $\{\varphi(\omega-k),\ k\in\mathbb{Z}\}\subset L^2(\mathbb{R})$ ist genau dann orthonormal, wenn für die Fourier-Transformierte f von φ gilt*

$$\sum_{j=-\infty}^{\infty}|f(\omega+2\pi j)|^2 = 1 \text{ für fast alle } \omega\in\mathbb{R} \text{ (vgl. Bem 2.2.16).}$$

Beweis: Bachman et al. (2000).

2.2.16 Bemerkung: In Satz 2.2.15 ist

$$''\sum_{j=-\infty}^{\infty}|f(\omega+2\pi j)|^2 = 1 \text{ für fast alle } \omega\in\mathbb{R}\ ''$$

gleichbedeutend damit, dass die durch $g(\omega):=\sum_{j=-\infty}^{\infty}|f(\omega+2\pi j)|^2,\ \omega\in\mathbb{R}$, definierte Funktion auf $L^2(\mathbb{R})$ mit der konstanten Funktion $h(\omega)\equiv 1$ übereinstimmt, dass also gilt $\int_{\mathbb{R}}|g(\omega)-1|^2d\omega = 0$ (Lebesgue-Integral).

2.2.17 Satz: *Es sei m_0 ein trigonometrisches Polynom, das die Bedingungen (a) - (d) erfüllt. Dann gilt für die gemäß (2.12) definierte Funktion Φ*

$$\sum_{j=-\infty}^{\infty} |\Phi(\omega + 2\pi j)|^2 = 1 \text{ für fast alle } \omega \in \mathbb{R}.$$

Beweis: Bachman et al. (2000).

2.2.18 Bemerkung: Es sei $\{\phi(\omega - k),\ k \in \mathbb{Z}\}$ eine Basis von V_0 und Φ die Fourier-Transformierte von ϕ. Weiter gelte

$$A \leq \sum_{j=-\infty}^{\infty} |\Phi(\omega + 2\pi j)|^2 \leq B \tag{2.14}$$

für Konstanten $A, B \in \mathbb{R}$. Man setze

$$\tilde{\Phi}(\omega) := \frac{\Phi(\omega)}{\sqrt{\sum_{j=-\infty}^{\infty} |\Phi(\omega - 2\pi j)|^2}}$$

und es bezeichne $\tilde{\phi}$ die inverse Fourier-Transformierte von $\tilde{\Phi}$. Dann ist

$$\{\tilde{\phi}(\omega - k),\ k \in \mathbb{Z}\}$$

eine ONB für V_0.

Ein Erzeugendensystem $\{\phi(\omega - k),\ k \in \mathbb{Z}\}$ von V_0 mit der Eigenschaft (2.14) heißt *Frame*. In der Praxis verzichtet man manchmal auf die Orthogonalitätseigenschaft und/oder auf die Minimalität des Erzeugendensystems $\{\phi(\omega - k),\ k \in \mathbb{Z}\}$. Man sucht eine Funktion ϕ, die die Scaling Identity (2.10), die Eigenschaft (d) sowie (2.14) erfüllt und erhält so anstelle einer ONB einen Frame (vgl. Daubechies (1988)).

2.2.4 Daubechies' Ansatz

Daubechies' Zielsetzung war die Konstruktion eines Mother Wavelets mit **kompaktem** Träger, wobei die Trägerweite möglichst gering sein sollte. Dies bedeutet notwendigerweise eine endliche, möglichst kleine Anzahl von Koeffizienten h_n mit $h_n \neq 0$. Daubechies hat gezeigt (vgl. Daubechies (1988)), dass, abgesehen von der Haar-Wavelet-Basis, **keine** symmetrische oder antisymmetrische Wavelet-Basis mit kompaktem Träger existiert. Entsprechend besitzt das von Daubechies entwickelte Wavelet eine eher irreguläre Form.

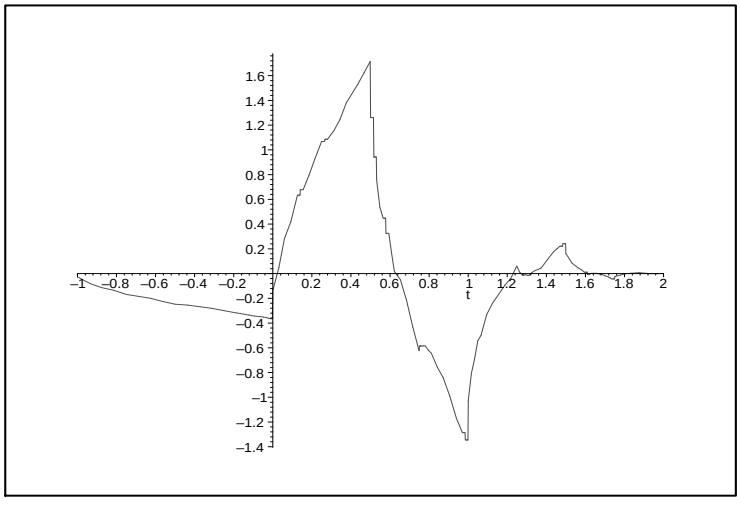

Abbildung 2.3: Daubechies' Mother Wavelet ψ

Bei der Entwicklung ihrer Wavelet-Basis ließ sich Daubechies u. A. von einem Prinzip inspirieren, das ursprünglich zur Bilddaten-Kompression verwendet wurde (Laplace'sches Pyramiden-Schema, P. Burt, E. Adelson). Im Laplace'schen Pyramiden-Schema ist $(H_k) =: (H_k^{(0)})$ eine Folge von Daten (Bildpixel-Werte), die in mehrere Folgen $(H_k^{(1)}), (H_k^{(2)}), \ldots$ aufgespalten wird, welche jeweils zu unterschiedlichen Auflösungen korrespondieren. Die Folge $(H_k^{(m)})$ wird dabei bestimmt durch

$$H_k^{(m)} := \sum_{j=-\infty}^{\infty} W(j-2k) H_j^{(m-1)}, \tag{2.15}$$

mit geeigneten, reellen Koeffizienten $W(j)$. Daubechies' Idee war nun die Konstruktion eines Funktionals $\mathcal{F}$ mit

$$(\mathcal{F}g)(t) := \sum_{k=0}^{N} h_k \sqrt{2} g(2t-k),\ g : \mathbb{R} \to \mathbb{R},$$

welches, mit $g = \mathbf{1}_{[0,1)}$, nach iterierter Anwendung eine Scaling Funktion ϕ ergibt. D.h.

$$\phi(t) := \lim_{n\to\infty} (\mathcal{F}^n g)(t),$$

wobei

$$\begin{aligned} (\mathcal{F}^2 g)(t) &= \sum_{k=0}^{N} h_k \sqrt{2} (\mathcal{F}g)(2t-k), \\ \text{und } (\mathcal{F}^n g)(t) &= \sum_{k=0}^{N} h_k \sqrt{2} (\mathcal{F}^{n-1} g)(2t-k),\ n \in \mathbb{N}. \end{aligned}$$

Um sicher zu stellen, dass man auf diese Weise eine Scaling Function erhält, die eine Multiresolution Analysis erzeugt, stellt Daubechies an die Koeffizienten h_k unter Anderem die folgenden Anforderungen:

1.) $\sum_{k=-\infty}^{\infty} h_k = \sqrt{2}$,

2.) $\sum_{k=-\infty}^{\infty} h_{k-2h} h_{k-2l} = \begin{cases} 1, & h = l, \\ 0, & \text{sonst}, \end{cases}$ für $h, l \in \mathbb{Z}$,

3.) Es existiert ein $\alpha > 0$, so dass $\sum_{k=-\infty}^{\infty} |h_k||k|^\alpha < \infty$,

sowie weitere Bedingungen, die die Konvergenz des Produktes (2.12), die Stetigkeit der Scaling Function und (2.7) gewährleisten. Diese sämtlichen Bedingungen werden z. B. von den folgenden Koeffizienten erfüllt:

$$\begin{aligned} h_0 &= \frac{1+\sqrt{3}}{4\sqrt{2}}, \\ h_1 &= \frac{3+\sqrt{3}}{4\sqrt{2}}, \\ h_2 &= \frac{3-\sqrt{3}}{4\sqrt{2}}, \\ h_3 &= \frac{1-\sqrt{3}}{4\sqrt{2}}, \\ h_k &= 0,\ k \notin \{0,1,2,3\}. \end{aligned} \tag{2.16}$$

Die Abbildung 2.4 veranschaulicht die iterierte Anwendung des Funktionals $\mathcal{F}$ auf die Funktion $g = \mathbf{1}_{[0,1)}$ mit den Koeffizienten aus (2.16).

Als Scaling Funktion (also für $n \to \infty$) erhält man *Daubechies' Building Block* (s. Abbildung 2.5).

Schließlich erhält man mit $g_k = (-1)^k h_{1-k}$ das Daubechies' Mother Wavelet, also

$$\psi(t) = h_3\phi(2t+2) - h_2\phi(2t+1) + h_1\phi(2t) - h_0\phi(2t-1),$$

welches in Abbildung 2.3 zu sehen ist.

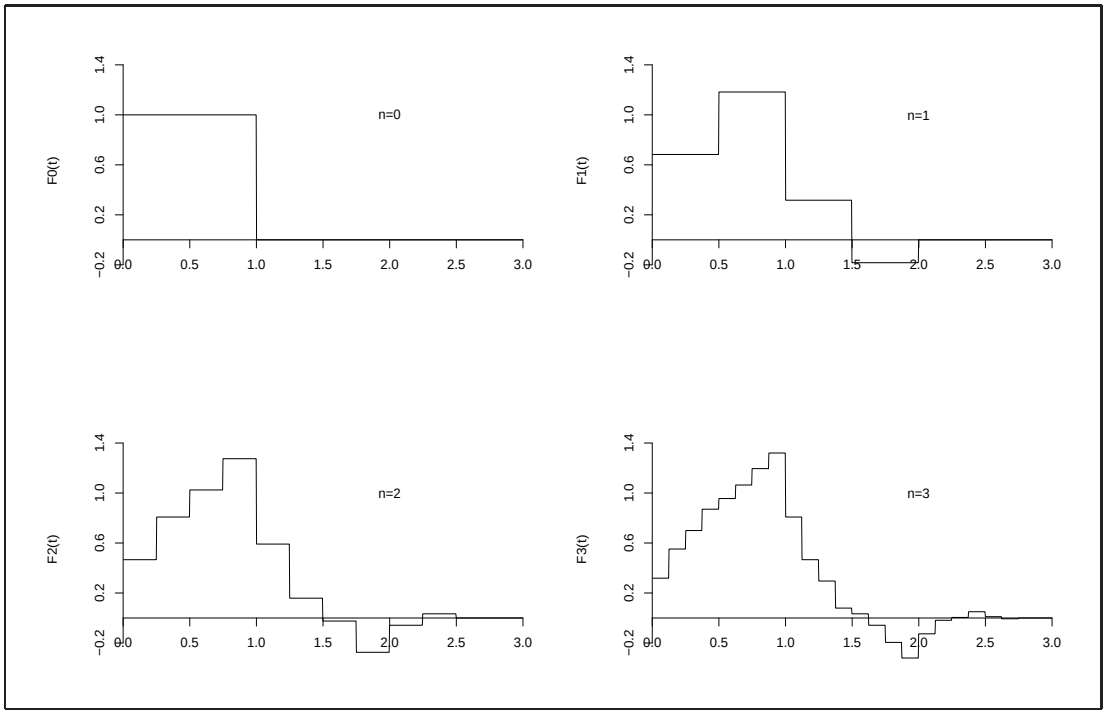

Abbildung 2.4: Daubechies' Konstruktion einer Scaling Function: Die Funktionen $\mathcal{F}^n \mathbf{1}_{[0,1)}$ mit $n = 0, 1, 2, 3$.

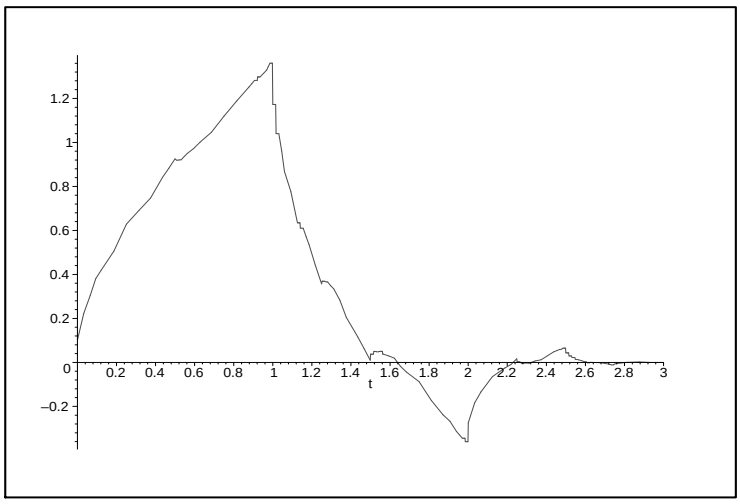

Abbildung 2.5: Daubechies' Building Block ϕ

Kapitel 3. Stochastische Prozesse

In diesem Kapitel werden wichtige Begriffe aus der Theorie stochastischer Prozesse eingeführt und ein Überblick über die Methoden der klassischen Zeitreihenanalyse gegeben. Das Ziel ist jedoch keine ausführliche Abhandlung zu diesem Themenbereich, vielmehr wird für ein vertieftes Studium auf die beiden Bücher von Brockwell und Davis (1991 bzw. 1996) verwiesen. Ersteres ist eher theoretisch aufgebaut, das Zweite anwendungsbezogen. In Brockwell / Davis (1991) findet man auch sämtliche in den Abschnitten 3.2 bis 3.4 nicht ausgeführten Beweise.

3.0.1 Definition: Ein *stochastischer Prozess* $(X_t)_{t\in T}$, ist eine Familie von Zufallsvariablen $X_t,\ t \in T$, mit Zeitparameter-Menge $T \subset \mathbb{R}$. Eine Realisierung von (X_t) wird *Pfad* des Prozesses genannt. Ist T diskret, so spricht man von einer *Zeitreihe.*

Im Folgenden sei $T = \mathbb{Z}$. Eine Zeitreihe (X_t) wird üblicherweise gemäß

$$X_t = m_t + s_t + Y_t$$

in eine rein deterministische Komponente, bestehend aus einem Trend (m_t) und gegebenenfalls einem Zyklus (s_t) und eine rein zufällige Komponente (Y_t) aufgespalten. In der Regel eliminiert man zunächst die deterministischen Anteile in der Hoffnung, im Anschluss mit einem *stationären* Prozess (s. Abschnitt 3.2) arbeiten zu können.

3.0.2 Bemerkung: Bei der Eliminierung des Trends m_t ist zu beachten, dass die vom geschätzten Trend $\hat{m}_t$ bereinigte Zeitreihe $(X_t - \hat{m}_t)$ in der Regel korreliert ist selbst im Falle einer unkorrelierten Zeitreihe (X_t).

3.1 Schätzung bzw. Eliminierung des Trends

3.1.1 Filtern der Zeitreihe mittels eines Moving Average

Liegt keine zyklische Komponente vor, so kann ein evtl. vorhandener Trend der Zeitreihe (X_t) durch einen *Moving Average*

$$\hat{m}_t := \frac{1}{2q+1} \sum_{j=-q}^{q} X_{t+j},\ t = q+1, \ldots, n-q, \tag{3.1}$$

mit geeignetem $q \in \mathbb{N}$ geschätzt werden. Der Moving Average bildet zu jedem Zeitpunkt t das Mittel über die $2q+1$ umliegenden Zeitpunkte und schätzt somit den Trend. Die Differenz $X_t - \hat{m}_t$ sollte daher annähernd von einem Trend befreit sein.

3.1.1 Bemerkung: Der Moving Average (3.1) ist ein *linearer Filter* (siehe Definition 3.6.4), der die hochfrequenten Anteile der Zeitreihe (X_t) herausfiltert und somit die Zeitreihe glättet. Dieser glättende Effekt lässt sich auch mit anderen linearen Filtern, z.B. einem gewichteten Moving Average $\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t+j}$ mit geeigneten Gewichten $a_j,\ j \in \mathbb{Z}$, erzielen (s. Brockwell / Davis (1991)).

3.1.2 Bemerkung: Bei Vorliegen einer zyklischen Komponente (z.B. bei jahreszeitabhängigen Daten) muss zunächst diese eliminiert werden, bevor der Trend mittels eines Moving Average geschätzt werden kann. Siehe hierzu Brockwell / Davis (1991).

3.1.2 Kleinste-Quadrate-Schätzung

In der Regel wird ein stationärer Prozess $(X_t)_{t\in T}$ nur an endlich vielen Stellen $t_1, \ldots, t_n$ beobachtet. Daher kann man

$$X_t = m_t + Y_t,\ t = t_1, \ldots, t_n,$$

auch als *lineares Modell* betrachten (siehe Abschnitt 5.1). Man setzt die funktionale Form des Trends

$$m_t = \beta_1 f_1(t) + \cdots + \beta_m f_m(t)$$

mit unbekannten Parametern $\beta_1, \ldots, \beta_m$ an. Nun kann mit den in Abschnitt 5.1 beschriebenen Methoden der Kleinste-Quadrate-Schätzer des Parametervektors $\beta = (\beta_1, \ldots, \beta_m)^\top$ bestimmt werden. Dieser Schätzer minimiert

$$\sum_{t=t_1}^{t_n} [X_t - (b_1 f_1(t) + \cdots + b_m f_m(t))]^2$$

bezüglich $b = (b_1, \ldots, b_m)^\top \in \mathbb{R}^m$. Es empfiehlt sich, $f_1 \equiv 1$ zu setzen.

3.1.3 Differenzenbildung

Gegeben sei eine Zeitreihe

$$X_t = m_t + Y_t,\ t \in T. \tag{3.2}$$

Man definiert nun den *Backward Shift Operator* B durch

$$\begin{aligned} BX_t &= X_{t-1}, \\ B^2 X_t &= X_{t-2}, \end{aligned}$$

allgemein

$$B^k X_t = X_{t-k},\ k \in \mathbb{N}, \tag{3.3}$$

sowie den *Differenzenoperator* ∇ durch

$$\begin{aligned} \nabla X_t &= (1-B)X_t = X_t - X_{t-1}, \\ \nabla^2 X_t &= (1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2}, \end{aligned}$$

allgemein

$$\nabla^k X_t = (1-B)^k X_t,\ k \in \mathbb{N}. \tag{3.4}$$

Besitzt die Zeitreihe (3.2) einen linearen Trend $m_t = at + b$, so besitzt $\nabla X_t = \nabla m_t + \nabla Y_t$ den konstanten Trend $\nabla m_t = b$. Ein linearer Trend lässt sich also durch Differenzenbildung eliminieren (bis auf eine Konstante).

Kann man den Trend durch einen quadratischen Term beschreiben, gilt also $m_t = at^2 + bt + c$, so ist

$$\nabla m_t = (2a)t + (b - a)$$

linear in t und

$$\nabla^2 m_t = 2a$$

konstant. Eine quadratische Erwartungswertfunktion kann also durch eine zweimalige Anwendung des Differenzenoperators ∇ eliminiert werden.

Ist allgemein der Trend m_t modellierbar durch ein Polynom k-ten Grades, so besitzt

$$\nabla^k X_t = (1-B)^k X_t$$

einen konstanten Trend.

3.1.3 Bemerkung: Betrachtet man die Zeitreihe (3.2) als lineares Modell (s. Abschnitt 5.1), so entspricht die Anwendung des Differenzenoperators ∇ der Bildung von Pseudo-Residuen, siehe Abschnitt 5.2.4.

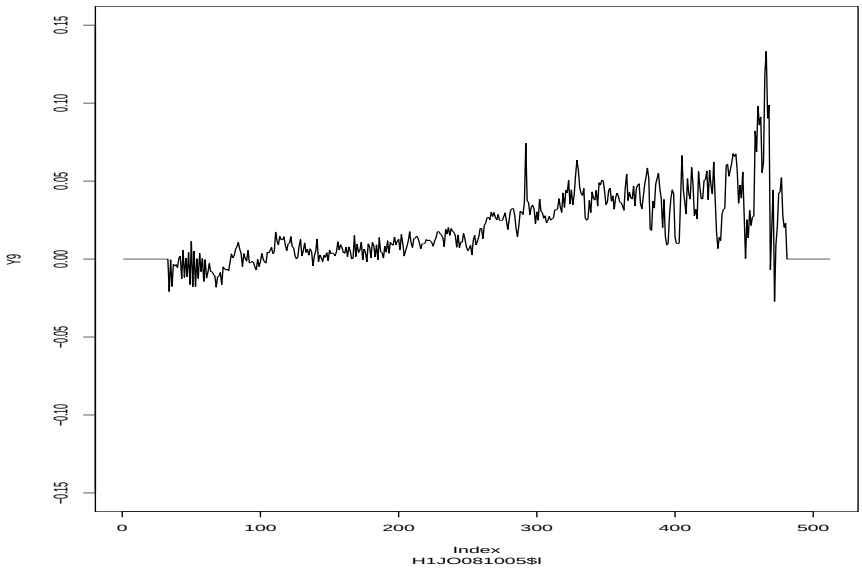

Abbildung 3.1: Die GPS-Zeitreihe H1JO081005

3.1.4 Trendschätzung mit Wavelets

Eine Multiresolution Analysis (s. Abschnitt 2.2.1) eines stochastischen Prozesses (X_t) spaltet den Prozess in verschiedene hoch- und niederfrequente Anteile auf. Dies ermöglicht sowohl ein Entfernen der hochfrequenten Anteile und somit ein Ausschalten des "Rauschens", als auch die Schätzung und Eliminierung des Trends. Dieses Vorgehen soll nun anhand der GPS-Zeitreihe aus Abbildung 3.1 und der Haar-Wavelet-Basis veranschaulicht werden.

Die Abbildungen 3.2 und 3.3 zeigen eine Multiresolutionanalysis der Zeitreihe aus Abbildung 3.1. Addiert man die zu niedrigen Frequenzen korrespondierenden Anteile auf, erhält man den geschätzten Trend (Abb. 3.4 oben). Die Summe der "hochfrequenten" Komponenten ergibt die vom Trend bereinigte Zeitreihe (Abb. 3.4 unten). Dabei muss ein geeignetes Kriterium gefunden werden, bei welchem Level die Grenze zwischen nieder- und hochfrequenten Anteilen gezogen wird.

3.2 Stationäre Prozesse

3.2.1 Definition: Ein stochastischer Prozess heißt *(schwach) stationär,* falls

(*i*) $\mathrm{E}(X_t^2) < \infty$ für alle $t \in \mathbb{Z}$,

(*ii*) $\mathrm{E}(X_t) = \mu$ für alle $t \in \mathbb{Z}$, (d.h. die *Erwartungswertfunktion* ist konstant),

(*iii*) $\mathrm{Cov}(X_{t+h}, X_t) = \mathrm{Cov}(X_h, X_0)$ für alle $t, h \in \mathbb{Z}$. Das bedeutet, dass die Kovarianz zwischen zwei Zeitpunkten nur von deren Abstand, nicht aber von t abhängt.

3.2.2 Bemerkung: Wurde der Trend der Zeitreihe mit Methoden aus den Abschnitten 3.1.1, 3.1.2 oder 3.1.4 geschätzt und eliminiert, ist der Erwartungswert der vom Trend bereinigten Zeitreihe gleich 0. In (*ii*) kann daher ohne Einschränkung $\mu = 0$ angenommen werden.

3.2.3 Bemerkung: Aus (*iii*) folgt insbesondere (setze $h = 0$), dass die Varianzfunktion $\sigma^2(t) := \mathrm{Var}(X_t)$ einer stationären Zeitreihe konstant gleich $\sigma^2 := \mathrm{Var}(X_0)$ ist.

Von besonderer Bedeutung in der Theorie stationärer Zeitreihen ist der *White Noise* Prozess.

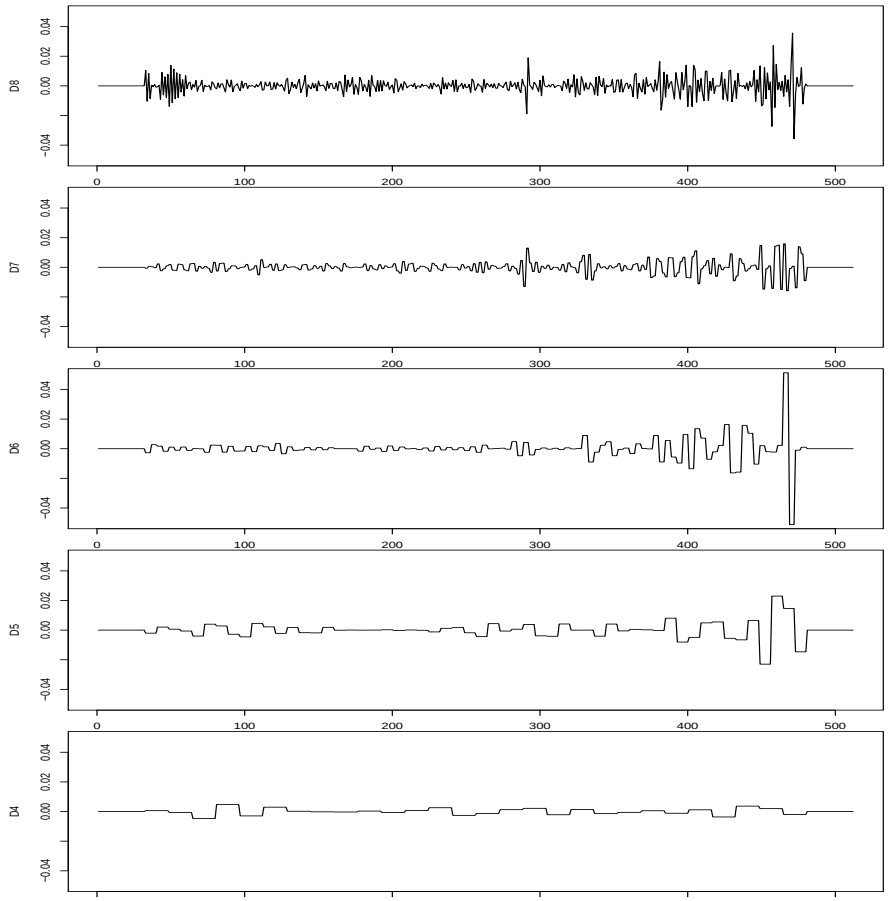

Abbildung 3.2: MRA der GPS-Zeitreihe H1JO081005 (1)

3.2.4 Definition: Ein stochastischer Prozess (Z_t) heißt *White Noise* Prozess, falls $\mathrm{E}(Z_t) = 0$ für alle $t \in \mathbb{Z}$ und

$$\mathrm{Cov}(Z_t, Z_s) = \begin{cases} 0, & s \neq t, \\ \sigma^2, & s = t. \end{cases}$$

Man schreibt $Z_t \sim WN(0, \sigma^2)$. Sind die $Z_t,\ t \in \mathbb{Z}$, sogar unabhängig und identisch verteilt, so schreibt man $Z_t \sim \mathit{IID}(0, \sigma^2)$.

3.2.1 *ARMA*-Prozesse

Eine wichtige Klasse stationärer Zeitreihen sind die *ARMA*-Prozesse. Eine stationäre Zeitreihe $(X_t),\ t \in \mathbb{Z}$, heißt *ARMA*(p, q)*-Prozess*, falls (X_t) darstellbar ist in der Form

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}. \tag{3.5}$$

Dabei sind die ϕ_i und θ_j reelle Parameter $(i = 1, \ldots, p,\ j = 1, \ldots, q;\ p, q \in \mathbb{N} \cup 0)$ und (Z_t) ist ein White Noise Prozess. Spezialfälle von *ARMA*-Prozessen sind *Moving Average (MA)* Prozesse

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}\ (\mathrm{MA}(q))$$

und *Autoregressive (AR)* Prozesse

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t\ (\mathrm{AR}(p)). \tag{3.6}$$

Auch ein White Noise Prozess ist ein spezieller *ARMA*-Prozess. Die Abbildungen 3.5, 3.6 und 3.7 zeigen Realisierungen eines White Noise Prozesses, eines Moving Average Prozesses bzw. eines autoregressiven Prozesses.

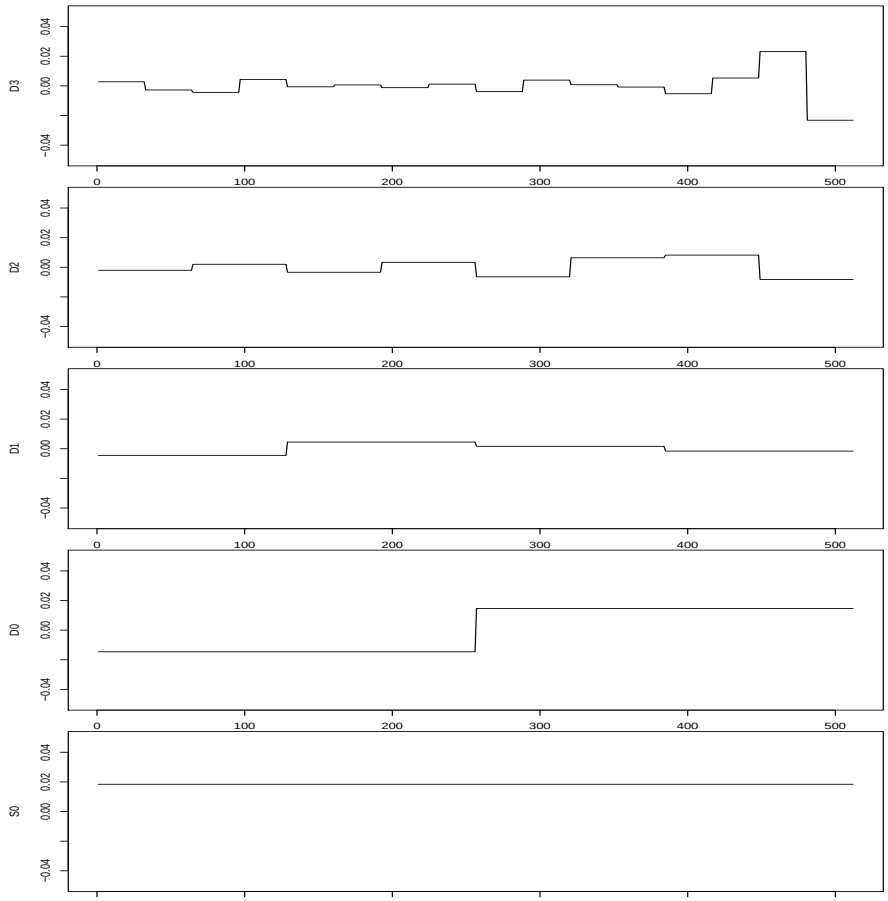

Abbildung 3.3: MRA der GPS-Zeitreihe H1JO081005 (2)

3.2.5 Definition: Für einen stationären Prozess $(X_t),\ t \in \mathbb{Z}$, wird durch

$$\gamma(h) := \mathrm{Cov}(X_h, X_0),\ h \in \mathbb{Z},\ \text{und}\ \rho(h) := \mathrm{Corr}(X_h, X_0) = \frac{\gamma(h)}{\gamma(0)}$$

die *Autokovarianz-* bzw. die *Autokorrelationsfunktion (ACF)* von (X_t) definiert.

3.2.6 Beispiel: Die Autokorrelationsfunktion eines White Noise Prozesses ist (unabhängig von σ^2)

$$\rho(h) = \begin{cases} 1, & h = 0, \\ 0, & h \neq 0. \end{cases}$$

3.2.7 Beispiel: Für die Autokorrelationsfunktion eines $MA(q)$-Prozesses

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2),$$

gilt stets $\rho(h) \neq 0$ für $h \leq q$ und $\rho(h) = 0$ für $h \geq q+1$. Ein $MA(1)$-Prozess z.B. besitzt die Autokorrelationsfunktion

$$\rho(h) = \begin{cases} 1, & h = 0, \\ \frac{\theta_1}{1+\theta_1^2}, & |h| = 1, \\ 0, & |h| \geq 2. \end{cases}$$

Ein autoregressiver Prozess $X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t$ hingegen besitzt eine betragsmäßig schwächer abnehmende Autokorrelationsfunktion. Dabei nimmt die ACF betragsmäßig umso schwächer ab, je größer p ist. Die ACF eines AR(1)-Prozesses z.B. ist gegeben durch

$$\rho(h) = \phi_1^h,\ h \in \mathbb{N} \cup \{0\}.$$

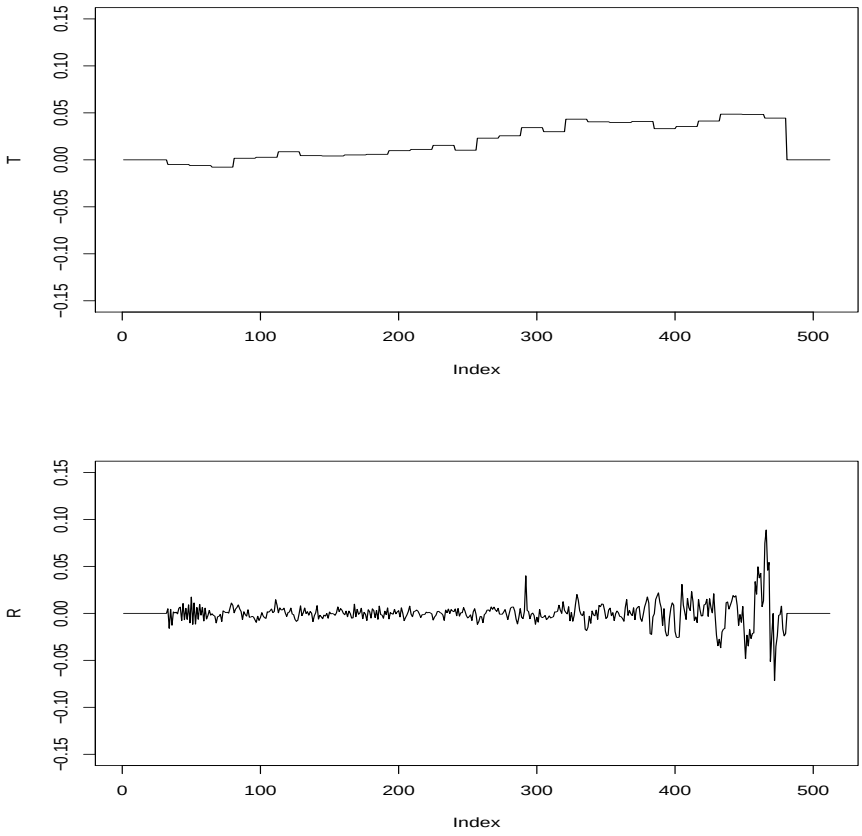

Abbildung 3.4: Trend und die vom Trend bereinigte Zeitreihe zu H1JO081005

Sind n Beobachtungen einer stationären Zeitreihe (X_t) gegeben, so schätzt man $\mathrm{E}(X_t)$ durch das empirische Mittel $\overline{X} := \frac{1}{n}\sum_{t=1}^{n} X_t$. Die Autokovarianz- und die Autokorrelationsfunktion können geschätzt werden durch die *empirische Autokovarianzfunktion*

$$\hat{\gamma}_n(h) := \frac{1}{n}\sum_{t=1}^{n-|h|}(X_{t+|h|} - \overline{X})(X_t - \overline{X}),\ -n < h < n, \tag{3.7}$$

bzw. die *empirische Autokorrelationsfunktion (sample ACF)*

$$\hat{\rho}_n(h) := \frac{\hat{\gamma}_n(h)}{\hat{\gamma}_n(0)}. \tag{3.8}$$

Die Abbildungen 3.8, 3.9 und 3.10 zeigen die empirischen Autokorrelationsfunktionen der Zeitreihen aus den Abbildungen 3.5, 3.6 bzw. 3.7.

3.2.2 Der lineare Prozess in seiner allgemeinen Form

Im Folgenden soll eine Zeitreihe (X_t) als *allgemeiner linearer Prozess* bezeichnet werden, falls eine Folge $(\psi_j) \subset \mathbb{R}$ existiert mit

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty, \tag{3.9}$$

so dass für alle $t \in T$ gilt

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},\ Z_t \sim WN(0, \sigma^2). \tag{3.10}$$

Zur Konvergenz der Reihe in (3.10) s. Brockwell / Davis (1991).

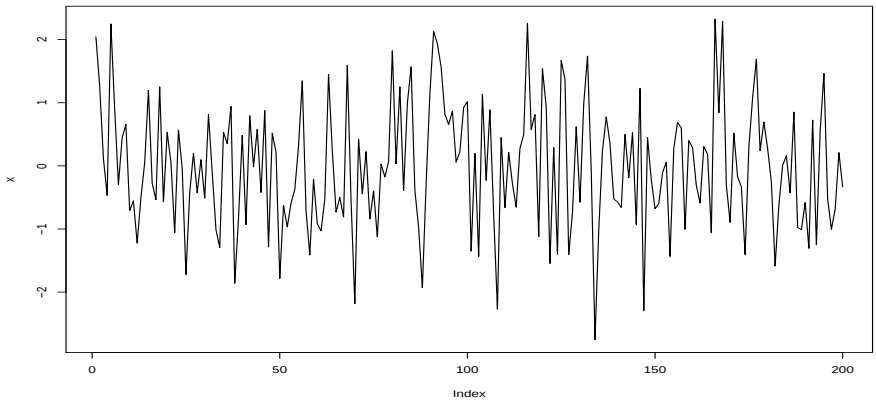

Abbildung 3.5: Ein White Noise Prozess mit $\sigma^2 = 1$

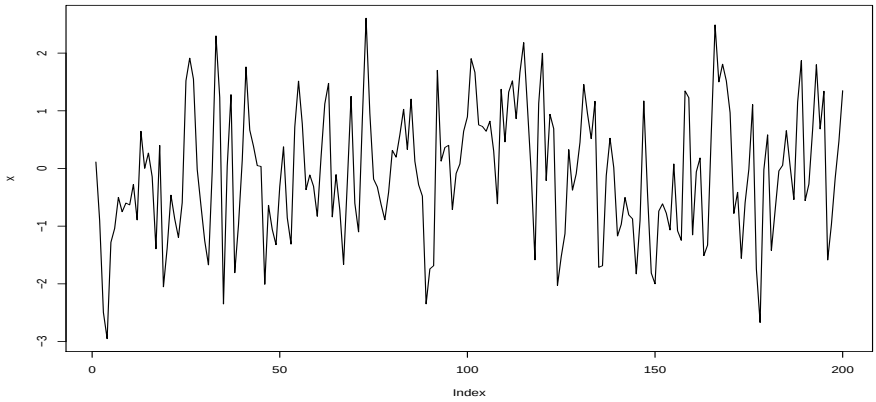

Abbildung 3.6: Ein $MA(1)$ Prozess mit $\theta_1 = \frac{1}{2}$ und $\sigma^2 = 1$

3.2.8 Bemerkung: Aufgrund der Stationarität des White Noise Prozesses ist der allgemeine lineare Prozess ebenfalls stationär mit (Beweis: Brockwell / Davis (1991))

$$\begin{aligned}
\mathrm{E}(X_t) &= \sum_{j=-\infty}^{\infty} \psi_j \cdot \mu_Z = 0 \ (\mu_Z := \mathrm{E}(Z_0)), \\
\mathrm{Var}(X_t) &= \sum_{j=-\infty}^{\infty} \psi_j^2 \cdot \sigma_Z^2 \ (\sigma_Z^2 := \mathrm{Var}(Z_0)), \\
\mathrm{Cov}(X_{t+h}, X_t) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \ \mathrm{Cov}(Z_{t+h-i}, Z_{t-j}) \\
&= \sum_{i=-\infty}^{\infty} \psi_i \psi_{i-h} \ \mathrm{Cov}(Z_{t+h-i}, Z_{t+h-i}) \\
&= \sum_{i=-\infty}^{\infty} \psi_i \psi_{i-h} \ \sigma_Z^2.
\end{aligned}$$

3.2.9 Beispiel: Ist (X_t) ein $ARMA(p, q)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \ (Z_t) \sim WN(0, \sigma^2)$$

mit

$$\phi(z) := 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \neq 0 \text{ für alle } z \in \mathbb{C} \text{ mit } |z| = 1,$$

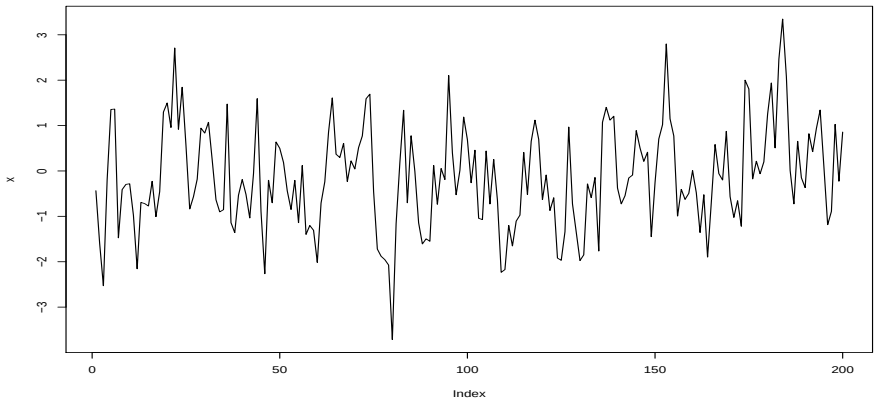

Abbildung 3.7: Ein AR(1) Prozess mit $\phi_1 = \frac{1}{2}$ und $\sigma^2 = 1$

dann existieren eindeutig bestimmte Koeffizienten $\psi_j,\ j \in \mathbb{Z}$, mit $X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, d.h. (X_t) ist ein allgemeiner linearer Prozess .

(Beweis: Brockwell / Davis (1991).)

3.2.10 Satz: *Erfüllt der allgemeine lineare Prozess*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},\ (Z_t) \sim \mathit{IID}(0, \sigma^2)$$

die Bedingungen

(i) $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ *und*

(ii) $E(Z_t^4) < \infty$,

so gilt für die empirische Autokorrelationsfunktion $\hat{\rho}(\cdot)$ *von* (X_t)

$$\sqrt{n} \begin{pmatrix} \hat{\rho}(1) \\ \vdots \\ \hat{\rho}(h) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\begin{pmatrix} \rho(1) \\ \vdots \\ \rho(h) \end{pmatrix}, W),\ h \in \mathbb{N}. \tag{3.11}$$

Dabei bedeutet "$\xrightarrow{\mathcal{D}}$" Verteilungskonvergenz (s. Anhang B) und W bezeichnet die Kovarianzmatrix, deren Einträge an den Stellen (i, j), $i, j = 1, \ldots, h$, gegeben sind durch

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \cdot \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}.$$

Im Falle $(X_t) \sim \mathit{IID}(0, \sigma^2)$ *ist W die Einheitsmatrix.*

Der Beweis ist in Brockwell / Davis (1991) zu finden.

3.2.11 Definition: In Bezug auf (3.11) sagt man auch, der Zufallsvektor $\hat{\rho} := (\hat{\rho}(1), \ldots, \hat{\rho}(h))^\top$ ist *asymptotisch normalverteilt* mit Erwartungswert-Vektor $\rho := (\rho(1), \ldots, \rho(h))^\top$ und Kovarianz-Matrix $\frac{1}{n}W$, in Zeichen

$$\hat{\rho} \sim \mathrm{AN}(\rho, \frac{1}{n}W).$$

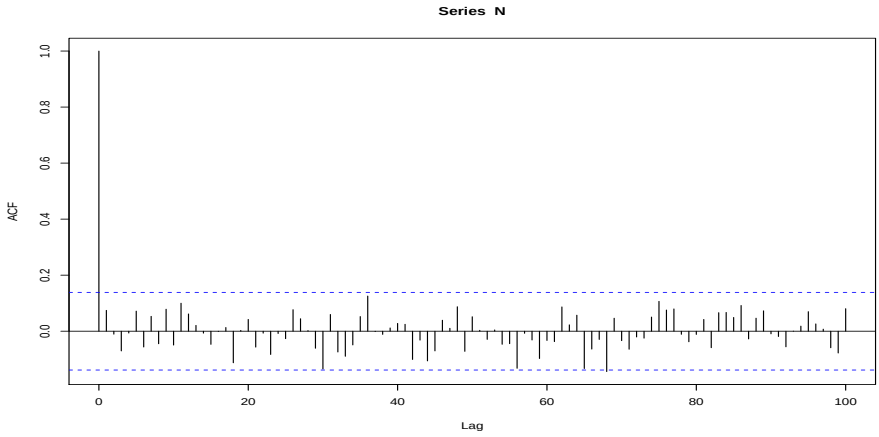

Abbildung 3.8: Empirische Autokorrelationsfunktion eines White Noise Prozesses

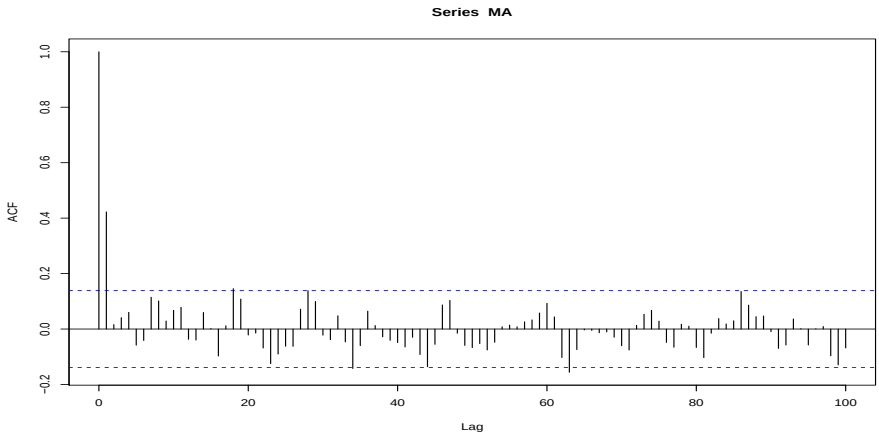

Abbildung 3.9: Empirische Autokorrelationsfunktion eines $MA(1)$-Prozesses

Die Kenntnis der asymptotischen Verteilungen der $\hat{\rho}(h)$ erlaubt es, Konfidenzbereiche für $\hat{\rho}(h)$ zu berechnen. Die meisten Statistikpakete geben zu einer Autokorrelationsfunktion stets die 95%-Konfidenzschranken $\pm 1.96 n^{-1/2}$ eines Gauß'schen White Noise Prozesses mit aus. Das entspricht einer oberen und unteren Grenze, die unter Vorliegen eines White Noise Prozesses in ca. 95% aller "Time Lags" h von der empirischen Korrelation $\hat{\rho}(h)$ nicht überschritten werden darf.

Abbildung 3.8 zeigt die empirische Autokorrelationsfunktion des White Noise Prozesses aus Abbildung 3.5 mit dem entsprechenden 95%-Konfidenzbereich (horizontale Linien).

In Abbildung 3.9 ist die empirische Autokorrelationsfunktion des $MA(1)$- Prozesses aus Abbildung 3.6 zu sehen. Wie aufgrund der theoretischen ACF zu erwarten ist, überschreitet $\hat{\rho}(h)$ für $h = 0$ und $h = 1$ die Grenzen des 95%-Konfidenzbereichs eines White-Noise-Prozesses. Für $h \geq 2$ bleibt $\hat{\rho}(h)$ in 95% aller Time Lags innerhalb dieser Grenzen. Abbildung 3.10 zeigt die empirische ACF des $AR(1)$-Prozesses aus Abbildung 3.7.

3.2.12 Definition: Ein $ARMA(p, q)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \ (Z_t) \sim WN(0, \sigma^2),$$

heißt *zukunftsunabhängig* oder *nicht vorgreifend* (engl. *causal*), falls eine reelle Folge (ψ_j) existiert mit $\sum_{j=0}^{\infty} |\psi_j| < \infty$ und

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \ t \in \mathbb{Z}. \tag{3.12}$$

Ein nicht vorgreifender Prozess ist also ein allgemeiner linearer Prozess mit $\psi_j = 0$ für $j < 0$. Der nicht vorgreifende Prozess ist somit unabhängig von der "Zukunft" des White Noise Prozesses (Z_t). Man nennt (3.12) auch einen *MA*(∞)*-Prozess.*

3.2.13 Lemma: *Der MA*(∞)*-Prozess (3.12) besitzt die Erwartungswertfunktion* $\mu \equiv 0$ *und die Kovarianzfunktion*

$$\gamma(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|} \ .$$

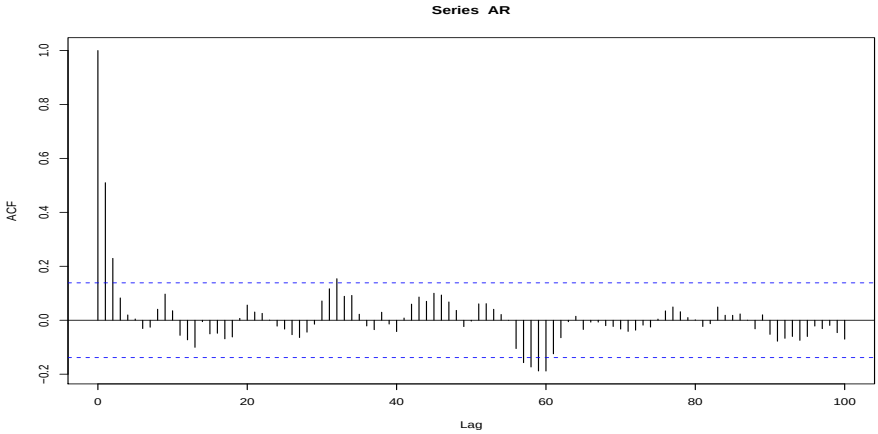

Abbildung 3.10: Empirische Autokorrelationsfunktion eines AR(1)*-Prozesses*

3.2.14 Satz: *Es sei* (X_t) *ein ARMA*(p, q)*-Prozess, für den die Polynome*

$$\begin{aligned} \phi(z) &:= 1 - \phi_1 z - \cdots - \phi_p z^p \\ \theta(z) &:= 1 + \theta_1 z + \cdots + \theta_q z^q \end{aligned}$$

keine gemeinsamen Nullstellen besitzen. Dann ist (X_t) *genau dann nicht vorgreifend, wenn*

$$\phi(z) \neq 0 \text{ für alle } z \in \mathbb{C} \text{ mit } |z| \leq 1,$$

d.h., wenn $\phi(\cdot)$ *keine Nullstelle auf der abgeschlossenen Einheitskreisscheibe besitzt.*

Die Koeffizienten ψ_j $(j = 1, 2, \dots)$ *in der Darstellung (3.12) sind gegeben durch*

$$\psi(z) := \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \ |z| \leq 1. \tag{3.13}$$

Beweis: Brockwell / Davis (1991).

Mit $\theta_0 := 1$, $\theta_j := 0$ für $j > q$ und $\phi_j := 0$ für $j > p$ lassen sich die Koeffizienten ψ_j, $j = 0, 1, 2, \dots$ aus (3.13) konkret berechnen gemäß (Koeffizientenvergleich, s. Brockwell / Davis (1991))

$$\begin{aligned} \psi_0 &= \theta_0 = 1, \\ \psi_1 &= \theta_1 + \psi_0 \phi_1 = \theta_1 + \phi_1 \\ \psi_2 &= \theta_2 + \psi_0 \phi_2 + \psi_1 \phi_1 = \theta_2 + \phi_2 + \theta_1 \phi_1 + \phi_1^2 \\ &\vdots \\ \psi_j &= \theta_j + \sum_{i=1}^{j} \phi_i \psi_{j-i}. \end{aligned} \tag{3.14}$$

3.2.15 Beispiel: Es sei (X_t) ein $AR(1)$-Prozess

$$X_t - \phi_1 X_{t-1} = Z_t,\ (Z_t) \sim WN(0, \sigma^2),$$

mit $\phi_1 \in (-1, 1)$. Dann ist $1 - \phi_1 z \neq 0$ für alle $z \in \mathbb{C}$ mit $|z| \leq 1$. Somit ist (X_t) nicht vorgreifend. Die Koeffizienten gemäß der Darstellung (3.12) sind gegeben durch $\psi_j = \phi_1^j,\ j \in \mathbb{N} \cup \{0\}$.

3.2.16 Definition: Ein $ARMA(p, q)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2),$$

heißt *invertierbar*, falls eine reelle Folge (π_j) existiert mit $\sum_{j=0}^{\infty} |\pi_j| < \infty$ und

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j},\ t \in \mathbb{Z}. \tag{3.15}$$

3.2.17 Satz: *Es sei (X_t) ein $ARMA(p, q)$-Prozess, für den die Polynome $\phi(z)$ und $\theta(z)$ keine gemeinsamen Nullstellen besitzen. Dann ist (X_t) genau dann invertierbar, wenn*

$$\theta(z) \neq 0 \text{ für alle } z \in \mathbb{C} \text{ mit } |z| \leq 1.$$

Die Koeffizienten π_j $(j = 1, 2, \dots)$ in der Darstellung (3.15) sind gegeben durch

$$\pi(z) := \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)},\ |z| \leq 1.$$

Den Beweis dieser zu Satz 3.2.14 symmetrischen Aussage findet man in Brockwell / Davis (1991).

Für die Modellierung von *ARMA*-Prozessen ist der nächste Satz von Bedeutung:

3.2.18 Satz: *Es sei (X_t) der $ARMA(p, q)$-Prozess*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2),$$

mit $\phi(z) \neq 0$ und $\theta(z) \neq 0$ für alle $z \in \mathbb{C}$ mit $|z| = 1$. Dann existiert ein Polynom $\tilde{\phi}(z)$ vom Grad p und ein Polynom $\tilde{\theta}(z)$ vom Grad q ohne Nullstellen auf der abgeschlossenen Einheitskreisscheibe $\{z \in \mathbb{C} : |z| \leq 1\}$ und ein White Noise Prozess (Z_t^*), so dass

$$X_t - \tilde{\phi}_1 X_{t-1} - \cdots - \tilde{\phi}_p X_{t-p} = Z_t^* + \tilde{\theta}_1 Z_{t-1}^* + \cdots + \tilde{\theta}_q Z_{t-q}^*,\ (Z_t^*) \sim WN(0, \tilde{\sigma}^2), \tag{3.16}$$

nicht vorgreifend und invertierbar ist. Man spricht in Bezug auf (3.16) von der nicht vorgreifenden und invertierbaren Darstellung von (X_t).

Beweis: Brockwell / Davis (1991).

3.2.19 Bemerkung: Ist

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2), \tag{3.17}$$

ein $ARMA(p, q)$-Prozess und bezeichnen $\phi(\cdot)$ und $\psi(\cdot)$ die Polynome

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \cdots - \phi_p z^p \\ \text{bzw. } \theta(z) &= 1 + \theta_1 z + \cdots + \theta_q z^q, \end{aligned}$$

so schreibt man an Stelle von (3.17) auch

$$\phi(B) X_t = \theta(B) Z_t,$$

wobei B der Backward Shift Operator (3.3) ist.

3.2.3 Long Memory Prozesse

Für $j \in \mathbb{N}$ sei $(X_t^{(j)})_{t\in T}$ ein stationärer Prozess mit "Short Memory", d.h. für "kleine" $h \in \mathbb{Z}$ sind $X_t^{(j)}$ und $X_{t+h}^{(j)}$ (positiv) korreliert. Man betrachte nun den kumulierten Prozess $(X_t)_{t\in T}$ mit

$$X_t := \sum_{j=1}^{\infty} X_t^{(j)}.$$

Hier überlagern sich Abhängigkeiten der individuellen Prozesse $(X_t^{(j)})$, so dass (X_t) sehr bizarre Eigenschaften besitzen kann, zum Beispiel:

a) Eine zeitweise (Tage / Wochen ...) nahezu konstante Mean Funktion, die sich jedoch im Laufe der Zeit verändert,

b) Betrachtet man relativ kurze Zeitabschnitte (einige Monate / Jahre), so scheint ein zyklisches Verhalten vorzuliegen. Der Prozess als Ganzes zeigt jedoch keinen kontinuierlichen Zyklus. Vielmehr scheinen (fast) alle "Frequenzen" in zufälliger Folge vorzukommen.

Zeitreihen, die ein solches Verhalten zeigen, sind zum Beispiel

1.) Jährliche Wasserstands-Minima des Nils. Historisch von Bedeutung sind die Daten aus den Jahren 622-1281 u.Z., gemessen in der Nähe von Kairo. Anhand dieser Daten wurden u.A. die oben beschriebenen Effekte untersucht sowie statistische Modelle für Prozesse mit diesen Eigenschaften eingeführt.

2.) Temperaturdaten, z.B. die monatliche Temperatur in Norwich, England, von 1854-1989,

3.) Viele ökonomische Zeitreihen.

Allen diesen Beispielen gemeinsam ist die offensichtliche Überlagerung veschiedener quasi-zyklischer Einflüsse. Die Beispiele machen aber auch deutlich, dass es unmöglich ist, alle individuellen Einflüsse zu modellieren, einige dieser Einflüsse können sogar vollständig unbekannt sein. Sucht man für einen solchen Prozess ein Modell, so bleibt im Allgemeinen nichts anderes übrig, als die bekannten Einflüsse zu modellieren und die entsprechenden Residuen wiederum als stochastischen Prozess zu betrachten und weiter zu untersuchen.

3.2.20 Definition: Es sei (X_t) ein stationärer Prozess mit Autokorrelationsfunktion ρ. Weiter existiere eine reelle Zahl $\beta \in (0,1)$ und ein $c_\rho > 0$ mit

$$\frac{\rho(h)}{h^{-\beta}} \overset{h\to\infty}{\longrightarrow} c_\rho.$$

Dann heißt (X_t) ein stationärer Prozess mit *Long Memory* oder *Long Range Korrelationen.*

Man beachte, dass die Definition von Long-Memory-Prozessen lediglich eine Aussage über die Korrelationen zweier Beobachtungen macht, deren zeitlicher Abstand gegen unendlich geht. Es wird dabei kein absoluter Wert angegeben, verlangt wird nur ein langsames Nachlassen der Korrelationen. Korrelationen zwischen Beobachtungen mit endlichem Abstand dürfen dabei beliebig klein sein. Somit kann selbst in einem Fall, in dem die empirischen Autokorrelationen wegen ihrer Geringfügigkeit einen Schluss auf Unkorreliertheit zulassen, ein Long Memory Prozess mit langsam abnehmenden Korrelationen vorliegen.

Literatur: z.B. Beran (1994).

3.3 Vorhersage stationärer Prozesse

Häufig besteht der Wunsch, das Verhalten eines stochastischen Prozesses in der Zukunft zu prognostizieren, d.h., aufgrund von vergangenen Beobachtungen $X_1, \ldots, X_n$ zukünftige Werte $X_{n+h},\ h \in \mathbb{N}$, vorherzusagen. In diesem Rahmen soll kurz auf die lineare Vorhersage $\hat{X}_{n+1}$ von X_{n+1} eingegangen werden. Für eine ausführliche Behandlung dieses Themas, auch für $h \geq 2$, wird auf Brockwell / Davis (1991) verwiesen.

Es sei (X_t) ein stationärer Prozess mit $\mathrm{E}X_t \equiv 0$ und Autokovarianzfunktion $\gamma(\cdot)$. Gesucht ist eine lineare Vorhersage für X_{n+1}, also eine Vorhersage der Form

$$\begin{aligned} \hat{X}_{n+1} &= \phi_{n1}X_n + \cdots + \phi_{nn}X_1 \\ &= \sum_{i=1}^{n} \phi_{ni}X_{n+1-i},\ n \geq 1,\ \phi_{ni} \in \mathbb{R}\ (i = 1, \ldots, n). \end{aligned} \tag{3.18}$$

Wegen der Existenz der Kovarianzfunktion γ sind sämtliche Zufallsvariablen $X_t,\ t \in T$, Elemente aus $L^2(P)$ (vgl. Anhang B.3). Bezeichnet $\mathcal{H}_{1,n} := \overline{\mathrm{sp}}\{X_1, \ldots, X_n\}$ den (abgeschlossenen) Untervektorraum des $L^2(P)$, der von den Zufallsvariablen $X_1, \ldots, X_n$ aufgespannt wird (also die Menge aller Linearkombinationen von $X_1, \ldots, X_n$), so ist offensichtlich $\hat{X}_{n+1}$ ein Element aus $\mathcal{H}_{1,n}$.

Wie kann man nun die Koeffizienten $\phi_{n1}, \ldots, \phi_{nn}$ in (3.18) so wählen, dass man eine möglichst gute Prognose bekommt? Was ist überhaupt eine "gute" Prognose?

Auf dem vollständigen normierten Raum $L^2(P)$ ist durch $\|X-Y\|$ (s. Anhang B.3) der Abstand zweier Elemente X und Y gegeben. Unter der "besten" linearen Vorhersage $\hat{X}_{n+1}$ versteht man nun jenes Element aus $\mathcal{H}_{1,n}$, das in $L^2(P)$ den geringsten Abstand zu X_{n+1} besitzt, für das also gilt

$$\|X_{n+1} - \hat{X}_{n+1}\| = \min_{Y \in \mathcal{H}_{1,n}} \|X_{n+1} - Y\|.$$

Nach dem Projektionssatz (s. Anhang B.3) ist $\hat{X}_{n+1}$ eindeutig bestimmt durch die Orthogonalprojektion von X_{n+1} auf $\mathcal{H}_{1,n}$ und $X_{n+1} - \hat{X}_{n+1}$ ist orthogonal zu $\mathcal{H}_{1,n}$. Somit ist $X_{n+1} - \hat{X}_{n+1}$ orthogonal zu $X_{n+1-j},\ j = 1, \ldots, n$, d.h.

$$\langle \hat{X}_{n+1} - X_{n+1}, X_{n+1-j} \rangle = 0,\ j = 1, \ldots n. \tag{3.19}$$

Dabei ist $\langle \cdot, \cdot \rangle$ das Innenprodunkt in $L^2(P)$, d.h. $\langle X_i, X_j \rangle = \mathrm{Cov}(X_i, X_j)$. (Man beachte $\mathrm{E}X_t \equiv 0$.) Somit ist die Orthogonalität in (3.19) gleichbedeutend mit Unkorreliertheit.

Die Aussage in (3.19) wiederum ist gleichbedeutend zu

$$\langle \sum_{i=1}^{n} \phi_{ni}\ X_{n+1-i}, X_{n+1-j} \rangle = \langle X_{n+1}, X_{n+1-j} \rangle,\ j = 1, \ldots, n. \tag{3.20}$$

Wegen der Linearität des Innenproduktes kann (3.20) geschrieben werden als

$$\begin{aligned} & \sum_{i=1}^{n} \phi_{ni}\ \langle X_{n+1-i}, X_{n+1-j} \rangle = \langle X_{n+1}, X_{n+1-j} \rangle,\ j = 1, \ldots, n, \\ \iff & \sum_{i=1}^{n} \phi_{ni}\ \gamma(i-j) = \gamma(j),\ j = 1, \ldots, n, \\ \iff & \Gamma_n \phi_n = \gamma_n \end{aligned} \tag{3.21}$$

$$\text{mit } \Gamma_n = (\gamma(i-j))_{i,j=1}^{n}\ ,\ \gamma_n = \begin{pmatrix} \gamma(1) \\ \vdots \\ \gamma(n) \end{pmatrix} \text{ und } \phi_n = \begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{nn} \end{pmatrix}. \tag{3.22}$$

Die Gleichungen (3.21) heißen *Yule-Walker-Gleichungen*. Das folgende Lemma gewährleistet schon unter sehr schwachen Bedingungen die Invertierbarkeit der Kovarianzmatrix Γ_n und damit die Existenz einer eindeutigen Lösung des linearen Gleichungssystems (3.21).

3.3.1 Lemma: *Gilt $\gamma(0) > 0$ und $\gamma(h) \to 0$ $(h \to \infty)$, dann ist die Kovarianzmatrix Γ_n von $(X_1, \ldots, X_n)^\top$ regulär für jedes $n \in \mathbb{N}$.*

Beweis: Brockwell / Davis (1991).

Damit ergibt sich sofort:

3.3.2 Satz: *Unter den Voraussetzungen von Lemma 3.3.1 ist die beste lineare Vorhersage $\hat{X}_{n+1}$ von X_{n+1} aufgrund $X_1, \ldots, X_n$ gegeben durch*

$$\hat{X}_{n+1} = \sum_{i=1}^{n} \phi_{ni}\ X_{n+1-i},\ n = 1, 2, \ldots$$

mit $\phi_n = (\phi_{n1}, \ldots, \phi_{nn})^\top = \Gamma_n^{-1}\gamma_n$, Γ_n und γ_n wie in (3.22). Weiter gilt

$$v_n := E(\hat{X}_{n+1} - X_{n+1})^2 = \gamma(0) - \gamma_n^\top \Gamma_n^{-1}\gamma_n. \tag{3.23}$$

Beweis: Es bleibt nur (3.23) zu zeigen:

$$\begin{aligned}
\mathrm{E}(\hat{X}_{n+1} - X_{n+1})^2 &= \langle \hat{X}_{n+1} - X_{n+1}, \hat{X}_{n+1} - X_{n+1}\rangle \\
&= \underbrace{\langle \hat{X}_{n+1} - X_{n+1}, \hat{X}_{n+1}\rangle}_{=0,\ da\ \hat{X}_{n+1}\in\mathcal{H}_{1,n}} - \langle \hat{X}_{n+1} - X_{n+1}, X_{n+1}\rangle \\
&= \langle X_{n+1}, X_{n+1}\rangle - \sum_{i=1}^{n} \phi_{ni}\ \langle X_{n+1-i}, X_{n+1}\rangle \\
&= \gamma(0) - \phi_n^\top \gamma_n \qquad (3.24) \\
&= \gamma(0) - \gamma_n^\top \Gamma_n^{-1}\gamma_n.
\end{aligned}$$

□

Für große n empfiehlt sich zur Berechnung von $\phi_n = (\phi_{n1}, \ldots, \phi_{nn})^\top$ ein rekursiver Algorithmus, der Durbin-Levinson-Algorithmus.

3.3.3 Satz: (Durbin-Levinson-Algorithmus) *Es sei (X_t) ein stationärer Prozess mit $EX_t \equiv 0$ und Kovarianzfunktion γ so, dass $\gamma(0) > 0$ und $\gamma(h) \to 0$ für $h \to \infty$. Dann lassen sich die mittlere quadratische Abweichung v_n und die Koeffizienten ϕ_{ni} aus Satz 3.3.2 rekursiv berechnen gemäß*

$$\begin{aligned}
\phi_{11} &= \frac{\gamma(1)}{\gamma(0)},\ v_0 = \gamma(0), \\
\phi_{nn} &= v_{n-1}^{-1} \cdot \left(\gamma(n) - \sum_{i=1}^{n-1} \phi_{n-1,i}\ \gamma(n-i)\right), \\
\begin{pmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{pmatrix} &= \begin{pmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{pmatrix} - \phi_{nn} \begin{pmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{pmatrix},
\end{aligned}$$

sowie

$$v_n = v_{n-1}(1 - \phi_{nn}^2).$$

Beweis: Brockwell / Davis (1991).

Ersetzt man die Basis $\{X_1, \ldots, X_n\}$ von $\mathcal{H}_{1,n}$ durch die Orthogonalbasis $\{X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n\}$ ($\hat{X}_1 := 0$), so wird deutlich, dass eine Darstellung der besten linearen Vorhersage in der Form

$$\hat{X}_{n+1} = \sum_{i=1}^{n} \theta_{ni}\ (X_{n+1-i} - \hat{X}_{n+1-i})\ (n \geq 1) \tag{3.25}$$

existiert mit Koeffizienten $\theta_{n1}, \ldots, \theta_{nn} \in \mathbb{R},\ n \in \mathbb{N}$. Die Zufallsvariablen $X_{n+1-i} - \hat{X}_{n+1-i}$, $i = 1, \ldots, n$, heißen *Innovationen*. Auch zur Berechnung der Koeffizienten θ_{ni} existiert ein rekursiver Algorithmus, der *Innovationsalgorithmus*.

3.3.4 Satz: (Innovations-Algorithmus) *Es sei (X_t) ein stochastischer Prozess mit $EX_t \equiv 0$ und Autokovarianzfunktion $\kappa(i,j) := EX_iX_j$. Die Matrix $(\kappa(i,j))_{i,j=1}^n$ sei regulär für alle $n \in \mathbb{N}$. Dann lassen sich die Koeffizienten θ_{ni} in der Darstellung (3.25) und die mittlere quadratische Abweichung $v_n = E(\hat{X}_{n+1} - X_{n+1})^2$ rekursiv berechnen, wobei*

$$\begin{aligned} v_0 &= \kappa(1,1), \\ \theta_{n,n-k} &= v_k^{-1} \cdot \left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\ v_j \right),\ k = 0,1,\ldots,n-1, \end{aligned}$$

und

$$v_n = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2\ v_j.$$

Beweis: Brockwell / Davis (1991).

Der Innovationsalgorithmus bietet gegenüber dem Durbin-Levinson-Algorithmus die Vorteile, dass er in seiner allgemeinen Form (Satz 3.3.4) auf Stationarität verzichtet und sich im Falle eines *ARMA*-Prozesses rechentechnisch erheblich vereinfacht. Für einen *ARMA*-Prozess gilt nämlich (Beweis s. Brockwell / Davis (1991))

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n = 1,\ldots,m-1, \\ \phi_1 X_n + \cdots + \phi_p X_{n+1-p} + \sum_{j=1}^q \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m \end{cases} \tag{3.26}$$

mit $m := \max\{p,q\}$.

Die Parameter $\theta_{nj},\ j = 1,\ldots,n$, erhält man, indem man den Innovationsalgorithmus auf die Zeitreihe (W_t) mit

$$W_t = \begin{cases} \sigma^{-1}X_t, & t = 1,\ldots m, \\ \sigma^{-1}(X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}), & t > m, \end{cases}$$

gemäß Satz 3.3.4 anwendet. Dabei ist

$$\kappa(i,j) = \begin{cases} \sigma^{-2}\gamma_X(i-j), & 1 \leq i,j \leq m, \\ \sigma^{-2}\left(\gamma_X(i-j) - \sum_{r=1}^p \phi_r \gamma_X(r-|i-j|)\right), & \min(i,j) \leq m < \max(i,j) \leq 2m, \\ \sum_{r=0}^q \theta_r\theta_{r+|i-j|}, & \min(i,j) > m, \\ 0, & \text{sonst.} \end{cases}$$

Man beachte, dass in diesem Zusammenhang der Innovationsalgorithmus nicht $v_n = \mathrm{E}(X_{n+1} - \hat{X}_{n+1})^2$, sondern die mittlere quadratische Abweichung

$$w_n = \mathrm{E}(W_{n+1} - \hat{W}_{n+1})^2 = \frac{v_n}{\sigma^2} \tag{3.27}$$

berechnet.

3.4 Parameterschätzung in *ARMA*-Modellen

3.4.1 Die partielle Autokorrelationsfunktion

Soll in der Praxis ein *ARMA*-Modell an eine beobachtete Zeitreihe angepasst werden, so gilt es, die Parameter p und q sowie die Koeffizientenvektoren $\phi = (\phi_1,\ldots,\phi_p)^\top$, $\theta = (\theta_1,\ldots,\theta_q)^\top$ und die White Noise Varianz σ^2 zu schätzen. Kann man speziell von einem $MA(q)$-Prozess ausgehen, so bietet die empirische Autokorrelationsfunktion einen Anhaltspunkt für die Schätzung von q (Beispiel 3.2.7 sowie Abbildung 3.9). Zur Bestimmung von p bei Anpassung eines $AR(p)$-Prozesses ist die empirische Autokorrelationsfunktion hingegen nicht sehr hilfreich. Denn wie bereits im Anschluss an Beispiel 3.2.7 angesprochen wurde, verschwindet die Autokorrelationsfunktion eines autoregressiven Prozesses **nicht** für $h \to \infty$. Wir veranschaulichen dieses Phänomen an einem $AR(1)$-Prozess (X_t):

Für ein beliebiges $t \in \mathbb{Z}$ ist $X_t = \phi_1 X_{t-1} + Z_t$ korreliert mit X_{t-1}. Wegen $X_{t-1} = \phi_1 X_{t-2} + Z_{t-1}$ gilt wiederum $X_t = \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t$. Somit ergeben sich "indirekte" Korrelationen zwischen X_t und X_{t-2}, u.s.w., die sich in der Autokorrelationsfunktion wiederspiegeln. Man greift daher zur Bestimmung des Parameters p eines autoregressiven Prozesses auf die *partielle Autokorrelationsfunktion (partial ACF)* zurück, die so konstruiert ist, dass sie "indirekte" Korrelationen nicht aufzeigt.

3.4.1 Definition: Für einen stationären Prozess (X_t) mit $\mathrm{E}X_t \equiv 0$ wird durch

$$\begin{aligned}\alpha(1) &:= \mathrm{Corr}(X_2, X_1) = \rho(1),\\ \alpha(h) &:= \mathrm{Corr}(X_{h+1} - \mathrm{pr}_{\mathcal{H}_{2,h}} X_{h+1}, X_1 - \mathrm{pr}_{\mathcal{H}_{2,h}} X_1),\ h \geq 1,\end{aligned}$$

die *partielle Autokorrelationsfunktion* $\alpha(\cdot)$ definiert. Dabei ist $\mathcal{H}_{2,h} := \overline{\mathrm{sp}}\{X_2, \ldots, X_h\}$ der (abgeschlossene) Unterraum des $L^2(P)$ (vgl. Anhang B.3), der von den Zufallsvariablen $X_2, \ldots, X_h$ aufgespannt wird, also die Menge aller Linearkombinationen $a_2X_2 + \cdots + a_hX_h,\ a_2, \ldots, a_h \in \mathbb{R}$. $\mathrm{pr}_{\mathcal{H}_{2,h}}$ bezeichnet die Orthogonalprojektion auf den Unterraum $\mathcal{H}_{2,h}$ (OGP vgl. Anhang B.3).

3.4.2 Bemerkung: Im Falle eines linearen Regressionsmodells $Y_n = X_n\beta + \epsilon_n$ (vgl. Abschnitt 5.1) ist die Orthogonalprojektion von Y_n auf den Bildraum der Designmatrix X_n der beste lineare erwartungstreue Schätzer für $X_n\beta$. Durch die Orthogonalprojektion von Y_n auf das **orthogonale Komplement** des Bildraumes von X_n erhält man den Vektor der Residuen, s. Abschnitt 5.1.

In Definition 3.4.1 ist die Orthogonalprojektion eines $X \in L^2(P)$ auf den Raum $\mathcal{H}_{2,h}$ die beste lineare Vorhersage von X bei gegebenen $X_2, \ldots, X_h$. Die Orthogonalprojektion von X auf das **orthogonale Komplement** von $\mathcal{H}_{2,h}$ ist $X - \mathrm{pr}_{\mathcal{H}_{2,h}} X$. Bei der besten linearen Vorhersage von X spielt also $X - \mathrm{pr}_{\mathcal{H}_{2,h}} X$ die Rolle des Residuenvektors bei linearen Modellen. Die partielle ACF an der Stelle h entspricht somit dem Korrelationskoeffizienten zwischen dem "Residuum" von X_1 und dem "Residuum" von X_{h+1}.

Durch diesen Übergang von den Zufallsvariablen auf die "Residuen" werden die "indirekten" Korrelationen eines $AR(p)$-Prozesses bei der Berechnung des Korrelationskoeffizienten außer acht gelassen. Bei einem $AR(p)$-Prozess verschwindet daher die partielle ACF für $h > p$: Es sei (X_t) der nicht vorgreifende $AR(p)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t\ ,\ (Z_t) \sim WN(0, \sigma^2),$$

mit $\mathrm{E}X_t \equiv 0$, also

$$X_{h+1} = \phi_1 X_h + \cdots + \phi_p X_{h+1-p} + Z_{h+1}\ . \tag{3.28}$$

Da (X_t) nicht vorgreifend ist, ist die beste lineare Vorhersage von X_{h+1} in Abhängigkeit von $X_2, \ldots, X_h$ für $h > p$ gegeben durch

$$\begin{aligned}\mathrm{pr}_{\mathcal{H}_{2,h}} X_{h+1} &= \phi_1 X_h + \cdots + \phi_p X_{h+1-p}\\ &= \sum_{j=1}^{p} \phi_j X_{h+1-j}\ .\end{aligned}$$

Damit ist für $h > p$

$$\begin{aligned}\alpha(h) &= \mathrm{Corr}(X_{h+1} - \sum_{j=1}^{p} \phi_j X_{h+1-j}, X_1 - \mathrm{pr}_{\mathcal{H}_{2,h}} X_1)\\ &= \mathrm{Corr}(Z_{h+1}, X_1 - \mathrm{pr}_{\mathcal{H}_{2,h}} X_1)\\ &= 0.\end{aligned}$$

Für $h \leq p$ gilt der folgende Satz:

3.4.3 Satz: *Es sei (X_t) ein stationärer Prozess mit $E(X_t) \equiv 0$ und Kovarianzfunktion γ so, dass $\gamma(h) \to 0$ für $h \to \infty$. Weiter sei $\mathcal{H}_{1,h} := \overline{sp}\{X_1, \ldots, X_h\}$ der (abgeschlossene) Unterraum des $L^2(P)$, der von den Zufallsvariablen $X_1, \ldots, X_h$ aufgespannt wird. Dann gilt für die partielle Autokorrelationsfunktion $\alpha(\cdot)$*

$$\alpha(h) = \phi_{hh},\ h \geq 1,$$

wobei ϕ_{hh} eindeutig bestimmt ist durch die folgende Darstellung der Orthogonalprojektion von X_{h+1} auf den abgeschlossenen Unterraum $\mathcal{H}_{1,h}$:

$$\begin{aligned}pr_{\mathcal{H}_{1,h}} X_{h+1} &= \phi_{h1} X_h + \cdots + \phi_{hh} X_1\\ &= \sum_{i=1}^{h} \phi_{hi} X_{h+1-i}\ .\end{aligned}$$

Beweis: Brockwell / Davis (1991).

Nach Satz 3.4.3 entspricht die partielle Autokorrelation $\alpha(h)$ also dem Koeffizienten ϕ_{hh} von X_1 in der Darstellung der besten linearen Vorhersage von X_{h+1} durch $X_1, \ldots, X_h$. Da diese beste lineare Vorhersage die Orthogonalprojektion von X_{h+1} auf $\mathcal{H}_{1,h}$ ist, ist $X_{h+1} - \mathrm{pr}_{\mathcal{H}_{1,h}} X_{h+1}$ orthogonal zu jedem Vektor, der $\mathcal{H}_{1,h}$ aufspannt, also zu X_j, $j = 1, \ldots, h$. So ergeben sich aus

$$\begin{aligned} & \langle X_{h+1} - \mathrm{pr}_{\mathcal{H}_{1,h}} X_{h+1}, X_j \rangle &&= 0, \ j = 1, \ldots, h, \\ \iff \quad & \textstyle\sum_{i=1}^{h} \phi_{hi} \langle X_{h+1-i}, X_j \rangle &&= \langle X_{h+1}, X_j \rangle, \ j = 1, \ldots, h, \end{aligned}$$

mit

$$\langle X_i, X_j \rangle = \mathrm{E} X_i X_j \overset{EX_t=0}{=} \mathrm{Cov}(X_i, X_j) = \gamma(i-j)$$

die Yule-Walker-Gleichungen $\Gamma_h \phi_h = \gamma_h$ (vgl. (3.21)). Nach Division durch $\gamma(0)$ erhält man das lineare Gleichungssystem

$$\begin{pmatrix} \rho(0) & \rho(1) & \rho(2) & \ldots & \rho(h-1) \\ \rho(1) & \rho(0) & \rho(1) & \ldots & \rho(h-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \ldots & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_{h1} \\ \phi_{h2} \\ \vdots \\ \phi_{hh} \end{pmatrix} = \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix}. \tag{3.29}$$

Schätzt man in (3.29) die Autokorrelationen empirisch, so erhält man durch Lösen des entsprechenden Gleichungssystems die geschätzten Koeffizienten $\hat{\phi}_{h1}, \ldots, \hat{\phi}_{hh}$. (Zur Lösbarkeit des LGS s. Lemma 3.4.6.) Die Funktion $\hat{\alpha}(\cdot)$ mit $\hat{\alpha}(h) := \hat{\phi}_{hh}, \ h = 1, \ldots, n$, heißt die *empirische partielle Autokorrelationsfunktion* der Zeitreihe (X_t). Abbildung 3.11 zeigt die empirische partielle Autokorrelationsfunktion des autoregressiven Prozesses aus Abbildung 3.7.

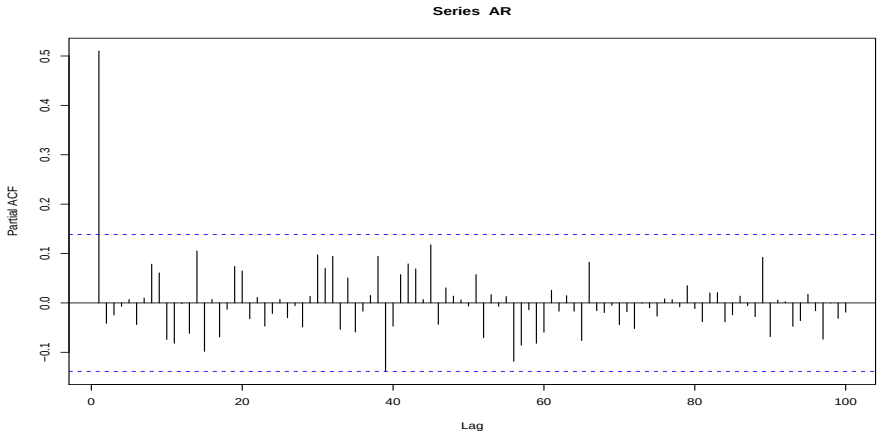

Abbildung 3.11: Empirische partielle ACF eines $AR(1)$-Prozesses

3.4.4 Bemerkung: Liegt kein $AR(p)$-Prozesses vor, so verschwindet die partielle ACF für große Time-Lags i.A. **nicht**, wie das folgende Beispiel zeigt:

Es sei (X_t) der $MA(1)$-Prozess

$$X_t = Z_t + \theta_1 Z_{t-1}, \ |\theta_1| < 1, \ (Z_t) \sim WN(0, \sigma^2).$$

Dann gilt (vgl. Brockwell / Davis (1991))

$$\begin{aligned} \alpha(1) &= \rho(1) = \frac{\theta_1}{1+\theta_1^2} \\ \alpha(h) &= \frac{-(-\theta_1)^h (1-\theta_1^2)}{1-\theta_1^{2(h+1)}}, \ h \geq 2. \end{aligned}$$

Abbildung 3.12 zeigt die empirische partielle Autokorrelationsfunktion des $MA(1)$-Prozesses aus Abbildung 3.6.

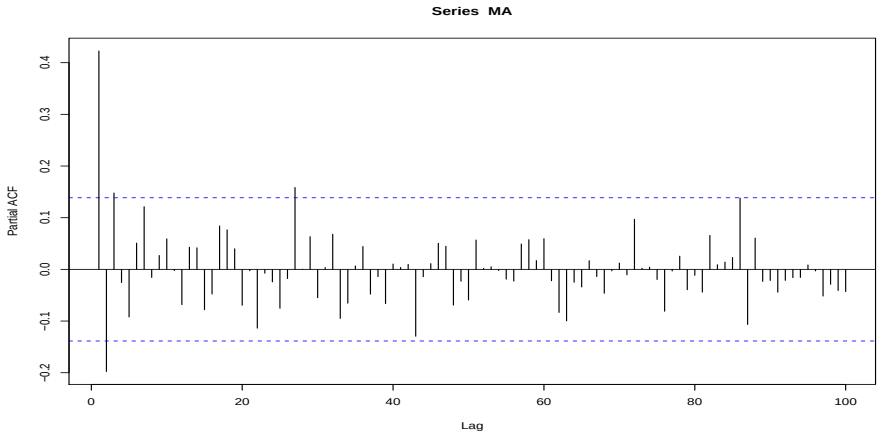

Abbildung 3.12: Empirische partielle ACF eines $MA(1)$-Prozesses

3.4.2 Schätzer für die Parameter p und q in $ARMA$-Modellen

Wie bereits in Abschnitt 3.2 erwähnt wurde, lässt sich im Falle eines $MA(q)$-Prozesses ein Schätzer für den Parameter q aus der empirischen Autokorrelationsfunktion ablesen. Man vergleiche hierzu Beispiel 3.2.7 und Abbildung 3.9. Im Falle eines $AR(p)$-Prozesses kann p mit Hilfe der empirischen partiellen ACF bestimmt werden.

Bei Modellannahme eines $ARMA(p,q)$-Prozesses hingegen ist die Bestimmung der Parameter p und q aufwändiger als bei Annahme eines Moving Average oder eines autoregressiven Prozesses. Zunächst kann man anhand der empirischen Autokorrelationsfunktion und der empirischen partiellen ACF p und q grob schätzen (s. Abschnitt 3.4.1). Anschließend geht man so vor, dass für mehrere in Frage kommende Werte von p und q ein $ARMA(p,q)$-Prozess angepasst wird, d.h. die Parametervektoren $\phi = (\phi_1, \ldots, \phi_p)^\top$, $\theta = (\theta_1, \ldots, \theta_q)^\top$ und die White Noise Varianz σ^2 geschätzt werden. Schließlich werden die angepassten Modelle auf die Güte der Anpassung hin überprüft.

Es ist zu beachten, dass ein Modell mit höheren Werten für p und q den Anschein haben kann, die gegebenen Daten besser zu modellieren als ein Modell mit kleinerem p bzw. q. Doch wie bei der Anpassung eines linearen Modells (vgl. Abschnitt 5.1) an einen gegebenen Datensatz muss man sich auch hier vorsehen, das Modell nicht zu überparametrisieren (*Overfitting*). Ein Beispiel aus der Theorie linearer Modelle soll das Problem des Overfittings veranschaulichen:

3.4.5 Beispiel: Gegeben seien 100 Beobachtungen des linearen Modells $Y_t = a + bt + \epsilon_t,\ (\epsilon_t) \overset{iid}{\sim} \mathcal{N}(0,1)$. Wird an diese 100 Beobachtungen ein Polynom vom Grade 99 angepasst, so wird das geschätzte Modell mit den Beobachtungen identisch sein, der Fit scheint somit perfekt. Doch das angepasste Modell enthält zuviele zufällige Einflüsse, eine andere Realisierung wird vollständig andere Ergebnisse liefern.

Um einerseits eine Überparametrisierung zu verhindern und andererseits eine gute Anpassung des Modells zu ermöglichen, wurden auf dem Gebiet der Zeitreihenanalyse Kriterien entwickelt, mittels derer angepasste $ARMA$-Modelle auf ihre Eignung hin verglichen werden können. Zu erwähnen sind in diesem Zusammenhang beispielhaft die Kriterien AIC und AICC von Akaike. Die Akaike-Kriterien ordnen jedem $ARMA$-Modell eine reelle Zahl AIC bzw. $AICC$ zu, die einerseits umso kleiner wird, je besser die Anpassung des Modells ist und andererseits umso größer, je mehr Parameter eingeführt werden. Die optimale Wahl von $\hat{p},\ \hat{q},\ \hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)^\top$ und $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)^\top$ ist somit jene, die den Wert des Kriteriums minimiert. Das Akaike-Kriterium AICC z.B. ist gegben durch

$$AICC(\phi, \theta) := -2 \ln L(\phi, \theta, S_n(\phi, \theta)) + \frac{2n(p+q+1)}{n-p-q-2}.$$

Dabei ist $L(\phi, \theta, S_n(\phi, \theta))$ die *Likelihood-Funktion* aus Abschnitt 3.4.3. Man beachte, dass bei vorgegebenem p und q die Minimierung des Akaike-Kriteriums einer Maximierung der Likelihood-Funktion entspricht.

3.4.3 Maximum Likelihood Schätzer für ϕ, θ und σ^2

Sind p und q fest vorgegeben, so lassen sich optimale Schätzer für die Parametervektoren $\phi = (\phi_1, \ldots, \phi_p)^\top$, $\theta = (\theta_1, \ldots, \theta_q)^\top$ und die White Noise Varianz σ^2 mit der Maximum Likelihood (ML) Methode (s. Anhang B) bestimmen.

Ist (X_t) ein nicht vorgreifender $ARMA(p,q)$-Prozess, so ist unter Normalverteilungsannahme die *Likelihood-Funktion* des Beobachtungsvektors $x = (x_1, \ldots, x_n)^\top$ in Abhängigkeit von den Parametern ϕ, θ und σ^2 gegeben durch (s. Brockwell / Davis (1991))

$$L(\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-n/2}(w_0 \cdots w_{n-1})^{-1/2} \cdot \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(x_j - \hat{X}_j)^2}{w_{j-1}}\right\}.$$

Dabei ist $\hat{X}_j = \hat{X}_j(\phi, \theta)$ die Vorhersage für X_j aus dem Innovationsalgorithmus bei gegebenen Beobachtungen $x_1, \ldots, x_{j-1}$ $(j = 2, \ldots, n,\ \hat{X}_1 := 0)$ und w_{j-1} wie in (3.27) definiert, $j = 1, \ldots, n$.

Es gilt also, $L(\phi, \theta, \sigma^2)$ oder (gleichbedeutend)

$$\ln L(\phi, \theta, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(x_j - \hat{X}_j)^2}{w_{j-1}} - \frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{j=1}^{n}\ln w_{j-1} \tag{3.30}$$

in ϕ, θ und σ^2 zu maximieren. Wir führen dies in zwei Schritten durch und leiten $\ln L$ zunächst nach σ^2 ab:

$$\frac{\partial \ln L(\phi, \theta, \sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^2}\left(\frac{1}{\sigma^2}\sum_{j=1}^{n}\frac{(x_j - \hat{X}_j)^2}{w_{j-1}} - n\right).$$

Dieser Ausdruck verschwindet für

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{(x_j - \hat{X}_j)^2}{w_{j-1}} =: S_n(\phi, \theta). \tag{3.31}$$

Da die zweite Ableitung an der Stelle $S_n(\phi, \theta)$ negativ ist, ist $S_n(\hat{\phi}_{ML}, \hat{\theta}_{ML})$ der Maximum Likelihood Schätzer für σ^2. Dabei bezeichnet $\hat{\phi}_{ML}$ und $\hat{\theta}_{ML}$ die ML-Schätzer für ϕ bzw. θ, welche den Ausdruck (ersetze in (3.30) σ^2 durch $S_n(\phi, \theta)$)

$$l(\phi, \theta) := \ln S_n(\phi, \theta) + \frac{1}{n}\sum_{j=1}^{n}\ln w_{j-1}(\phi, \theta) \tag{3.32}$$

in ϕ und θ minimieren. Diese Minimierung erfolgt i.d.R. über einen nichtlinearen Optimierungsalgorithmus, für den $\hat{X}_j = \hat{X}_j(\phi, \theta)$ und $w_{j-1} = w_{j-1}(\phi, \theta)$ mittels des Innovationsalgorithmus berechnet werden. Hierzu werden initiale Schätzer für ϕ, θ und σ^2 benötigt. Zur Bestimmung von Initialschätzern, die den optimalen Werten schon relativ nahe sind, wird auf die Abschnitte 3.4.5 bis 3.4.7 verwiesen.

3.4.4 Least Squares Schätzer für ϕ, θ und σ^2

Nach Anpassung eines $ARMA(p,q)$-Modells an eine stationäre Zeitreihe werden (mit den Bezeichnungen aus Abschnitt 3.4.3) durch

$$R_j := \frac{x_j - \hat{X}_j(\phi, \theta)}{\sqrt{w_{j-1}(\phi, \theta)}} \tag{3.33}$$

Residuen nach Anpassung des Modells definiert. Es können also auch diejenigen Schätzer für ϕ und θ in einem gewissen Sinne als optimal betrachtet werden, die die Summe der quadrierten Residuen

$$nS_n(\phi, \theta) = \sum_{j=1}^{n}\frac{(x_j - \hat{X}_j)^2}{w_{j-1}}$$

in ϕ und θ minimieren. Die so erhaltenen Schätzer $\hat{\phi}_{LS}$ und $\hat{\theta}_{LS}$ für ϕ bzw. θ heißen *Kleinste-Quadrate-Schätzer* (*Least Squares (LS) Estimator*) für ϕ bzw. θ. Der LS-Schätzer für σ^2 ist dann gegeben durch (siehe Brockwell / Davis (1991))

$$\widehat{\sigma^2}_{LS} = \frac{n}{n-p-q} S_n(\hat{\phi}_{LS}, \hat{\theta}_{LS}).$$

Man beachte, dass auch für dieses Optimierungsverfahren initiale Schätzer für ϕ, θ und σ^2 benötigt werden, um die Vorhersagen (3.26) und die mittlere quadratische Abweichnung (3.27) berechnen zu können.

3.4.5 Initiale Parameterschätzung in $AR(p)$-Modellen

Es sei p fest vorgegeben und (X_t) der nicht vorgreifende $AR(p)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t, \ (Z_t) \sim WN(0, \sigma^2), \tag{3.34}$$

mit $\mathrm{E}X_t \equiv 0$ und unbekannten Koeffizienten $\phi_1, \ldots, \phi_p$, sowie White Noise Varianz σ^2.

Da (X_t) nicht vorgreifend ist, existiert eine Darstellung

$$X_t = \sum_{i=0}^{\infty} \psi_i Z_{t-i} \ ,$$

bzw. für $j = 0, 1, \ldots, p$,

$$X_{t-j} = \sum_{i=0}^{\infty} \psi_i Z_{t-j-i} = \sum_{i=j}^{\infty} \psi_{i-j} Z_{t-i} \ . \tag{3.35}$$

Bildet man für $j = 0, 1, \ldots, p$ in (3.34) das Skalarprodukt mit X_{t-j}, so erhält man unter Berücksichtigung von (3.35)

$$\begin{aligned} \langle X_t, X_{t-j}\rangle - \phi_1\langle X_{t-1}, X_{t-j}\rangle - \cdots - \phi_p\langle X_{t-p}, X_{t-j}\rangle &= \langle Z_t, X_{t-j}\rangle \\ &= \sum_{i=j}^{\infty} \psi_{i-j}\langle Z_t, Z_{t-i}\rangle. \end{aligned} \tag{3.36}$$

Mit

$$\langle Z_t, Z_{t-i}\rangle = \begin{cases} \sigma^2, & i = 0, \\ 0, & \text{sonst}, \end{cases}$$

ergibt sich aus (3.36) mit (3.14)

$$\begin{aligned} \gamma(0) - \phi_1\gamma(1) - \cdots - \phi_p\gamma(p) &= \psi_0\sigma^2 = \sigma^2, \\ \phi_1\gamma(j-1) + \cdots + \phi_p\gamma(j-p) &= \gamma(j), \ j = 1, \ldots, p. \end{aligned}$$

Das entspricht den Yule-Walker-Gleichungen

$$\begin{aligned} \Gamma_p\phi_{(p)} &= \gamma_p \ , \\ \sigma^2 &= \gamma(0) - \phi_{(p)}^{\top}\gamma_p \ . \end{aligned} \tag{3.37}$$

Dabei ist Γ_p und γ_p wie in (3.22) definiert (setze $n = p$) und $\phi_{(p)} = (\phi_1, \ldots, \phi_p)^{\top}$. Durch den Übergang zu der empirischen ACF $\hat{\gamma}(\cdot)$ erhält man Schätzer $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ für $\phi_1, \ldots, \phi_p$ durch Lösen des linearen Gleichungssystems

$$\hat{\Gamma}_p\hat{\phi}_p = \hat{\gamma}_p \ (\hat{\Gamma}_p = (\hat{\gamma}(i-j))_{i,j=1}^{p} \text{ und } \hat{\gamma}_p = (\hat{\gamma}(1), \ldots, \hat{\gamma}(p))^{\top}) \tag{3.38}$$

und mit diesen einen Schätzer $\widehat{\sigma^2}$ für σ^2 mittels

$$\widehat{\sigma^2} = \hat{\gamma}(0) - \hat{\phi}_p^{\top}\hat{\gamma}_p \ .$$

Die Schätzer $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ und $\widehat{\sigma^2}$ heißen *Yule-Walker-Schätzer* für $\phi_1, \ldots, \phi_p$ bzw. σ^2. Das folgende Lemma gewährleistet die Lösbarkeit des linearen Gleichungssystems (3.38).

3.4.6 Lemma: *Es sei (X_t) ein stationärer Prozess mit $EX_t \equiv 0$ und $\hat{\gamma}(\cdot)$ die empirische Autokorrelationsfunktion von (X_t). Ist $\hat{\gamma}(0) > 0$, dann ist $\hat{\Gamma}_k$ regulär für alle $k \in \mathbb{N}$.*

Beweis: Brockwell / Davis (1991).

Die Analogie von (3.38) zu (3.21) legt es nahe, die Yule-Walker-Schätzer $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ und $\widehat{\sigma^2}$ mit Hilfe des Durbin-Levinson-Algorithmus zu berechnen. (Man ersetze hierzu $\gamma(\cdot)$ durch $\hat{\gamma}(\cdot)$ und v_j durch $\hat{v}_j,\ j = 1, \ldots, p$.) Es gilt dann (vgl. (3.37) mit (3.24)) $\widehat{\sigma^2} = \hat{v}_p$.

Schließlich gilt es noch zu überprüfen, ob $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ und $\widehat{\sigma^2}$ "sinnvolle" Schätzer für $\phi_1, \ldots, \phi_p$ bzw. σ^2 sind. Eine Minimalanforderung an einen "sinnvollen" Schätzer $\hat{\beta}_n$ für β ist, neben der Erwartungstreue, die stochastische Konvergenz gegen den zu schätzenden Parameter β, also

$$\hat{\beta}_n \xrightarrow{P} \beta. \tag{3.39}$$

Ein Schätzer $\hat{\beta}$, der (3.39) erfüllt, heißt *konsistent* für β (vgl. Anhang B).

3.4.7 Lemma: *Die Yule-Walker-Schätzer $\hat{\phi}_{p1}, \ldots, \hat{\phi}_{pp}$ und $\widehat{\sigma^2}$ sind konsistente Schätzer für $\phi_1, \ldots, \phi_p$ bzw. σ^2, d.h.*

$$\hat{\phi}_{pj} \xrightarrow{P} \phi_j,\ (j = 1, \ldots, p) \text{ und } \widehat{\sigma^2} \xrightarrow{P} \sigma^2.$$

Beweis: Brockwell / Davis (1991).

3.4.8 Bemerkung: Im Prinzip kann die empirische partielle Autokorrelation $\hat{\alpha}(m) = \hat{\phi}_{mm}$ als Nebenprodukt bei der Anpassung eines $AR(m)$-Prozesses, $m \in \mathbb{N}$, betrachtet werden, wobei m sukzessive erhöht wird. Auch mit diesem Ansatz liegt es nahe, p durch jenen Wert $\tilde{p}$ zu schätzen, für den $\hat{\alpha}(m) = \hat{\phi}_{m,m}$ "klein" ist für alle $m > \tilde{p}$, d.h. innerhalb geeigneter Konfidenzschranken liegt. Diese Konfidenzschranken werden von den gängigen Statistikpaketen bei der Schätzung der partiellen ACF mit ausgegeben. Für die theoretischen Grundlagen zur Berechnung der Konfidenzschranken wird auf Brockwell / Davis (1991) verwiesen.

3.4.6 Initiale Parameterschätzung in $MA(q)$-Modellen

Ist der Parameter q fest vorgegeben, so gilt es noch, die Koeffizienten $\theta_1, \ldots, \theta_q$ und die White Noise Varianz σ^2 zu schätzen. Setzt man

$$\mathrm{E}X_t \equiv 0 \tag{3.40}$$

voraus, so leistet der Innovationsalgorithmus aus Abschnitt 3.3 auch hier gute Dienste. Denn wie die Zufallsvariablen $Z_t,\ t \in \mathbb{Z}$, des White Noise Prozesses, so besitzen auch die Innovationen $X_{n+1} - \hat{X}_{n+1},\ n = 0, 1, \ldots$, die Eigenschaft, in $L^2(P)$ zueinander orthogonal zu sein. Ist (X_t) der invertierbare $MA(q)$-Prozess

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2), \tag{3.41}$$

so lässt sich sogar zeigen (vgl. Brockwell / Davis (1991)), dass der $L^2(P)$-Abstand von Z_n zu der Innovation $X_n - \hat{X}_n$ für $n \to \infty$ verschwindet, also

$$\|(X_n - \hat{X}_n) - Z_n\| \overset{n\to\infty}{\longrightarrow} 0. \tag{3.42}$$

Es liegt also nahe, in der Darstellung (3.41) Z_{t-j} durch $X_{t-j} - \hat{X}_{t-j}$ zu ersetzen $(j = 0, \ldots, q)$. Auf diese Weise erhält man die Darstellung

$$\hat{X}_t = \sum_{i=0}^{q} \theta_i\ (X_{t-i} - \hat{X}_{t-i}),\ \theta_0 := 1,$$

in welcher die Parameter $\theta_i, \ldots, \theta_q$ mit Hilfe des Innovationsalgorithmus rekursiv berechnet werden können. Zum Zwecke der Parameterschätzung wähle man zunächst eine Folge $m(n),\ n \in \mathbb{N}$, in Abhängigkeit des Stichprobenumfangs n so, dass

$$m = m(n) \in \mathbb{N},\ m(n) < n,\ m(n) \to \infty\ (n \to \infty) \text{ und } m(n) = o(n^{1/3})\ (n \to \infty). \tag{3.43}$$

Dabei bedeutet die Schreibweise $m(n) = o(n^{1/3})$ $(n \to \infty)$, dass

$$\frac{m(n)}{n^{1/3}} \overset{n\to\infty}{\longrightarrow} 0.$$

Anschließend ersetze man in Satz 3.3.4 (Innovationsalgorithmus) die Kovarianzfunktion $\kappa(i,j) = \mathrm{E}(X_i X_j)$ durch die empirische Autokovarianzfunktion $\hat{\gamma}(i-j)$, die mittlere quadratische Abweichung v_n durch $\hat{v}_m$ und die Vorhersagekoeffizienten θ_{nj} durch Schätzer $\hat{\theta}_{mj}$, $j = 1, \ldots, m$. Einen Schätzer für σ^2 erhält man durch $\widehat{\sigma^2_m} := \hat{v}_m$.

Das Iterationsverfahren wird so lange fortgesetzt, bis sich die Schätzfolgen $(\hat{\theta}_{mj})_m$ stabilisieren, d.h. sich im Verhältnis zu $\left(\frac{1}{n}\sum_{i=1}^{j-1}\hat{\theta}^2_{mi}\right)^{1/2}$ nur noch geringfügig verändern $(j = 1, \ldots, q)$. Die Wahl der Folge $m(n)$, $n \in \mathbb{N}$, gemäß (3.43) sichert dabei die Konsistenz der Schätzer $\hat{\theta}_{mj}$, $j = 1, \ldots, p$, und $\widehat{\sigma^2}$.

3.4.7 Initiale Parameterschätzung in $ARMA(p,q)$-Modellen

Sind p und q fest vorgegeben, so wird ein $ARMA(p,q)$-Modell wie folgt angepasst:

Gegeben sei das nicht vorgreifende $ARMA(p,q)$-Modell

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim WN(0, \sigma^2).$$

Da (X_t) nicht vorgreifend ist, existiert eine Darstellung

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \tag{3.44}$$

mit (vgl. 3.14)

$$\begin{aligned} \psi_0 &= 1, \\ \psi_j &= \theta_j + \sum_{i=1}^{j} \phi_i \psi_{j-i},\ j = 1, 2, \ldots \end{aligned} \tag{3.45}$$

Dabei ist $\theta_j := 0$ für $j > q$ und $\psi_j := 0$ für $j > p$. Die Ausführungen in Abschnitt 3.4.6 bezüglich der Orthogonalität der Innovationen, insbesondere (3.42), legen auch hier eine Schätzung der Koeffizienten ψ_j in (3.44) mittels des Innovationsalgorithmus nahe. Sind $\hat{\psi}_{mj}$ die so erhaltenen Schätzer für ψ_j, $j = 1, 2, \ldots$, und ist $\hat{v}_m$ die mittlere quadratische Abweichung aus dem Innovationsalgorithmus (wie in Abschnitt 3.4.6 beschrieben), so gilt der folgende Satz:

3.4.9 Satz: *Es sei (X_t) der nicht vorgreifende und invertierbare $ARMA(p,q)$-Prozess*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ (Z_t) \sim IID(0, \sigma^2),$$

mit $E(Z_t^4) < \infty$ und $\psi(z)$ wie in (3.13), wobei $\psi_0 = 1$. Dann gilt für jedes $k \in \mathbb{N}$ und jede beliebige Folge $(m_n)_{n\geq 1}$, $m_n \in \mathbb{N}$ wie in (3.43),

$$\sqrt{n}(\hat{\psi}_{m1} - \psi_1, \ldots, \hat{\psi}_{mk} - \psi_k)^\top \overset{\mathcal{D}}{\longrightarrow} \mathcal{N}(0, A).$$

Dabei ist $A = (a_{ij})_{i,j=1}^k$ mit

$$a_{ij} = \sum_{r=1}^{\min(i,j)} \psi_{i-r}\psi_{j-r}.$$

Weiter gilt

$$\hat{v}_m \overset{P}{\longrightarrow} \sigma^2.$$

Beweis: Brockwell und Davis (1988).

Mit Lemma B.4.9 folgt aus Satz 3.4.9, dass für jede Folge (m_n) wie in (3.43)

$$\hat{\psi}_{mj} \xrightarrow{P} \psi_j,\ j = 1, 2 \ldots$$

D.h., $(\hat{\psi}_{mj})_{m_n}$ ist eine konsistente Schätzfolge für $\psi_j\ ,\ j = 1, 2 \ldots .$

Aus den Schätzern $\hat{\psi}_{mj}$ für ψ_j lassen sich wie folgt Schätzer für $\phi_1, \ldots, \phi_p$ und $\theta_1, \ldots, \theta_q$ berechnen:

Man ersetze in (3.45) ψ_j durch $\hat{\psi}_{mj}$ für $j = q+1, \ldots, q+p$ ($\phi_j = 0,\ j > p$, und $\theta_j = 0,\ j > q$) und löse die so erhaltenen Gleichungen

$$\begin{aligned}
\hat{\psi}_{m,q+1} &= \sum_{i=1}^{q+1} \phi_i \hat{\psi}_{m,q+1-i} \\
\hat{\psi}_{m,q+2} &= \sum_{i=1}^{q+2} \phi_i \hat{\psi}_{m,q+2-i} \\
&\vdots \\
\hat{\psi}_{m,q+p} &= \sum_{i=1}^{q+p} \phi_i \hat{\psi}_{m,q+p-i}
\end{aligned}$$

nach $\phi_1, \ldots, \phi_p$. Mit den so erhaltenen Schätzern $\hat{\phi}_{m1}, \ldots, \hat{\phi}_{mp}$ erhält man anschließend $\hat{\theta}_{mj}\ ,\ j = 1, \ldots, q$, durch

$$\hat{\theta}_{mj} = \hat{\psi}_{mj} - \sum_{i=1}^{\min(j,p)} \hat{\phi}_{mi} \hat{\psi}_{m,j-i}$$

(vgl. (3.45) für $j = 1, \ldots, q$) und schließlich

$$\widehat{\sigma^2} = \hat{v}_m.$$

Mit (m_n) wie in (3.43) sind die Schätzer $\hat{\phi}_{m1}, \ldots, \hat{\phi}_{mp}$, $\hat{\theta}_{m1}, \ldots, \hat{\theta}_{mq}$ und $\widehat{\sigma^2}$ konsistent für $\phi_1, \ldots, \phi_p$, $\theta_1, \ldots, \theta_q$ bzw. σ^2.

3.4.8 Überprüfen der Modellanpassung

Nach Schätzung von $p,\ q,\ \phi,\ \theta$ und σ^2 können die Vorhersagen $\hat{X}_j$ $(j = 2, \ldots, n)$ und die mittleren quadratischen Abweichungen w_{j-1} $(j = 1, \ldots, n)$ mit Hilfe des Innovationsalgorithmus berechnet werden (s. (3.26) bzw. (3.27)). Anschließend sollte das angepasste $ARMA(p, q)$-Modell auf seine Eignung hin geprüft und ggf. modifiziert werden. Üblicherweise werden zum Überprüfen der Modellanpassung die *Residuen*

$$R_j := \frac{x_j - \hat{X}_j(\phi, \theta)}{\sqrt{w_{j-1}(\phi, \theta)}} \tag{3.46}$$

einem Test für WN- bzw. IID-Prozesse unterzogen. Verwirft der Test die Hypothese *iid*-verteilter Residuen nicht, so kann man davon ausgehen, dass der angepasste Prozess die beobachtete Zeitreihe hinreichend gut modelliert. Diverse solche Tests sind in Abschnitt 6.2 zu finden. Für weitere Methoden des "Model Checking" wird auf die Bücher von Brockwell und Davis (1991 und 1996) verwiesen.

3.4.10 Beispiel: Eine geodätische GPS-Messreihe kann durch ein lineares Regressionsmodell (vgl. Abschnitt 5.1) beschrieben werden. Modelliert das gewählte Regressionsmodell die Daten hinreichend gut, so sollten die Residuen des Modells (s. Abschnitt 5.2.1) annähernd einen White Noise Prozess beschreiben. Abbildung 3.13 zeigt die Least Squares Residuen doppeltdifferenzierter (DD) Daten einer GPS-Messreihe aus der antarktischen Halbinsel. Der Fit des für die gemessenen Daten gewählten Modells kann nun durch die Betrachtung von empirischer ACF und empirischer partieller ACF der Residuen-Zeitreihe überprüft werden.

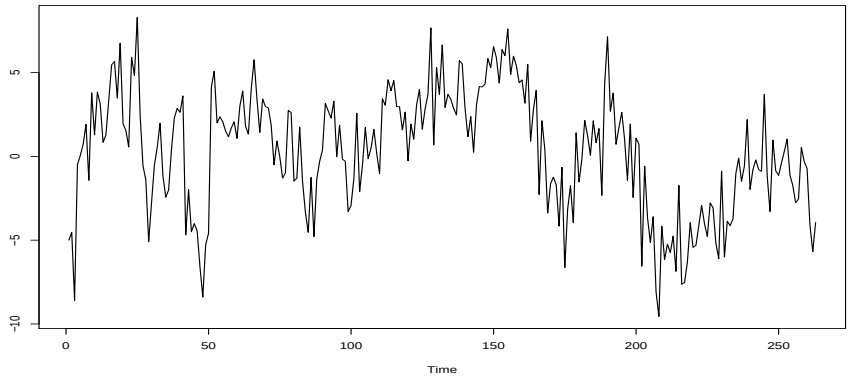

Abbildung 3.13: Least Squares Residuen einer antarktischen GPS-Messreihe

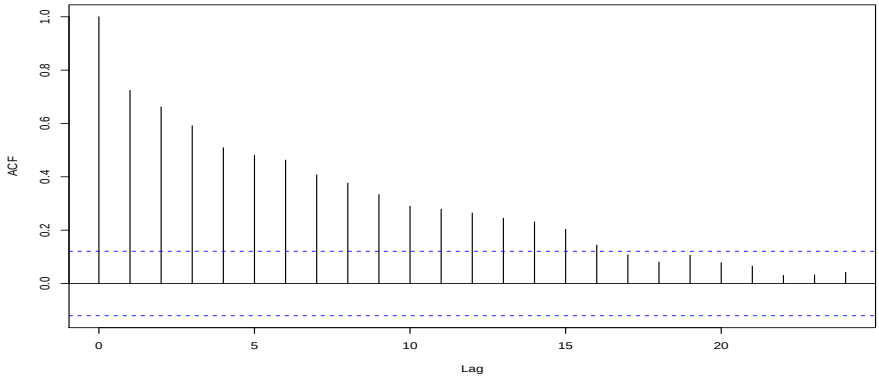

Abbildung 3.14: Empirische ACF der Residuenzeitreihe aus Abb. 3.13

Die Abbildungen 3.14 bzw. 3.15 zeigen die empirische Autokorrelationsfunktion und die empirische partielle Autokorrelationsfunktion der Zeitreihe aus Abbildung 3.13, jeweils mit dem 95%-Konfidenzbereich eines White Noise Prozesses.

Wie man in Abbildung 3.14 sieht, überschreiten die ersten 16 Time Lags den 95%-Konfidenzbereich und weisen somit klar auf zeitliche Korrelationen hin. Der 95%-Konfidenzbereich wird auch von der empirischen partiellen ACF überschritten, vgl. Abbildung 3.15.

Eine Minimierung des Akaike-Wertes ergibt die optimalen Parameterschätzer $\hat{p} = \hat{q} = 3$ und

$$\begin{array}{ll} \hat{\phi}_1 = -0.079 & \hat{\theta}_1 = \ 0.578 \\ \hat{\phi}_2 = -0.076 & \hat{\theta}_2 = \ 0.637 \\ \hat{\phi}_3 = \ 0.882 & \hat{\theta}_3 = -0.379. \end{array} \tag{3.47}$$

Nun können gemäß 3.46 Residuen nach Anpassung der Zeitreihe aus Abb. 3.13 an das ARMA(3,3)-Modell mit den Parametern (3.47) berechnet werden, sie sind in Abbildung 3.16 zu sehen. Die Abbildungen 3.17 und 3.18 zeigen die empirische ACF bzw. empirische partielle ACF der Residuen aus Abbildung 3.16. Alle Time Lags variieren innerhalb des 95%-Konfidenzbereiches eines White Noise Prozesses, was auf eine hinreichend gute Modellierung der Daten durch das ARMA(3,3)-Modell mit den Parametetern (3.47) schließen lässt.

In Abschnitt 6.2 werden die Residuen-Zeitreihen aus den Abbildungen 3.13 und 3.16 statistischen Tests zur Überprüfung der White Noise Annahme unterzogen.

Hinweis: Die anwendungsbezogene Fassung Brockwell / Davis (1996) enthält eine CD-Rom, auf der Software u.a. zur Parameterschätzung in *ARMA*-Modellen zu finden ist.

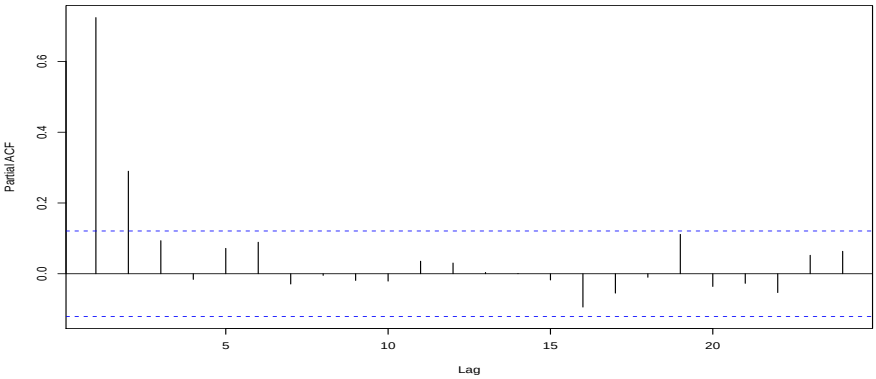

Abbildung 3.15: Empirische parteille ACF der Residuenzeitreihe aus Abb. 3.13

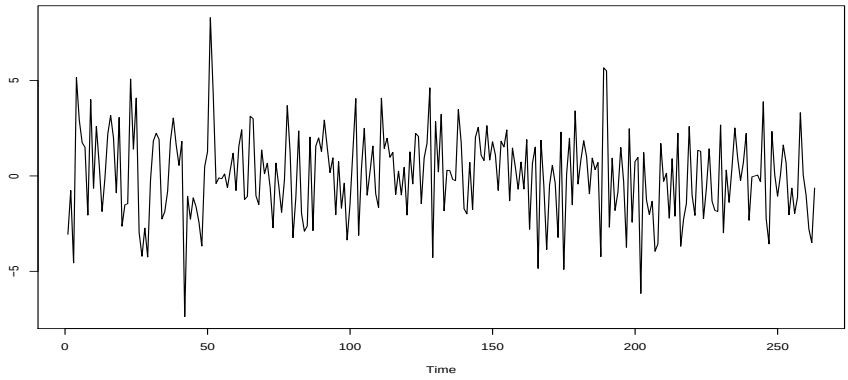

Abbildung 3.16: Gemäß 3.46 berechnete Residuen der Zeitreihe aus Abb. 3.13

3.5 Nichtstationäre Prozesse

3.5.1 Ein Beispiel aus der Geodäsie

Die Theorie in den Abschnitten 3.2 bis 3.4 ist auf der Annahme eines stationären Prozesses aufgebaut. In der Praxis jedoch enthalten viele Zeitreihen einen Trend oder einen Zyklus, oder sie sind aus anderen Gründen offensichtlich nicht stationär, z.B. indem sie eine schwankende Varianz aufweisen, wie Abbildung 3.19 anhand einer Residuen-Zeitreihe geodätischer GPS-Messdaten zeigt. (Es handelt sich um DD Beobachtungen aus der Gegend um Karlsruhe mit einer Basislinienlänge von ca. 15 km.)

In diesem Fall sollte zunächst versucht werden, wenigstens annähernd Stationarität zu erzeugen, z.B. durch Eliminierung von Trend bzw. Zyklus, vgl. Abschnitt 3.1. Die GPS-Zeitreihe aus Abbildung 3.19 wurde mit einer geeigneten Funktion gewichtet, um Homoskedastizität (Konstanz der Varianzfunktion) zu erzeugen. Das Thema Gewichtung wird in Abschnitt 7.2 ausführlich behandelt.

Abbildung 3.20 zeigt die gewichtete Zeitreihe aus Abbildung 3.19.

3.5.1 Bemerkung: Auch bei der GPS-Residuen-Zeitreihe aus Abbildung 3.13 handelt es sich um eine gewichtete Residuen-Zeitreihe.

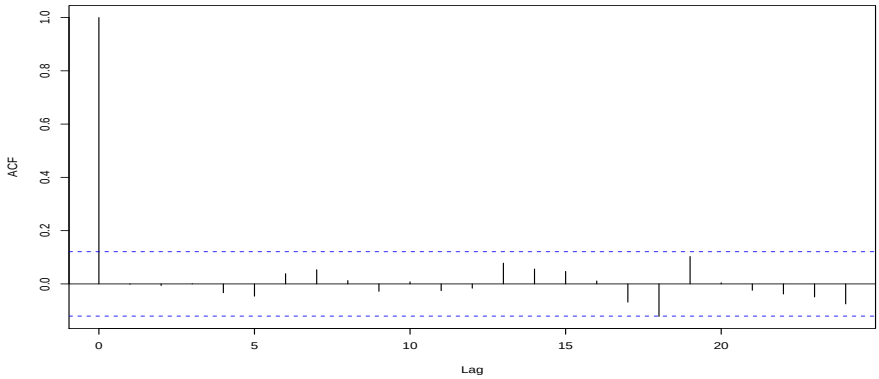

Abbildung 3.17: Empirische ACF der Residuen aus Abb. 3.16

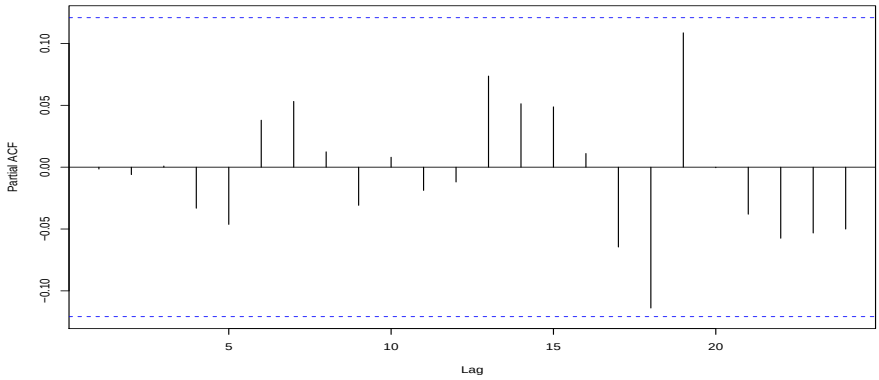

Abbildung 3.18: Empirische partielle ACF der Residuen aus Abb. 3.16

3.5.2 *ARIMA*-Modelle

3.5.2 Definition: Ist B der Backward Shift Operator aus (3.3) und $d \geq 0$, so heißt (X_t) ein *ARIMA*(p, d, q)-*Prozess*, falls

$$Y_t := (1 - B)^d X_t = \nabla^d X_t \tag{3.48}$$

ein zukunftsunabhängiger *ARMA*(p, q)-Prozess ist. Dabei bezeichnet ∇ den Differenzenoperator aus (3.4).

3.5.3 Bemerkung: Ist in (3.48) $d \geq 1$, so ist der *ARIMA*(p, d, q)-Prozess (X_t) **nicht** stationär. Der differenzierte Prozess (Y_t) ist jedoch definitionsgemäß ein stationärer *ARMA*-Prozess, dessen Autokovarianzfunktion und Parameter mittels der in den vorhergehenden Abschnitten beschriebenen Methoden geschätzt werden können.

3.5.4 Bemerkung: Es sei (X_t) ein *ARIMA*$(p, 1, q)$-Prozess. Wegen $Y_t = (1 - B)X_t = X_t - X_{t-1}$ lässt sich der *ARIMA*$(p, 1, q)$-Prozess darstellen gemäß

$$X_t = X_1 + \sum_{j=2}^{t} (X_j - X_{j-1}) = X_1 + \sum_{j=2}^{t} Y_j.$$

Damit ergibt sich für die Kovarianzen des *ARIMA*$(p, 1, q)$-Prozesses (ohne Einschränkung sei $s \leq t$ und γ bezeichne die Kovarianzfunktion von (Y_t))

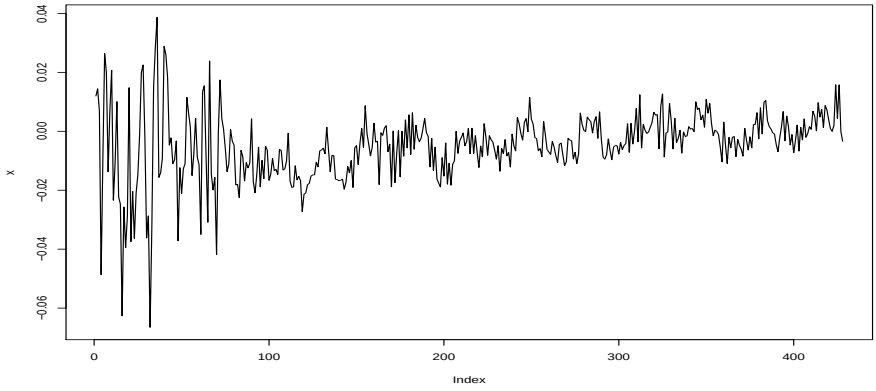

Abbildung 3.19: Die ungewichtete GPS-Zeitreihe H1JO022610

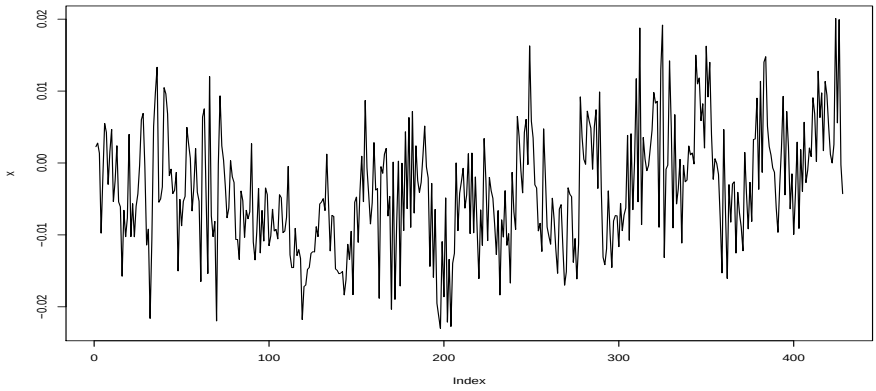

Abbildung 3.20: Die gewichtete GPS-Zeitreihe H1JO022610

$$\begin{aligned}\mathrm{Cov}(X_s, X_t) &= \mathrm{Var}(X_1) + \mathrm{Cov}(X_1, \sum_{j=2}^{s} Y_j) + \mathrm{Cov}(X_1, \sum_{j=2}^{t} Y_j) + \mathrm{Cov}(\sum_{i=2}^{s} Y_i, \sum_{j=2}^{t} Y_j) \\ &= \mathrm{Var}(X_1) + 2\sum_{j=2}^{s} \mathrm{Cov}(X_1, Y_j) + \sum_{j=s+1}^{t} \mathrm{Cov}(X_1, Y_j) + \sum_{i=2}^{s}\sum_{j=2}^{t} \gamma(|j-i|).\end{aligned}$$

3.5.5 Bemerkung: Wie bereits in Abschnitt 3.1.3 angesprochen, wird ein durch ein Polynom d-ten Grades modellierbarer Trend durch die Differenzenbildung in (3.48) eliminiert. *ARIMA*-Modelle bieten daher eine Möglichkeit, Daten mit trendartigem Verhalten zu beschreiben.

3.5.6 Beispiel: Es sei $\phi_1 \in (-1, 1)$ und

$$(1 - \phi_1 B)(1 - B)X_t = Z_t,\ (Z_t) \sim WN(0, \sigma^2).$$

Dann ist $Y_t = (1 - B)X_t$ ein nicht vorgreifender $AR(1)$-Prozess mit $Y_t = \sum_{j=0}^{\infty} \phi_1^j Z_{t-j}$, vgl. Beispiel 3.2.15. Also ist (X_t) ein $ARIMA(1, 1, 0)$-Prozess.

Die Abbildungen 3.21 und 3.22 zeigen einen Pfad des $ARIMA(1, 1, 0)$-Prozesses (X_t) aus Beispiel 3.5.6 mit $\phi_1 = -\frac{1}{2}$ und $\sigma^2 = 1$ bzw. den entsprechenden differenzierten Prozess (Y_t).

Obwohl die Autokorrelationsfunktion nur für stationäre Prozesse definiert ist, lässt sich die empirische ACF gemäß (3.8) für jede beliebige Zeitreihe berechnen. Die empirische ACF des Prozesses aus Abbildung 3.21 ist in Abbildung 3.23 zu sehen. Das langsame Abfallen der empirischen ACF ist für einen *ARIMA*-Prozess typisch. Abbildung 3.24 zeigt die empirische partielle Autokorrelationsfunktion von (X_t). Nach Differenzierung erhält man eine für einen $AR(1)$-Prozess mit $\phi_1 \in (-1, 0)$ typische empirische Autokorrelationsfunktion bzw. empirische partielle Autokorrelationsfunktion, vgl. Abb. 3.25 und Abb. 3.26.

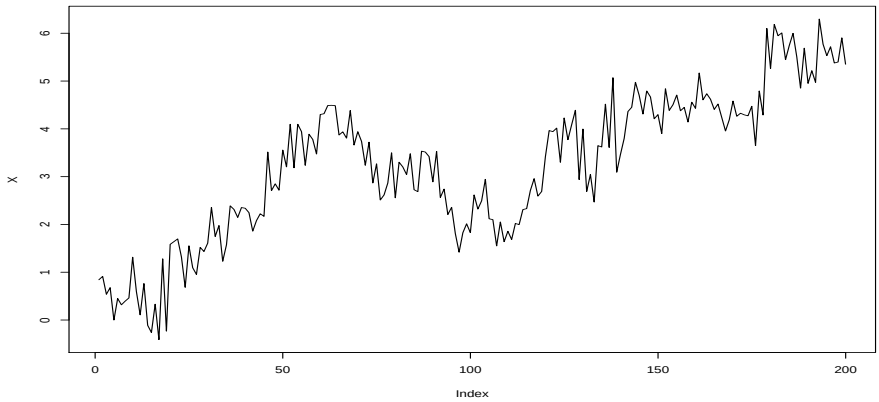

Abbildung 3.21: Pfad des $ARIMA(1,1,0)$-Prozesses (X_t)

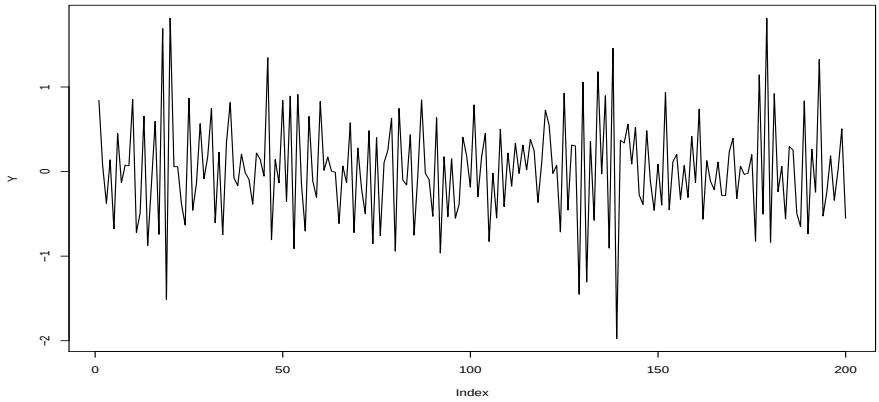

Abbildung 3.22: Der differenzierte Prozess (Y_t)

3.5.3 Die Brown'sche Bewegung

Ein wichtiger nichtstationärer Prozess mit stetigem Zeitparameter ist die *Brown'sche Bewegung*. Die Brown'sche Bewegung gehört zur Klasse der *Gauß-Prozesse*.

3.5.7 Definition: Es sei (X_t) ein stochastischer Prozess auf einem Intervall $[a,b]$. Ist für endlich viele $a \leq t_1 \leq \cdots \leq t_n \leq b$ der Vektor $(X_{t_1}, \ldots, X_{t_n})^\top$ stets n-dimensional normalverteilt, so heißt (X_t) ein *Gauß-Prozess*. Gilt zudem $\mathrm{E}(X_t) = 0$ für alle $t \in [a,b]$, so wird (X_t) *zentrierter* Gauß-Prozess genannt.

Ein Gauß-Prozess (X_t) ist durch seine Erwartungswertfunktion $\mu(t) := \mathrm{E}(X_t)$ und seine Kovarianzfunktion $g(s,t) := \mathrm{Cov}(X_s, X_t)$ eindeutig bestimmt. Ein spezieller Gauß-Prozess, der künftig von Bedeutung sein wird, ist die Brown'sche Bewegung im $\mathbb{R}^1$.

3.5.8 Definition: Es sei $B := (B_z)_{z\in[0,1]}$ ein reellwertiger stochastischer Prozess mit

(*i*) Für $0 \leq z_0 \leq z_1 \leq \cdots \leq z_k \leq 1$ sind die Zufallsvariablen $B(z_1) - B(z_0), \ldots, B(z_k) - B(z_{k-1})$ stochastisch unabhängig (s. Anhang B.1),

(*ii*) für $i < j$ ist die Zufallsvariable $B(z_j) - B(z_i)$ normalverteilt mit Erwartungswert 0 und Varianz $(z_j - z_i)$,

dann heißt B eine *(reelle) Brown'sche Bewegung*. Gilt außerdem $B(0) = 0$ P-f.s. (P-fast sicher, vgl. Anhang B.1), so heißt die Brown'sche Bewegung *normal*.

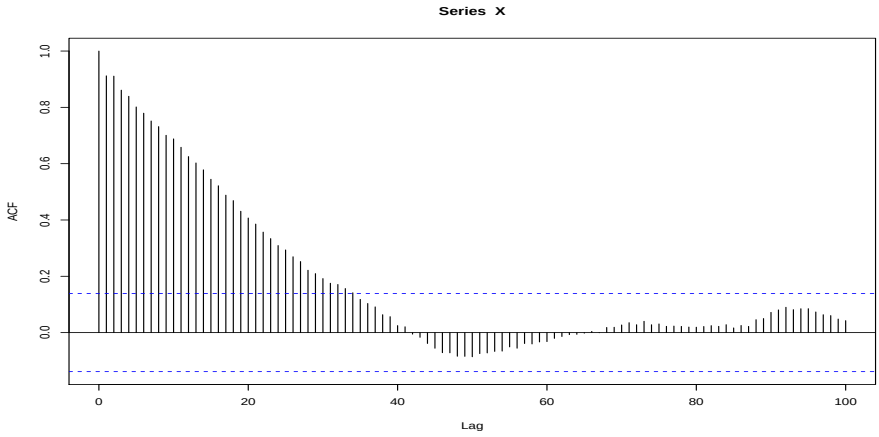

Abbildung 3.23: Die empirische ACF des Prozesses aus Abb. 3.21

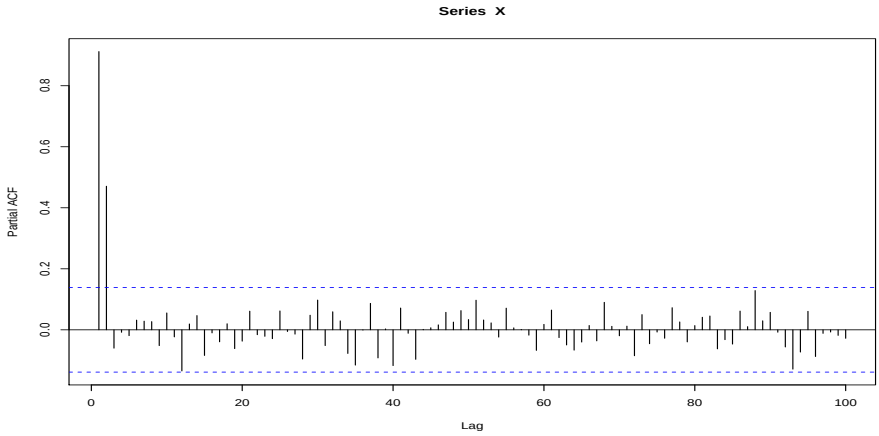

Abbildung 3.24: Die empirische partielle ACF des Prozesses aus Abb. 3.21

Die Abbildung 3.27 zeigt einen Pfad einer reellen, normalen Brown'schen Bewegung.

3.5.9 Lemma: *Eine reelle, normale Brown'sche Bewegung ist ein zentrierter Gauß-Prozess mit Kovarianzfunktion*

$$g(z_1, z_2) := Cov(B(z_1), B(z_2)) = \min\{z_1, z_2\}.$$

Beweis: Es seien $k \in \mathbb{N}$, $(z_1, \ldots, z_k) \in [0,1]^k$ und $(a_1, \ldots, a_k) \in \mathbb{R}^k$ beliebig. Setze $b_j := \sum_{i=j}^{k} a_i,\ j = 1, \ldots, k$. Somit ergibt sich $b_k = a_k$ und $b_j = b_{j+1} + a_j$, $j = 1, \ldots, k-1$, und es gilt

$$\begin{aligned} \sum_{i=1}^{k} a_i B(z_i) &= \sum_{i=1}^{k-1} (b_i - b_{i+1}) B(z_i) + b_k B(z_k) \\ &= b_1 B(z_1) + \sum_{i=2}^{k} b_i (B(z_i) - B(z_{i-1})). \end{aligned}$$

Da die $B(z_1), B(z_2) - B(z_1), \ldots, B(z_k) - B(z_{k-1})$ unabhängig und jeweils normalverteilt sind, ist $\sum_{i=1}^{k} a_i B(z_i)$ ebenfalls normalverteilt. Da die $a_i,\ i = 1, \ldots, k$, beliebig gewählt waren, besitzt der Vektor $(B(z_1), \ldots, B(z_k))$ eine k-dimensionale Normalverteilung. Somit ergibt sich, da auch k und die z_i beliebig sind, dass B ein Gauß-Prozess ist.

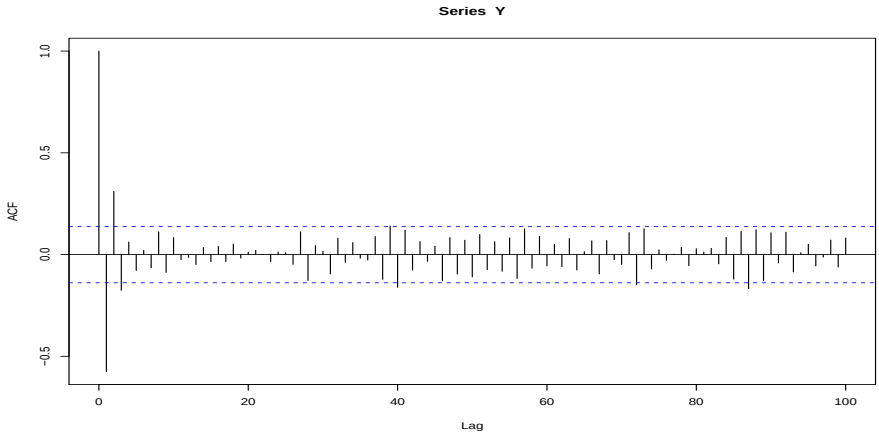

Abbildung 3.25: Die empirische ACF des Prozesses aus Abb. 3.22

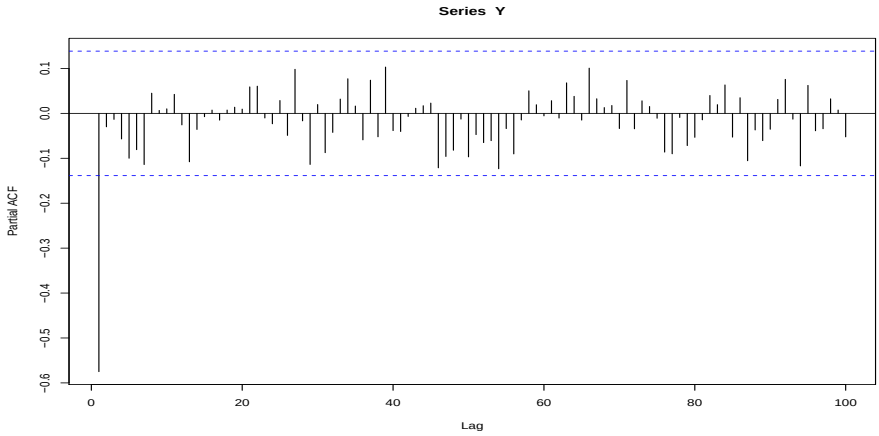

Abbildung 3.26: Die empirische partielle ACF des Prozesses aus Abb. 3.22

Nach Definition der normalen Brown'schen Bewegung ist

$$0 = \mathrm{E}(B(z) - B(0)) = \mathrm{E}(B(z)) - \mathrm{E}(B(0)) = \mathrm{E}(B(z)),\ z \in [0,1],$$

also $(B(z))_{z\in[0,1]}$ ist ein zentrierter Gauß-Prozess. Weiter gilt (ohne Einschränkung sei $z_1 \leq z_2$)

$$\begin{aligned} \mathrm{Cov}(B(z_1), B(z_2)) &= \mathrm{E}[B(z_1)B(z_2)] \\ &= \mathrm{E}[(B(z_2) - B(z_1))B(z_1)] + \mathrm{E}[B(z_1)^2] \\ &= \mathrm{Var}(B(z_1)) = z_1 = \min\{z_1, z_2\}. \end{aligned} \tag{3.49}$$

□

Aus Lemma 3.5.9 ergibt sich insbesondere

$$\mathrm{E}(B(z)) = 0 \text{ und } \mathrm{Var}(B(z)) = z \text{ für alle } z \in [0,1].$$

3.5.10 Bemerkung: Betrachtet man den Pfad der Brown'schen Bewegung aus Abbildung 3.27, so ließe sich - wäre die Verteilung nicht bekannt - ein Trend in den Beobachtungen vermuten. Tatsächlich aber besitzt die Brown'sche Bewegung eine konstante Erwartungswertfunktion $\mathrm{E}(B(z)) = 0,\ z \in [0,1]$. Das trendähnliche Verhalten wird also durch Korrelationen erzeugt. Dieses Beispiel macht klar, dass bei unbekannter Verteilung eines stochastischen Prozesses der Einfluss von Korrelationen nicht von einem echten Trend zu unterscheiden ist.

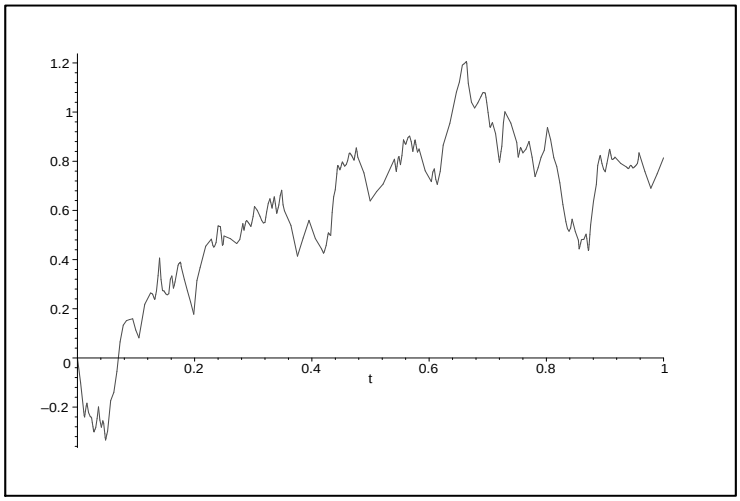

Abbildung 3.27: Ein Pfad einer reellen Brown'schen Bewegung

Um sich die Bedeutung der Brown'schen Bewegung im $\mathbb{R}^1$ zu veranschaulichen, betrachte man den folgenden Random Walk:

Ein Teilchen mache zu jedem Zeitpunkt $\frac{i}{n},\ i = 1, \ldots, n$, einen Schritt auf der reellen Zahlengerade. Die Schrittweite werde durch eine Zufallsvariable $\frac{1}{\sqrt{n}}\xi_i$ bestimmt, wobei die Zufallsvariablen $\xi_1, \ldots, \xi_n$ unabhängig und identisch $\mathcal{N}(0,1)$-verteilt sind. Bezeichnet $[z]$ die größte ganze Zahl kleiner oder gleich $z \in \mathbb{R}$, so beschreibt die Zufallsvariable

$$X_n(z) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[nz]} \xi_i \tag{3.50}$$

den Zustand des Teilchens zur Zeit $z \in [0,1]$. Ein Pfad des Prozesses $(X_n(z))_{z\in[0,1]}$ ist keine stetige Funktion. Dennoch kann man sich für sehr große n $(X_n(z))_{z\in[0,1]}$ als Näherung für die reelle, normale Brown'sche Bewegung vorstellen.

Präziser und allgemeiner wird dies im folgenden Satz formuliert. Dazu wird $(X_n(z))$ zunächst leicht modifiziert, so dass die Pfade stetig werden.

3.5.11 Satz: (Satz von Donsker) *Die Zufallsvariablen $\xi_i,\ i \in \mathbb{N}$, seien unabhängig und identisch verteilt mit Erwartungswert 0 und Varianz $\sigma^2 > 0$. Weiter sei $[z]$ das größte Ganze von z, d.h. $[z] = \max\{k \in \mathbb{Z} : k \le z\}$. Dann gilt für den stochastischen Prozess $(X_n(z))$ mit $X_n(z) := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{[nz]} \xi_i + (nz - [nz])\frac{1}{\sigma\sqrt{n}}\xi_{[nz]+1}$:*

$$(X_n(z)) \xrightarrow{\mathcal{D}} B \text{ in } C[0,1]. \tag{3.51}$$

Dabei ist B die reelle, normale Brown'sche Bewegung.

Die Konvergenz in (3.51) bezieht sich auf die Konvergenz im Raum $C[0,1]$ aller auf dem Intervall $[0,1]$ stetigen Funktionen (bezüglich der Maximumsnorm (A.2), siehe Anhang A.1).

3.5.12 Bemerkung: Der Prozess $(X_n(z))$ aus Satz 3.5.11 stimmt in den Punkten $[iz],\ i = 1, \ldots, n$, mit dem Prozess (3.50) überein. Die zweite additive Komponente von $X_n(z)$ aus Satz 3.5.11 verbindet die Punkte $X_n([iz]),\ i = 1, \ldots n$, linear. Damit ist $(X_n(z))$ stetig.

Man kann auf die in Bemerkung 3.5.12 hingewiesene, den Prozess $(X_n(z))$ stetig machende Komponente verzichten, wenn man sich eines geeigneten Funktionenraumes bedient, der auch nichtstetige Funktionen enthält:

3.5.13 Definition: Es bezeichnet $D[0,1]$ den Raum aller Funktionen $\varphi : [0,1] \to \mathbb{R}$ mit

(*i*) $\lim_{s\downarrow t} \varphi(s) = \varphi(t),\ t \in [0,1)$, d.h. φ ist rechtsseitig stetig,

(*ii*) $\lim_{s\uparrow t} \varphi(s)$ existiert für alle $t \in (0,1]$.

Der Raum $D[0,1]$ wird mit der *Skorohod Topologie* (s. BILLINGSLEY (1968)) versehen. Damit konvergiert eine Funktionenfolge $(\varphi_n) \in D[0,1]$ genau dann gegen ein $\varphi \in D[0,1]$, wenn eine Funktionenfolge (λ_n) aus

$$\begin{aligned}\Lambda \quad := \quad & \{\lambda : [0,1] \to [0,1] \ : \ \lambda \text{ ist streng monoton wachsend, stetig und bijektiv}\} \\ & \text{(N.B.: Es gilt } \lambda(0) = 0 \text{ sowie } \lambda(1) = 1)\end{aligned}$$

existiert mit

$$\begin{aligned}& \sup_{t\in[0,1]} |\varphi_n(\lambda_n(t)) - \varphi(t)| \quad \stackrel{n\to\infty}{\longrightarrow} 0 \\ \text{und} \qquad & \sup_{t\in[0,1]} |\lambda_n(t) - t| \qquad \stackrel{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Wir können jetzt den Satz von Donsker in einer einfacheren Form formulieren, so dass er direkt auf den stochastischen Prozess $(X_n(z))$ mit $X_n(z)$ wie in (3.50) angewendet werden kann:

3.5.14 Satz: (Satz von Donsker) *Die Zufallsvariablen ξ_i, $i \in \mathbb{N}$, seien unabhängig und identisch verteilt mit Erwartungswert 0 und Varianz $\sigma^2 > 0$. Dann gilt für den stochastischen Prozess $(X_n(z))$ mit $X_n(z) := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{[nz]} \xi_i$:*

$$(X_n(z)) \xrightarrow{\mathcal{D}} B \text{ in } D[0,1].$$

Dabei ist B die reelle, normale Brown'sche Bewegung.

Beweis: BILLINGSLEY (1968).

3.5.15 Bemerkung: Der in Definition 3.5.8 definierte Prozess ist nicht eindeutig bestimmt. Ist z.B. (B_t) eine Brown'sche Bewegung und (X_t) ein stochastischer Prozess mit

$$P(X_t = B_t) = 1 \text{ für alle } t \in [0,1], \tag{3.52}$$

so ist (X_t) ebenfalls eine Brown'sche Bewegung, deren Pfade jedoch u.U. andere Eigenschaften besitzen. Man spricht daher in Bezug auf (3.52) von verschiedenen *Versionen* der Brown'schen Bewegung.

3.5.16 Bemerkung:

a) Es existiert eine Version der Brown'schen Bewegung, deren Pfade P-fast sicher stetig sind.

b) Die Pfade einer Brown'schen Bewegung sind P-fast sicher in **keinem** Punkt des Intervalles $[0,1]$ differenzierbar.

c) Die Pfade einer Brown'schen Bewegung sind P-fast sicher von **un**beschränkter Variation auf jedem endlichen Intervall $[a,b] \subset [0,1]$.

d) Eine Brown'sche Bewegung ist ein *selbstähnlicher* Prozess (Definition s. unten).

Die Bemerkungen 3.5.16 a) bis c) werden z. B. in ASH / GARDNER (1994) bewiesen. Zu d) siehe BERAN (1994).

3.5.17 Definition: Ein stochastischer Prozess $(X_t)_{t\in[0,1]}$ heißt *selbstähnlich* mit *Selbstähnlichkeitsparameter* H , falls für alle $c > 0$ gilt

$$\frac{1}{c^H} X(ct) \sim X(t), \ t \in [0, \frac{1}{c}]. \tag{3.53}$$

Für $H = 1$ bedeuted dies, dass die Vergrößerung eines (beliebig kleinen) Ausschnittes dieselbe Verteilung besitzt, wie der Prozess über den gesamten Zeitbereich.

3.5.18 Definition: Ist (B_z) eine reelle, normale Brown'sche Bewegung auf $[0,1]$, so heißt der Prozess $(B_0(z))$ mit

$$B_0(z) := B(z) - zB(1)$$

eine *Brown'sche Brücke.*

Für eine Brown'sche Brücke gilt stets $B_0(1) = 0$ P-f.s. Man kann eine Brown'sche Brücke daher als eine im Punkt $z = 1$ an der Zeitachse "festgebundene" Brown'sche Bewegung betrachten. Abbildung 3.28 zeigt die Brown'sche Bewegung aus Abbildung 3.27 als Brown'sche Brücke. Man beachte die unterschiedliche Skalierung.

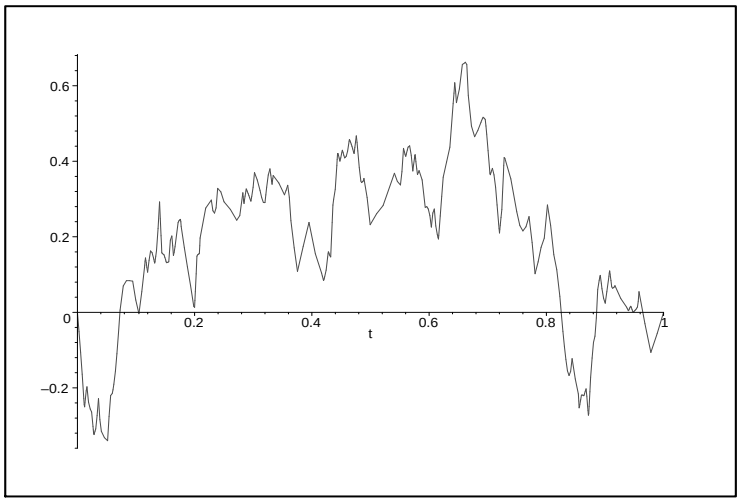

Abbildung 3.28: Ein Pfad einer Brown'schen Brücke

3.5.19 Lemma: *Eine Brown'sche Brücke auf* $[0,1]$ *ist ein zentrierter Gauß-Prozess mit Kovarianzfunktion*

$$\tilde{g}(z_1, z_2) := Cov(B_0(z_1), B_0(z_2)) = \min\{z_1, z_2\} - z_1 z_2.$$

Die Behauptung ergibt sich sofort aus Lemma 3.5.9.

3.6 Filter

3.6.1 Lineare Filter

3.6.1 Definition: Man sagt, der Prozess (X_t) sei der Output eines *linearen Filters,* angewandt auf einen stochastischen Prozess (U_t), falls eine Funktion k auf $\mathbb{R}^2$ existiert, so dass

$$X_t = \int_{-\infty}^{\infty} k(t,u) U_u du \ . \tag{3.54}$$

Üblicherweise stellt man einen Filter als "Blackbox" dar, die ein eingehendes Signal, z.B. einen stochastischen Prozess (U_t) (Input), in ein anderes Signal (X_t) (Output) umwandelt (s. Abbildung 3.29).

Abbildung 3.29: Linearer Filter

Der Filter heißt *zeitinvariant,* falls die Funktion $k(t, t-u)$ konstant in t ist. In diesem Falle setzt man $k(u) := k(t, t-u)$, und (3.54) nimmt die Form

$$X_t = \int_{-\infty}^{\infty} k(u) U_{t-u} du \tag{3.55}$$

$$= (k * U)(t) \tag{3.56}$$

an. Dabei bezeichnet $k * U$ die *Faltung* von k und U (vgl. Anhang A.2).

Man nennt die Funktion k *Impulse Response Funktion.* Sie beschreibt den Output, falls das Eingangssignal der Dirac-Impuls ist (s. Anhang A.2). Im zeitinvarianten Fall ergibt sich also

$$X_t = \int_{-\infty}^{\infty} k(u)\delta(t-u)du = k(t).$$

Die Fourier-Transformierte von k

$$\Gamma(\omega) = \int_{-\infty}^{\infty} k(u) e^{-i\omega u} du$$

heißt *Transfer-Funktion* des Filters. Sie existiert, falls k stetig und von beschränkter Variation ist. Ein Filter ist sowohl durch seine Impulse Response Funktion als auch durch seine Transfer-Funktion eindeutig spezifiziert.

3.6.2 Beispiel: Besitzt das in den Filter eingehende Signal die Form

$$U_t = A \cdot e^{i\omega t} = A \cdot (\cos(\omega t) + i \sin(\omega t)),$$

so erhält man bei einem zeitinvarianten Filter das ausgehende Signal

$$X_t = \int_{-\infty}^{\infty} k(u) \cdot A \cdot e^{i\omega(t-u)} du = A \cdot \Gamma(\omega) \cdot e^{i\omega t}.$$

Die Transfer-Funktion beschreibt also die Art und Weise, wie der Filter auf das eingehende Signal einwirkt. Der Filter verstärkt Signale aus einem Frequenzbereich, auf dem Γ betragsmäßig groß ist und dämpft jene, auf denen $|\Gamma|$ kleine Werte annimmt.

3.6.3 Bemerkung: Die Spektraldichte $f_X(\omega)$ des gefilterten Prozesses (X_t) ist gegeben durch (vgl. Priestley (1981[1]))

$$f_X(\omega) = f_U(\omega) \cdot |\Gamma(\omega)|^2,$$

wobei f_U die Spektraldichte des Eingangssignals (U_t) bezeichnet.

3.6.4 Definition: Im Falle $T = \mathbb{Z}$ nennt man X_t den Output eines linearen Filters, falls Konstanten $\psi_{t,j}$ $(t, j \in \mathbb{Z})$ existieren mit

$$X_t = \sum_{j=-\infty}^{\infty} \psi_{t,j} U_j \ (t \in \mathbb{Z}). \tag{3.57}$$

Der Filter heißt *zeitinvariant*, falls die Gewichte $\psi_{t,t-j}$ unabhängig vom Zeitpunkt t sind. in diesem Falle besitzt (3.57) mit $\psi_j := \psi_{t,t-j}$ die Form

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j U_{t-j}. \tag{3.58}$$

3.6.5 Beispiel: Der allgemeine lineare Prozess

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

ist ein gefilterter White Noise Prozess.

Im zeitinvarianten Fall mit $T = \mathbb{Z}$ ist die Impulse Response Funktion gegeben durch die Folge (ψ_j) und die Transfer-Funktion durch

$$\Gamma(\omega) = \sum_{j=-\infty}^{\infty} \psi_j e^{-i\omega j}.$$

3.6.6 Satz: *Ist der Prozess (U_t) stationär und gilt $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, so ist der Output (X_t) des linearen Filters $X_t = \sum_{j=-\infty}^{\infty} \psi_j U_{t-j}$ ebenfalls stationär mit*

$$E(X_t) = \left(\sum_{j=-\infty}^{\infty} \psi_j \right) E(U_t)$$

und

$$\begin{aligned} \gamma_X(h) = Cov(X_t, X_{t-h}) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j Cov(U_{t-i}, U_{t-j-h}) \\ &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma_U(h+j-i). \end{aligned}$$

Beweis: Brockwell / Davis (1991).

3.6.2 Kalman-Filter

Da die GPS-Auswertesoftware GIPSY/OASIS zur Auswertung der Daten einen Kalman-Filter verwendet, soll hier das Prinzip des Kalman-Filters mit diskretem Zeitparameter erläutert werden. Für weiterführende Informationen wird auf Kalman (1960), Priestley (1981[2]) oder Brockwell / Davis (1991) verwiesen.

Die Anwendung eines Kalman-Filters erfordert eine bestimmete Darstellung, die *State Space Darstellung* der Zeitreihe. Zur Einführung soll ein autoregressiver Prozess auf seine State Space Darstellung gebracht werden.

3.6.7 Beispiel: Es sei (X_t) der $AR(2)$-Prozess

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t,\ Z_t \sim WN(0, \sigma^2). \tag{3.59}$$

Setze nun

$$\begin{aligned} X_t^{(1)} &:= \phi_2 X_{t-1}, \\ X_t^{(2)} &:= X_t\ . \end{aligned} \tag{3.60}$$

Damit ist (3.59) in Verbindung mit (3.60) äquivalent zu

$$\underbrace{\begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix}}_{=:\mathbf{X}_t} = \underbrace{\begin{pmatrix} 0 & \phi_2 \\ 1 & \phi_1 \end{pmatrix}}_{=:A_t} \underbrace{\begin{pmatrix} X_{t-1}^{(1)} \\ X_{t-1}^{(2)} \end{pmatrix}}_{=\mathbf{X}_{t-1}} + \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix} Z_t}_{=:\mathbf{Z}_t}. \tag{3.61}$$

Der Vektor $\mathbf{X}_t$ heißt *Zustandsvektor* (*State Vector*), die Darstellung (3.61) *State Space Darstellung* der Zeitreihe (X_t). Die State Space Darstellung einer Zeitreihe besitzt die Eigenschaft, dass jeder Zustand $\mathbf{X}_t$ des Prozesses $(\mathbf{X}_t)$ lediglich vom vorangehenden Zustand $\mathbf{X}_{t-1}$ abhängt, nicht jedoch von der "Vergangenheit" $\mathbf{X}_{t-2}, \mathbf{X}_{t-3}, \ldots$ Diese Eigenschaft nennt man *Markov-Eigenschaft*, ein entsprechender Prozess heißt *Markov-Prozess*. Man beachte, dass X_t aus $\mathbf{X}_t$ erhalten werden kann durch

$$X_t = (0,1) \cdot \begin{pmatrix} X_t^{(1)} \\ X_t^{(2)} \end{pmatrix}.$$

Entsprechend lässt sich auch ein beliebiger $ARMA(p,q)$-Prozess

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \psi_1 Z_{t-1} + \cdots + \psi_q Z_{t-q},\ (Z_t) \sim WN(0,\sigma^2),$$

in der Form eines (höherdimensionalen) Markov-Prozesses, also in der State Space Form

$$\mathbf{X}_t = A_t \mathbf{X}_{t-1} + \mathbf{Z}_t$$

mit geeigneten Zufallsvektoren $\mathbf{X}_t$ und $\mathbf{Z}_t$ sowie einer Matrix A_t schreiben. Näheres hierzu s. Brockwell / Davis (1991).

3.6.8 Definition: Ein n-dimensionales *lineares dynamisches Modell* einer Zeitreihe (X_t) besteht aus

(i) einer Differenzengleichung (Systemgleichung) in State Space Form

$$\mathbf{X}_t = A_t \mathbf{X}_{t-1} + \mathbf{Z}_t,\ t \in \mathbb{Z}, \tag{3.62}$$

wobei $\mathbf{X}_t$, $t \in \mathbb{Z}$, ein n-dimensionaler Zufallsvektor und A_t eine bekannte $n \times n$-Matrix ist, die vom Zeitparameter t abhängen kann. $\mathbf{Z}_t$ ist ein n-dimensionaler $WN(0,\Xi_t)$ Fehlervektor ($t \in \mathbb{Z}$), genannt *Prozessrauschen*.

(ii) einer Messgleichung

$$\mathbf{Y}_t = H_t \mathbf{X}_t + \epsilon_t,\ t \in \mathbb{Z}. \tag{3.63}$$

Dabei ist $\mathbf{Y}_t$ ein m-dimensionaler Vektor ($m \leq n$), der Messungen des Prozesses (X_t) enthält, H_t eine bekannte $m \times n$-Matrix und ϵ_t ein m-dimensionaler $WN(0,\Sigma_t)$ Fehlervektor, genannt *Messrauschen*, der weder mit $(\mathbf{X}_t)$ noch mit $(\mathbf{Z}_t)$ korreliert ist.

Das lineare dynamische Modell beschreibt also eine Situation, in der der Prozess $(\mathbf{X}_t)$ u.U. nur indirekt beobachtet werden kann. Insbesondere unterliegen die Beobachtungen einem Messrauschen, das nicht zum Prozess gehört und weder mit $(\mathbf{X}_t)$ noch mit $(\mathbf{Z}_t)$ korreliert ist. Im Folgenden wird die Existenz von $\mathrm{E}[\mathbf{X_t X_t}^\top]$ und $\mathrm{E}[\mathbf{Y_t Y_t}^\top]$ vorausgesetzt.

Die Problemstellung ist nun, aufgrund der Messungen $(\mathbf{y}_t)$ von $(\mathbf{Y}_t)$ die Werte $(\mathbf{x}_t)$ der Zeitreihe $(\mathbf{X}_t)$ (und damit (X_t)) "herauszufiltern" (*Filtering Problem*). Konkret besteht die Aufgabe darin, aus den Realisierungen $\mathbf{y}_t$, $t = t_0,\ldots,t_l$, der Zeitreihe $(\mathbf{Y}_t)$ den Wert $\mathbf{x}_{t_l+k}$ ($k \in \mathbb{N}\cup\{0\}$), zu schätzen. Man beachte, dass die Aufgabenstellung sowohl für $k=0$ (Filterungs-Problem), als auch für $k>0$ (Vorhersage-Problem) Sinn macht. Der *Kalman-Filter* ist eine rekursive Methode zur Berechnung des optimalen (optimal i.S.v. Kleinste-Quadrate) linearen Schätzers $\hat{\mathbf{X}}_{t_l+k}$ für $\mathbf{X}_{t_l+k}$, also jenes linearen Schätzers mit

$$\mathrm{E}\left(\|\hat{\mathbf{X}}_{t_l+k} - \mathbf{X}_{t_l+k}\|^2\right) = \min_{\mathbf{x}\in\mathcal{Y}(t_0,\ldots,t_l)} \mathrm{E}\left(\|\mathbf{x} - \mathbf{X}_{t_l+k}\|^2\right).$$

Dabei ist

$$\mathcal{Y}(t_0,\ldots,t_l) := \left\{ \sum_{t=t_0}^{t_l} \mathbf{a}_t^\top \mathbf{y}_t \ :\ \mathbf{a}_t \in \mathbb{R}^m \right\}.$$

Geometrisch betrachtet entspricht der Schätzer $\hat{\mathbf{X}}_{t_l+k}$ der Orthogonalprojektion von $\mathbf{X}_{t_l+k}$ auf $\mathcal{Y}(t_0,\ldots,t_l)$.

Im Folgenden bezeichne $\hat{\mathbf{X}}(t+k|t)$ ($t \geq t_0,\ k \in \mathbb{N}\cup\{0\}$) den optimalen linearen Schätzer für $\mathbf{X}_{t+k}$ bei gegebenen $\mathbf{y}_{t_0},\ldots,\mathbf{y}_{t-1},\mathbf{y}_t$ und entsprechend $\hat{\mathbf{Y}}(t+k|t)$ den optimalen linearen Schätzer für $\mathbf{Y}_{t+k}$ bei gegebenen $\mathbf{y}_{t_0},\ldots,\mathbf{y}_{t-1},\mathbf{y}_t$. Es soll nun der Kalman-Filter für $k=0$ (gesucht ist $\hat{\mathbf{X}}(t|t),\ t \geq t_0$) beschrieben werden. Für den Beweis sowie den Kalman Vorhersage-Algorithmus (gesucht sind $\hat{\mathbf{Y}}_{t+k}$ und $\hat{\mathbf{X}}_{t+k}$ für $k>0$) wird auf Brockwell / Davis (1991) verwiesen.

3.6.9 Bemerkung: Der Kalman-Filter benötigt einen initialen Schätzer $\hat{\mathbf{X}}(t_0|0)$ für $\mathbf{X}_{t_0}$. Hierzu sei $\mathbf{Y}_0$ ein m- dimensionaler Zufallsvektor mit

$$\mathrm{E}[\mathbf{Y}_0 \cdot \mathbf{Z_t}] = \mathrm{E}[\mathbf{Y}_0 \cdot \epsilon_t] = 0 \text{ für alle } t.$$

Dann ist $\hat{\mathbf{X}}(t_0|0)$ gegeben durch

$$\hat{\mathbf{X}}(t_0|0) \ =\ \mathrm{E}[\mathbf{X}_{t_0}\mathbf{Y}_0^\top]\ \{\mathrm{E}[\mathbf{Y}_0\mathbf{Y}_0^\top]\}^-\ \mathbf{Y}_0,$$

wobei A^- eine verallgemeinerte Inverse (s. Def. 5.1.7) der Matrix A bezeichnet. Häufig kann $\mathbf{Y}_0 = \mathbf{1}_m = (1,\ldots,1)^\top$ gewählt werden (näheres s. Brockwell / Davis (1991)). In diesem Falle ist

$$\hat{\mathbf{X}}(t_0|0) = \mathrm{E}(\mathbf{X}_{t_0}).$$

Der Kalman-Filter

Für $t = t_0$ sei $\hat{\mathbf{X}}(t|t-1)$ der initiale Schätzer aus Bemerkung 3.6.9 und $\hat{\mathbf{Y}}(t|t-1) = H_t \cdot \hat{\mathbf{X}}(t|t-1)$. Für $t > t_0$ bezeichne $\hat{\mathbf{X}}(t-1|t-1)$ den optimalen linearen Schätzer von $\mathbf{X}_{t-1}$ aufgrund der Daten $\mathbf{y}_{t_0}, \ldots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}$. In diesem Falle sind die optimalen linearen Schätzer von $\mathbf{X}_t$ und $\mathbf{Y}_t$ aufgrund der Daten $\mathbf{y}_{t_0}, \ldots, \mathbf{y}_{t-2}, \mathbf{y}_{t-1}$ gegeben durch

$$\begin{aligned}\hat{\mathbf{X}}(t|t-1) &= A_t \cdot \hat{\mathbf{X}}(t-1|t-1)\\ \hat{\mathbf{Y}}(t|t-1) &= H_t \cdot \hat{\mathbf{X}}(t|t-1) = H_t \cdot A_t \cdot \hat{\mathbf{X}}(t-1|t-1)\end{aligned}$$

mit A_t und H_t wie in (3.62) bzw. (3.63).

Liegt nun eine weitere Beobachtung $\mathbf{y}_t$ vor ($t \geq t_0$), so wird der Vorhersagefehler

$$\mathbf{y}_t - \hat{\mathbf{Y}}(t|t-1) = \mathbf{y}_t - H_t\hat{\mathbf{X}}(t|t-1)$$

berechnet und der Schätzer für $\mathbf{X}_t$ angepasst ("Updating") gemäß

$$\hat{\mathbf{X}}(t|t) = \hat{\mathbf{X}}(t|t-1) + K_t \cdot \underbrace{\Big(\mathbf{y}_t - H_t \cdot \hat{\mathbf{X}}(t|t-1)\Big)}_{Vorhersagefehler\ von\ Y_t}, \tag{3.64}$$

wobei die Matrix

$$\begin{aligned}K_t := \quad & \underbrace{\mathrm{E}\left[\Big(\mathbf{X}_t - \hat{\mathbf{X}}(t|t-1)\Big)\Big(\mathbf{X}_t - \hat{\mathbf{X}}(t|t-1)\Big)^\top\right]}_{=:\Omega_t = Cov-Matrix\ des\ Vorhersage-Fehlers\ von\ \mathbf{X}_t}\\ & \cdot H_t^\top \cdot \underbrace{\mathrm{E}\left[\Big(\mathbf{Y}_t - \hat{\mathbf{Y}}(t|t-1)\Big)\Big(\mathbf{Y}_t - \hat{\mathbf{Y}}(t|t-1)\Big)^\top\right]}_{=:\Delta_t = Cov-Matrix\ des\ Vorhersage-Fehlers\ von\ \mathbf{Y}_t}^{-}\end{aligned}$$

den Vorhersagefehler gewichtet und *Kalman Gain Matrix* genannt wird. Dabei bezeichne, wie in Bem 3.6.9, A^- eine verallgemeinerte Inverse der Matrix A. Die Kovarianz-Matrix Δ_t des Vorhersagefehlers für $\mathbf{Y_t}$ ist gegeben durch

$$\begin{aligned}\Delta_t &= \mathrm{E}\left[\Big(\mathbf{Y}_t - \hat{\mathbf{Y}}(t|t-1)\Big)\Big(\mathbf{Y}_t - \hat{\mathbf{Y}}(t|t-1)\Big)^\top\right]\\ &= \mathrm{E}\left[\{H_t\Big(\mathbf{X}_t - \hat{\mathbf{X}}(t|t-1)\Big) + \epsilon_t\} \cdot \{\Big(\mathbf{X}_t - \hat{\mathbf{X}}(t|t-1)\Big)^\top H_t^\top + \epsilon_t^\top\}\right]\\ &= H_t\Omega_t H_t^\top + \Sigma_t,\end{aligned}$$

mit $\Sigma_t = \mathrm{Cov}(\epsilon_t)$. Die Kovarianzmatrix Ω_t des Vorhersagefehlers für $\mathbf{X_t}$ erhält man durch die initialen Bedingungen

$$\Pi_{t_0} := \mathrm{E}[\mathbf{X_{t_0}}\mathbf{X_{t_0}}^\top],\ \Psi_{t_0} := \mathrm{E}[\hat{\mathbf{X}}(t_0|0)\hat{\mathbf{X}}(t_0|0)^\top],\ \Omega_{t_0} := \Pi_{t_0} - \Psi_{t_0}$$

und die Rekursionen

$$\begin{aligned}\Pi_{t+1} &= A_t\Pi_t A_t^\top + \Xi_t,\ \Xi_t = \mathrm{Cov}(\mathbf{Z_t}),\\ \Psi_{t+1} &= A_t\Psi_t A_t^\top + (A_t\Omega_t H_t^\top)\Delta_t^-(A_t\Omega_t H_t^\top)^\top,\\ \Omega_{t+1} &= \Pi_{t+1} - \Psi_{t+1}.\end{aligned}$$

Mit (3.64) kann nach Beobachtung von $\mathbf{y}_{t+1}$ das Vorgehen wiederholt werden, um den Kleinste-Quadrate-Schätzer für $\mathbf{X}_{t+1}$ zu erhalten ...

3.6.10 Bemerkung: Es sei $((\mathbf{X}_s), (\mathbf{Y}_s))$ ein lineares dynamisches Modell wie in Definition 3.6.8 beschrieben, $s \leq s_0$. Weiter bezeichne

$$\hat{\mathbf{Y}}(s|s-1) = H_s \hat{\mathbf{X}}(s|s-1) = H_s A_s \hat{\mathbf{X}}(s-1|s-1)$$

den linearen Schätzer von $\mathbf{Y}_s$ aufgrund der Daten $\ldots, \mathbf{y}_{s-2}, \mathbf{y}_{s-1}$ und

$$\mathbf{U}_s := \mathbf{y}_s - \hat{\mathbf{Y}}(s|s-1)$$

den Vorhersagefehler von $\mathbf{Y}_s$ (One-Step Prediction Error). Dann ist

$$\begin{aligned}
\hat{\mathbf{X}}(s|s) &= \hat{\mathbf{X}}(s|s-1) + K_s \mathbf{U}_s \\
&= A_s \hat{\mathbf{X}}(s-1|s-2) + A_s K_{s-1} \mathbf{U}_{s-1} + K_s \mathbf{U}_s \\
&\vdots \\
&= \cdots + A_s A_{s-1} K_{s-2} \mathbf{U}_{s-2} + A_s K_{s-1} \mathbf{U}_{s-1} + K_s \mathbf{U}_s.
\end{aligned}$$

Es existieren also Konstanten $\psi_0, \psi_1, \psi_2, \ldots \in \mathbb{R}^{n \times m}$ mit

$$\hat{\mathbf{X}}(s|s) = \psi_0 \mathbf{U}_s + \psi_1 \mathbf{U}_{s-1} + \psi_2 \mathbf{U}_{s-2} + \ldots$$

Somit ist der Kalman-Filter ein linearer Filter mit dem One-Step Prediction Error als eingehendes Signal.

3.6.11 Bemerkung: Der Kalman-Filter für Prozesse mit stetigem Zeit-Parameter ist auf der umfangreichen Theorie stochastischer Differentialgleichungen aufgebaut. Der interessierte Leser wird hierfür auf NEUBURGER (1972) verwiesen.

Kapitel 4. Fourier-Theorie stationärer Prozesse

4.1 Die Spektraldichte eines stationären Prozesses

Die Spektralanalyse stellt ein schlagkräftiges Instrument zur Analyse stationärer Prozesse dar. Sie kann z.B. bei der Klassifizierung von Zeitreihen behilflich sein oder auch deren zyklische Komponenten ermitteln.

4.1.1 Spektraldichte und Spektral-Verteilungsfunktion

Für einen stationären Prozess $(X_t)_{t\in T}$ mit identisch verschwindender Erwartungswertfunktion $\mu_t := \mathrm{E}(X_t)$ und Autokovarianzfunktion γ mit

$$\int_{-\infty}^{\infty} |\gamma(t)|dt < \infty, \tag{4.1}$$

wird die *(nicht-normierte) Spektraldichte von* (X_t) definiert durch

$$g(\omega) := \frac{1}{2\pi}\int_{-\infty}^{\infty} \gamma(t)e^{-i\omega t}dt. \tag{4.2}$$

Das bedeutet, die (nicht-normierte) Spektraldichte eines stochastischen Prozesses ist die Fourier-Transformierte seiner Autokovarianzfunktion.

Umgekehrt kann die Autokovarianzfunktion gemäß (1.17) dargestellt werden durch

$$\gamma(t) = \int_{-\infty}^{\infty} g(\omega)e^{i\omega t}d\omega. \tag{4.3}$$

Aus (4.3) ergibt sich sofort für die Varianz $\sigma_X^2 = \mathrm{Var}(X_0)$:

$$\sigma_X^2 = \gamma(0) = \int_{-\infty}^{\infty} g(\omega)d\omega.$$

Ist (4.1) erfüllt und $\gamma(0) = \mathrm{Var}(X_0) \neq 0$, so wird die *normierte Spektraldichte von* (X_t) definiert durch

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \rho(t)e^{-i\omega t}dt = \frac{g(\omega)}{\gamma(0)} = \frac{g(\omega)}{\sigma_X^2}.$$

Ist (X_t) ein reellwertiger, stationärer Prozess, so sind Autokovarianzfunktion und Autokorrelationsfunktion gerade, denn

$$\gamma(-t) = \mathrm{Cov}(X_{-t}, X_0) = \mathrm{Cov}(X_0, X_t) = \gamma(t).$$

Somit ist die Spektraldichte reellwertig (vgl. Bem. 1.2.1) und die normierte Spektraldichte besitzt die folgenden grundlegenden Eigenschaften:

(*i*) f ist gerade, d.h. $f(-\omega) = f(\omega),\ \omega \in \mathbb{R}$,

(*ii*) $f(\omega) \geq 0$,

(*iii*) $\int_{-\infty}^{\infty} f(\omega)d\omega = 1$.

Damit ist f eine Lebesgue-Wahrscheinlichkeitsdichte.

Falls (4.1) erfüllt ist und $\gamma(0) \neq 0$, kann auch die Autokorrelationsfunktion mittels der normierten Spektraldichte dargestellt werden (*Spektraldarstellung* der Autokorrelationsfunktion):

$$\rho(h) = \int_{-\infty}^{\infty} f(\omega)e^{i\omega h}d\omega.$$

Somit ist ρ die *inverse Fourier-Transformierte* von f.

4.1.1 Definition: Ist (X_t) ein reellwertiger stationärer Prozess mit (4.1), so heißt die Funktion

$$F(\omega) := \int_{-\infty}^{\omega} f(\lambda)d\lambda,\ \omega \in \mathbb{R},$$

Spektral-Verteilungsfunktion oder *integriertes (normiertes) Spektrum.*

Die Spektral-Verteilungsfunktion besitzt die Eigenschaften

(i) $0 \leq F(\omega) \leq 1$ für alle $\omega \in \mathbb{R}$,

(ii) $F(-\infty) = 0$ und $F(+\infty) = 1$,

(iii) F ist monoton nicht-fallend und stetig.

F ist also eine Wahrscheinlichkeits-Verteilungsfunktion.

4.1.2 Bemerkung: Die Definitionen von Fourier-Transformierte, inverser FT und Spektral-Verteilungsfunktion machen deutlich, dass die Autokovarianz- bzw. die Autokorrelationsfunktion einerseits und die Fourier-Transformierte, inverse FT sowie Spektral-Verteilungsfunktion andererseits exakt dieselben Informationen über den Prozess (X_t) beinhalten. Genauer gesagt besitzen zwei stationäre Zeitreihen mit verschwindender Erwartungswertfunktion genau dann dieselbe Fourier-Transformierte, wenn sie dieselbe Autokovarianzfunktion besitzen (vgl. Brockwell / Davis (1991)).

4.1.3 Bemerkung: Die charakteristische Funktion φ einer Zufallsvariablen mit Lebesgue-Dichte f ist laut Definition

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)dx$$

die inverse Fourier-Transformierte der Dichte f. Die inverse Fourier-Transformierte der Spektraldichte eines stochastischen Prozesses ist dessen Autokorrelationsfunktion.

Ebenso, wie in der Klasse der der Wahrscheinlichkeits-Verteilungsfunktionen nicht jede Verteilung eine Dichte besitzt, gibt es auch stochastische Prozesse, die keine Spektraldichte besitzen. Wie bei diskreten Verteilungen, die an höchstens abzählbar vielen Stellen Masse besitzen, gibt es auch Prozesse, die dem Einfluss höchstens abzählbar vieler Frequenzkomponenten ausgesetzt sind (*harmonische Prozesse*, vgl. Priestley (1981[1])). Man spricht dann von einem *diskreten Spektrum*. Schließlich gibt es analog zu Verteilungen, die man aus einer diskreten Verteilung und einer Verteilung mit Dichte zusammensetzen kann, auch hier Prozesse mit gemischtem Spektrum. Die Definition 4.1.1 muss daher verallgemeinert werden. Wir tun dies indirekt durch das unten folgende, wichtige Wiener-Khintchine Theorem. Bevor dieser Satz formuliert wird, sollen jedoch die Begriffe *stochastische Stetigkeit* und *Riemann-Stieltjes-Integral* definiert werden:

4.1.4 Definition: Ein stochastischer Prozess (X_t) heißt *stochastisch stetig (im quadratischen Mittel)* an der Stelle t_0, genau wenn

$$\lim_{t \to t_0} \mathrm{E}\left[(X_t - X_{t_0})^2\right] = 0,$$

also im Falle

$$X_{t_0} = \text{l.i.m.}_{t \to t_0}\ X_t \text{ in } L^2(P) \text{ (vgl. Anhang B).}$$

Ist (X_t) stochastisch stetig an jeder Stelle $t \in T$, so sagt man, (X_t) sei stochastisch stetig.

4.1.5 Lemma: *Ein stationärer Prozess mit Varianz $\gamma(0)$ ist genau dann stochastisch stetig, wenn seine Autokorrelationsfunktion ρ stetig ist an der Stelle $t = 0$, d.h. wenn $\lim_{t \to 0} \rho(t) = 1$.*

Beweis: Aufgrund der Stationarität von (X_t) genügt es, die Behauptung für ein $t_0 \in T$ zu zeigen. Der Prozess (X_t) sei also stationär und stochastisch stetig an der Stelle t_0. Wegen

$$\begin{aligned} \mathrm{E}\left[(X_t - X_{t_0})^2\right] &= \mathrm{Var}(X_t) + \mathrm{Var}(X_{t_0}) - 2\mathrm{Cov}(X_t, X_{t_0}) \\ &= 2\gamma(0) \cdot (1 - \rho(t - t_0)) \end{aligned}$$

ist (X_t) genau dann stochastisch stetig an der Stelle t_0, wenn $\lim_{t \to t_0} \rho(t - t_0) = \lim_{s \to 0} \rho(s) = 1$. □

4.1.6 Definition: Es seien f und g zwei reellwertige Funktionen auf $[a,b]$ und $Z_n := \{x_0, x_1, \ldots, x_n\}$ eine Zerlegung von $[a,b]$ mit zugehörigem Zwischenvektor $\xi^{(n)} := (\xi_1, \ldots, \xi_n)$. Dann heißt

$$S(Z_n, \xi^{(n)}) := \sum_{k=1}^{n} f(\xi_k)[g(x_k) - g(x_{k-1})] \tag{4.4}$$

eine *Riemann-Stieltjes-Summe* für f bezüglich g. Ist (Z_n) eine Zerlegungsnullfolge mit $\max_{i=1,\ldots n} |x_i - x_{i-1}| \to 0$, so heißt $(S(Z_n, \xi^{(n)}))$ eine *Riemann-Stieltjes-Folge.* Strebt jede Riemann-Stieltjes-Folge (also unabhängig von der Zerlegung und dem Zwischenvektor) gegen denselben Grenzwert, so sagt man, f sei auf $[a,b]$ bezüglich g *Riemann-Stieltjes-integrierbar.* Den Grenzwert nennt man *Riemann-Stieltjes-Integral (RS-Integral)* und schreibt

$$\int_a^b f(x)dg(x).$$

4.1.7 Beispiel: Im Fall $g(x) = x \ \forall x$ ist das Riemann-Stieltjes-Integral identisch mit dem herkömmlichen Riemann-Integral. Ist g eine Treppenfunktion auf $[a,b]$ mit Sprüngen der Größe $g_1, \ldots, g_m$ an den Stellen $x_1, \ldots, x_m$, dann ist für jede stetige Funktion f das Integral $\int_a^b f(x)dg(x) = \sum_{k=1}^{m} f(x_k) \cdot g_k$.

4.1.8 Lemma: *Eine hinreichende Bedingung für die Existenz des RS-Integrals $\int_a^b f(x)dg(x)$ ist*

- *f ist stetig auf $[a,b]$ und*
- *g ist von beschränkter Variation auf $[a,b]$.*

Existiert $\int_a^b f(x)dg(x)$, so existiert auch $\int_a^b g(x)df(x)$ und es ist

$$\int_a^b g(x)df(x) = f(b)g(b) - f(a)g(a) - \int_a^b f(x)dg(x).$$

Beweis: Heuser (1986[1]).

Für Riemann-Stieltjes-Integrale gelten analoge Rechenregeln wie für herkömmliche Riemann-Integrale, näheres hierzu s. Heuser (1986[1]). Insbesondere gilt

4.1.9 Lemma: *Ist f auf $[a,c]$ Riemann-Stieltjes-integrierbar bezüglich g, so ist f für jeden Punkt $b \in [a,c]$ RS-integrierbar auf $[a,b]$ und $[b,c]$ und es ist*

$$\int_a^c f(x)dg(x) = \int_a^b f(x)dg(x) + \int_b^c f(x)dg(x).$$

Beweis: Heuser (1986[1]).

Außerdem gilt

4.1.10 Lemma: *(a) Ist f RS-integrierbar bzgl. $g_1 + g_2$, so ist f RS-integrierbar bzgl. g_1 und g_2 und es ist $\int_a^b f(x)d(g_1+g_2)(x) = \int_a^b f(x)dg_1(x) + \int_a^b f(x)dg_2(x)$.*

(b) Ist g differenzierbar und $f \cdot g'$ Riemann-integrierbar, so gilt $\int_a^b f(x)dg(x) = \int_a^b f(x)g'(x)dx$.

Teil (a) wird sofort klar bei Betrachtung der Riemann-Stieltjes-Summe (4.4). Den Beweis von (b) findet man in Heuser (1986[1]).

Nun der bereits angekündigte Satz.

4.1.11 Satz: (Wiener-Khintchine Theorem) *Die Funktion $\rho(t),\ t \in \mathbb{R}$, ist genau dann die Autokorrelationsfunktion eines stationären, stochastisch stetigen Prozesses (X_t), wenn eine rechtsseitig stetige, nicht fallende Funktion F auf $\mathbb{R}$ existiert, so dass*

$$\begin{aligned} F(-\infty) &= 0 \text{ und } F(\infty) = 1, \text{ sowie} \\ \rho(t) &= \int_{-\infty}^{\infty} e^{i\omega t} dF(\omega) \text{ für alle } t \in \mathbb{R}. \end{aligned} \tag{4.5}$$

In Verallgemeinerung von Definition 4.1.1 nennt man F *Spektral-Verteilungsfunktion* von (X_t). Für den Beweis des Satzes wird der Leser auf Priestley (1981[1]) verwiesen.

Bisher wurden ausschließlich Prozesse mit stetigem Zeitparameter betrachtet. Nun folgen die entsprechenden Definitionen und Aussagen für Zeitreihen mit $T = \mathbb{Z}$.

4.1.12 Definition: Gegeben sei eine stationäre Zeitreihe (X_t) mit $\mathrm{E}(X_t) = 0 \; \forall \; t \in \mathbb{Z}$ und Kovarianzfunktion γ so, dass gilt

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty. \tag{4.6}$$

Dann definiert man die nicht-normierte Spektraldichte durch

$$g(\omega) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}. \tag{4.7}$$

Im Falle (4.6) und $\gamma(0) \neq 0$ ist die normierte Spektraldichte einer stationären Zeitreihe definiert als die Fourier-Transformierte ihrer Autokorrelationsfunktion ρ. Somit ist f gegeben durch

$$f(\omega) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{-i\omega h} = \frac{g(\omega)}{\gamma(0)}. \tag{4.8}$$

Der nun folgende Satz von Wold (Satz von Herglotz) ist ein Analogon für Zeitreihen zum Wiener-Khintchine-Theorem.

4.1.13 Satz: (Wold's Theorem / Herglotz's Theorem) *Die Folge $\rho(h),\ h \in \mathbb{Z}$, ist genau dann die Autokorrelationsfunktion einer stationären Zeitreihe $(X_t),\ t \in \mathbb{Z}$, falls eine rechtsseitig stetige, nicht fallende Funktion F auf $\mathbb{R}$ existiert, so dass*

$$\begin{aligned} F(-\pi) &= 0 \text{ und } F(\pi) = 1, \quad \text{sowie} \\ \rho(h) &= \int_{-\pi}^{\pi} e^{i\omega h} dF(\omega) \text{ für alle } h \in \mathbb{Z}. \end{aligned} \tag{4.9}$$

Beweis: Brockwell / Davis (1991).

4.1.14 Bemerkung: Ist (X_t) so beschaffen, dass F überall differenzierbar ist, so existiert die normierte Spektraldichte f und es gilt

$$f(\omega) = \frac{dF(\omega)}{d\omega}$$

sowie

$$\rho(h) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega h} d\omega.$$

In diesem Falle ist auch

$$\gamma(h) = \int_{-\pi}^{\pi} g(\omega) e^{i\omega h} d\omega.$$

Sämtliche oben genannten Eigenschaften der Spektraldichte eines Prozesses mit stetigem Zeit-Parameter gelten entsprechend auch für die Spektraldichte einer Zeitreihe. Wird über den Frequenzbereich integriert, so verlaufen die Integrationsgrenzen für ω jedoch von $-\pi$ bis π, bei Integration über den Zeitparameter-Bereich ist das Integral stets durch eine Summe zu ersetzen.

4.1.2 Der Aliasing-Effekt

Naturgemäß kommen bei Zeitreihen sehr hohe Frequenzen nicht vor. Die Definition der Spektraldichte einer Zeitreihe ist also für Frequenzen außerhalb eines beschränkten Intervalles nicht sinnvoll. Wegen

$$e^{2hi\pi} = e^{-2hi\pi} = 1 \text{ für alle } h \in \mathbb{Z} \tag{4.10}$$

gilt für die Spektraldichte g einer Zeitreihe mit $T = \mathbb{Z}$

$$g(\omega + 2\pi) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \left(e^{-i\omega h} \cdot e^{-i2\pi h}\right) = g(\omega),$$

d.h. g ist 2π-periodisch. Eine Zeitreihe mit $T = \mathbb{Z}$ enthält also sämtliche spektralen Informationen im Intervall $[-\pi, \pi]$. Man definiert daher, ohne Informationsverlust, die Spektraldichte einer solchen Zeitreihe auf dem Intervall $[-\pi, \pi]$.

Für die Spektraldichte eines Prozesses in stetiger Zeit gilt die o.g. Aussage nicht, denn Gleichung (4.10) gilt nur für $h \in \mathbb{Z}$. Wie verhält es sich nun, wenn ein stochastischer Prozess $(X_t)_{t\in\mathbb{R}}$ in stetiger Zeit lediglich an diskreten Zeitpunkten beobachtet wird? Was geschieht mit den hochfrequenten Anteilen des stochastischen Prozesses, wenn die Spektraldichte des diskreten Beobachtungsprozesses außerhalb eines beschränkten Intervalles nicht definiert ist? Zur Illustration betrachte man eine hochfrequente Welle, die zu äquidistanten Zeitpunkten $t_0, t_0 + \Delta, t_0 + 2\Delta, \ldots$ beobachtet wird. Dabei bezeichne Δ den Abstand der äquidistanten Beobachtungspunkte. In Abbildung 4.1 ist $t_0 = 0$ und $\Delta = \pi$, die hochfrequente Welle ist als durchgezogene Linie eingezeichnet. In diesem synthetischen Beispiel werden zu den Zeitpunkten $0, \pi, 2\pi, \ldots, 7\pi$ die Werte $1, 0, -1, 0, 1, 0, -1, 0$ beobachtet.

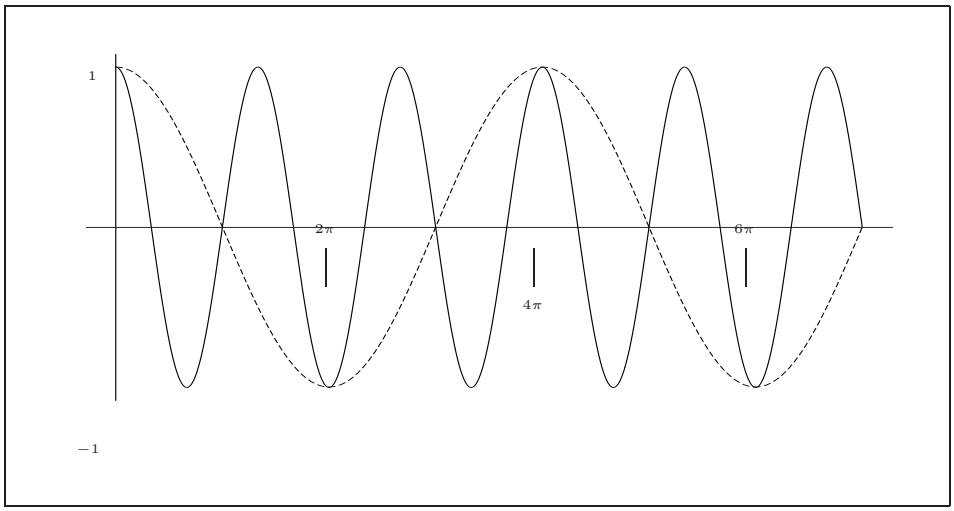

Abbildung 4.1: Der Aliasing-Effekt

Diese beobachteten Werte können jedoch auch durch eine niederfrequentere Welle erzeugt werden, die in Abbildung 4.1 gestrichelt eingezeichnet ist. Wird aufgrund der Beobachtungen zu den Zeitpunkten $t_0, t_0 + \Delta, t_0 + 2\Delta, \ldots$ eine Spektralanalyse durchgeführt, so wird die niederfrequente, nicht jedoch die tatsächliche Frequenzkomponente aufgespürt. Dieser Effekt wird *Aliasing-Effekt* genannt.

4.1.15 Bemerkung: Ist $(X_t),\ t \in \mathbb{R}$, ein stochastischer Prozess mit stetigem Zeitparameter, der zu diskreten Zeitpunkten $k \cdot \Delta$, $k \in \mathbb{Z}$, $\Delta \in \mathbb{R}$, beobachtet wird, so enthält der diskrete Beobachtungsprozess $(X_{k\cdot\Delta}),\ k \in \mathbb{Z}$, sämtliche spektralen Informationen im Intervall $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$, s. Priestley (1981[1]).

4.1.16 Bemerkung: Die Bezeichnung "Aliasing-Effekt" rührt daher, dass jede Frequenzkomponente $\omega \notin [-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ bei einer Spektralanalyse als Frequenz **in** $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ interpretiert wird. Sie hat also einen "Alias" im Intervall $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$.

Es sei nun (Y_t) der durch $Y_t := X_{t\cdot\Delta}\ ,\ t \in \mathbb{Z}$, definierte **diskrete**, also der beobachtete Prozess. Wie sieht nun die Spektraldichte von (Y_t) aus? Wie Abbildung 4.2 zeigt, wird die Spektraldichte von (X_t) an den Rändern

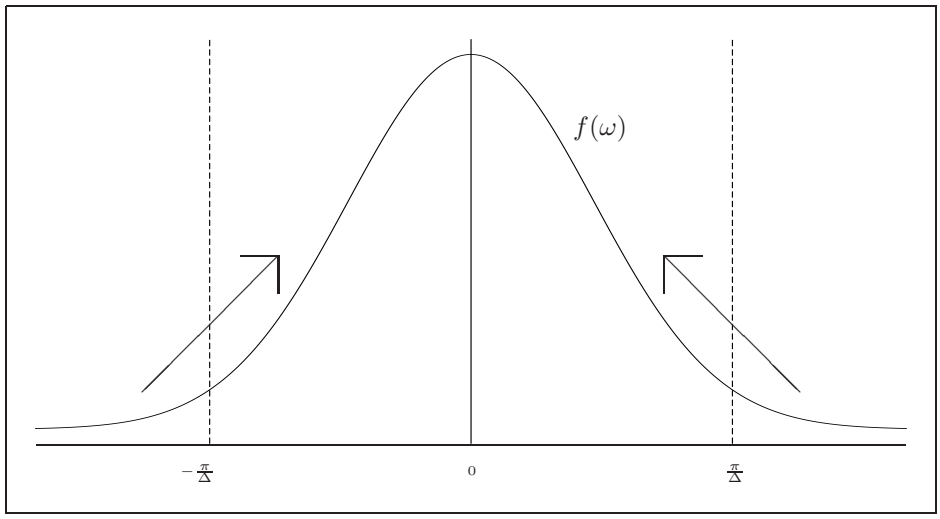

Abbildung 4.2: Durch den Aliasing-Effekt wird die Spektraldichte "gefaltet"

des Intervalles $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ "gefaltet". Dabei überlagern sich die außerhalb gelegenen Frequenzkomponenten mit jenen des Intervallinneren additiv. Die Spektraldichte des diskreten Prozesses (Y_t) ergibt sich, wenn man die sich überlagernden Komponenten addiert und durch den Abstand der Beobachtungs-Zeitpunkte dividiert.

Wegen der Symmetrie (man beachte, dass auch die nicht-normierte Spektraldichte die Eigenschaft (i) der normierten Spektraldichte besitzt) liegt somit die ganze Information über das Spektrum des Prozesses im Intervall $[0, \frac{\pi}{\Delta}]$. Die Frequenz $\omega = \frac{2\pi}{\Delta}$ heißt *Faltungs-Frequenz.*

Hat man also die Spektraldichte eines in diskreter Zeit beobachteten, stetigen Prozesses geschätzt (vergleiche Kapitel 4.2) und hat die geschätzte Spektraldichte einen Peak bei einer Frequenz $\omega_0 \in [0, \frac{\pi}{\Delta}]$, so bedeutet dies, dass eine oder mehrere der Frequenzen $\omega_0 \pm 2k\frac{\pi}{\Delta}$, $-\omega_0 \pm 2k\frac{\pi}{\Delta}$ $(k \in \mathbb{Z})$ zum Spektrum des Prozesses beitragen. **Welche** dieser Frequenzen einen Beitrag liefern, ist anhand der Spektraldichte **nicht** erkennbar. Diese Frage muss mit anderen stochastischen Methoden gelöst werden.

Zur Abschwächung des Aliasing-Effekts kann der Zeitabstand Δ möglichst klein gewählt werden, so dass das Intervall $[-\frac{\pi}{\Delta}, \frac{\pi}{\Delta}]$ möglichst groß wird.

Gänzlich vermeiden kann man diesen Effekt, wenn man die Beobachtungs-Zeitpunkte, falls möglich, zufällig gemäß eines *Poisson-Prozesses* (N_t) wählt. Salopp gesagt, beschreibt ein Poisson-Prozess (N_t) das wiederholte, **voneinander unabhängige** Eintreten eines bestimmten Ereignisses im zeitlichen Verlauf. Zu den Ereignissen, deren Eintreten durch einen Poisson-Prozess beschrieben werden können, gehören z.B. die (voneinander unabhängige) Ankunft von Fahrzeugen an einer Kreuzung, die (voneinander unabhängige) Ankunft von Personen an einem Schalter sowie die Emission von α-Teilchen von einer radioaktiven Substanz. Dabei ist N_t die Anzahl der Ereignisse im Intervall $[0, t]$, also bis zur Zeit t. Treten im Durchschnitt λ Ereignisse im Intervall $[0, 1]$ ein, so besitzt N_t für festes t eine Poisson-Verteilung mit Parameter λt (in Zeichen $\mathrm{Po}(\lambda t)$), d.h.

$$\mathrm{P}(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!},\ k = 0, 1, \ldots$$

(N_t) heißt dann ein *Poisson-Prozess mit Parameter* λ. Die Umgehung des Aliasing-Effektes durch zufällige Wahl der Beobachtungs-Zeitpunkte wird einsichtig, wenn man sich vergegenwärtigt, dass der Übergang zu einer diskreten, äquidistanten Zeitparameter-Menge in der Regel einen **systematischen** Informationsverlust bedeutet. Wählt man die Beobachtungs-Zeitpunkte zufällig, so ist der Informationsverlust unsystematisch. Zum Beispiel wählt der Poisson-Prozess gelegentlich auch nahe beieinander liegende Beobachtungs-Zeitpunkte, wodurch der Beobachter **direkte** Informationen über hohe Frequenzen erhält (anstatt der indirekten Aliase).

4.1.3 Spektraldichten spezieller stochastischer Prozesse

Am Ende dieses Abschnitts sollen die Spektraldichten einiger spezieller stochastischer Prozesse beschrieben werden.

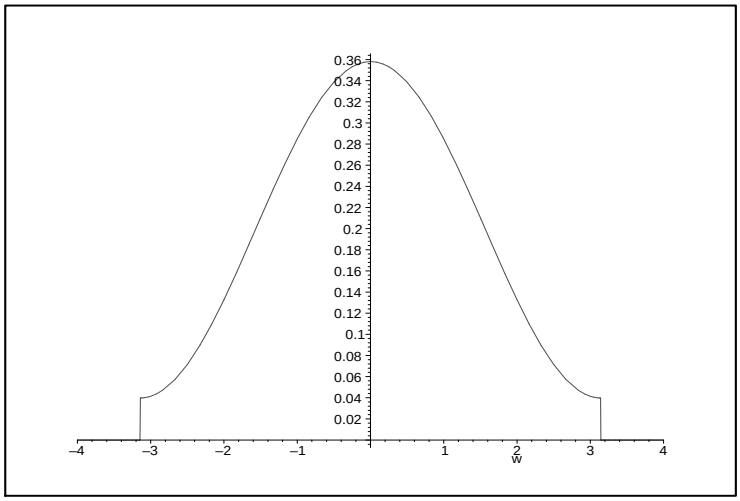

Abbildung 4.3: Spektraldichte eines diskreten $AR(1)$-Prozesses ($\phi_1 = \frac{1}{2},\ \sigma^2 = 1$)

4.1.17 Bemerkung: Ein White Noise Prozess $(W_t),\ t \in \mathbb{Z}$, zeichnet sich gerade durch seine "absolute Zufälligkeit" aus. Er besitzt die Autokorrelationsfunktion

$$\rho(h) = \begin{cases} 1 & ,\ h = 0, \\ 0 & ,\ h \in \mathbb{Z}\backslash\{0\}, \end{cases}$$

und somit eine normierte Spektraldichte, welche auf $[-\pi, \pi]$ konstant den Wert $\frac{1}{2\pi}$ und sonst 0 annimmt. Das bedeutet, dass jede Frequenz-Komponente aus $[-\pi, \pi]$ gleich viel zu (W_t) beiträgt. Dies ist vergleichbar mit "weißem" Licht, zu dem jede Farb-Komponente (also Frequenz-Komponente) des sichtbaren Lichtes den gleichen Anteil beiträgt. Auf der Zeitachse führt dies zu einer additiven Überlagerung von Wellen der unterschiedlichen Wellenlängen, im Frequenzbereich zu einem konstanten Spektrum. Diese Parallele zwischen Lichtspektrum und Spektraldichte von Zeitreihen gab dem "White" Noise Prozess seinen Namen.

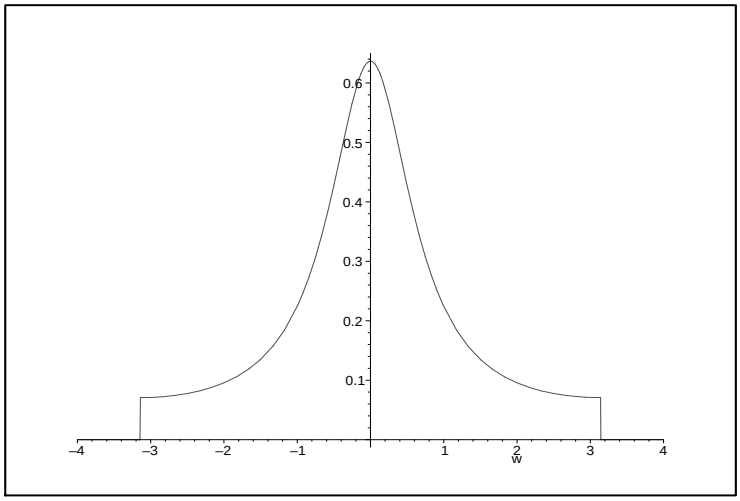

Abbildung 4.4: Spektraldichte eines diskreten $MA(1)$-Prozesses ($\theta_1 = \frac{1}{2},\ \sigma^2 = 1$)

Über die Spektraldichte eines $ARMA(p, q)$-Prozesses gibt der folgende Satz Auskunft:

4.1.18 Satz: *Es sei (X_t) ein $ARMA(p, q)$-Prozess mit*

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},\ Z_t \sim WN(0, \sigma^2).$$

Besitzen die Polynome

$$\begin{aligned} \phi(z) &:= 1 - \phi_1 z - \cdots - \phi_p z^p \\ \text{und } \theta(z) &:= 1 + \theta_1 z + \cdots + \theta_q z^q \end{aligned}$$

keine gemeinsamen Nullstellen und hat $\phi(z)$ keine Nullstellen im Einheitskreis, so besitzt (X_t) die Spektraldichte

$$f(\omega) = \frac{\sigma^2}{2\pi} \cdot \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2},\ \omega \in [-\pi, \pi].$$

Beweis: Brockwell / Davis (1991).

Die Abbildungen 4.3 bis 4.5 zeigen die Spektraldichten eines $AR(1)$- sowie eines $MA(1)$-Prozesses in diskreter Zeit beziehungsweise eines $AR(2)$-Prozesses mit stetigem Zeitparameter.

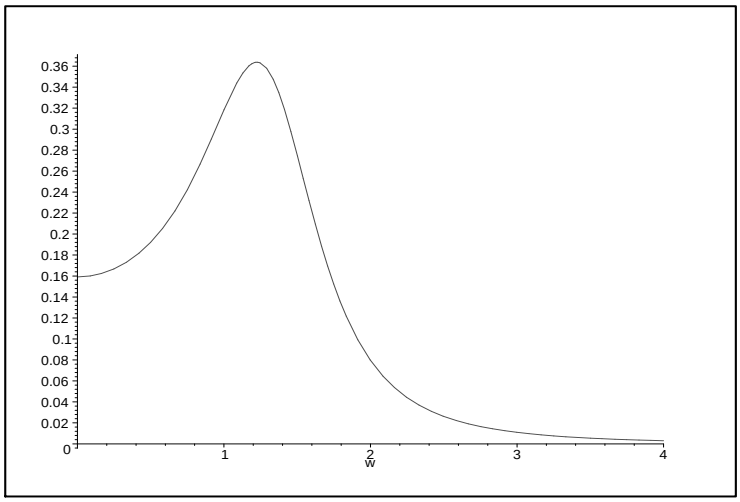

Abbildung 4.5: Spektraldichte eines $AR(2)$-Prozesses

Ist (X_t) ein Lomg Memory Prozess, so existiert ein $\beta \in (0, 1)$ und eine Konstante $c_g > 0$, so dass für die nicht-normierte Spektraldichte g des Prozesses gilt

$$\frac{g(\omega)}{|\omega|^{-\beta}} \xrightarrow{\omega \to 0} c_g,$$

d.h., nahe $\omega = 0$ verhält sich die nicht-normierte Spektraldichte ähnlich einer Funktion mit einem Pol der Form $\frac{c_g}{|\omega|^\beta}$.

4.2 Schätzer für die nicht-normierte Spektraldichte

4.2.1 Das Periodogramm

Aufgrund der Darstellung (4.7) für die nicht-normierte Spektraldichte liegt es nahe, g mit Hilfe der empirischen Autokovarianzfunktion $\hat{\gamma}$ (vgl. (3.7)) zu schätzen. Man nennt

$$I_N^*(\omega) := \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) e^{-i\omega h} \tag{4.11}$$

das *(modifizierte) Periodogramm.* (Als *Periodogramm* bezeichnet man die Summe in (4.11) ohne den Vorfaktor $\frac{1}{2\pi}$.) Wegen $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ ist

$$\begin{aligned} I_N^*(\omega) &= \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) \left(\cos(-\omega h) + i\sin(-\omega h)\right) \\ &\overset{\hat{\gamma} \text{ gerade}}{=} \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h)\cos(\omega h) \\ &= \frac{1}{2\pi}\hat{\gamma}(0) + \frac{1}{\pi}\sum_{h=1}^{N-1} \hat{\gamma}(h)\cos(\omega h). \end{aligned}$$

Ist (X_t) der lineare Prozess

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j},\ (Z_t) \sim \mathit{IID}(0, \sigma^2), \tag{4.12}$$

und γ die Autokovarianzfunktion von (X_t) mit

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{4.13}$$

so ist das Periodogramm (als natürlicher Schätzer für g) asymptotisch erwartungstreu, d.h.

$$\mathrm{E}[I_N^*(\omega)] \overset{N\to\infty}{\longrightarrow} \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h} = g(\omega),$$

s. Brockwell / Davis (1991). Ein großer Nachteil des modifizierten Periodogramms ist jedoch, dass die Korrelation zwischen $I_N^*(\omega_1)$ und $I_N^*(\omega_2)$ für nahe beieinander gelegene Frequenzen ω_1 und ω_2 **nachlässt**, falls N erhöht wird:

4.2.1 Satz: *Es sei (X_t) der lineare Prozess (4.12) mit (4.13).*

a) Ist $g(\omega) > 0$ für alle $\omega \in [-\pi, \pi]$ und $0 < \omega_1 < \cdots < \omega_k < \pi$, dann konvergiert der Zufallsvektor $(I_N^*(\omega_1), \ldots, I_N^*(\omega_k))^\top$ nach Verteilung gegen einen Vektor unabhängiger Zufallsvariablen Y_i, $i = 1, \ldots, k$, mit $E(Y_i) = g(\omega_i)$.

b) Ist $\sum_{j=-\infty}^{\infty} |\psi_j||j|^{\frac{1}{2}} < \infty$ sowie $E(Z_1^4) < \infty$, so gilt für $\omega_i = \frac{2\pi i}{N} \in (0, \pi)$

$$Cov(I_N^*(\omega_i), I_N^*(\omega_j)) = \begin{cases} g^2(\omega_i) + O(\frac{1}{\sqrt{N}}), & \omega_i = \omega_j, \\ O(\frac{1}{N}), & \omega_i \neq \omega_j. \end{cases}$$

Die Schreibweise $a_n = O(\frac{1}{n})$ bedeutet, anschaulich gesprochen, dass das Wachstumsverhalten der Folge $(a_n)_{n\in\mathbb{N}}$ asymptotisch mit jenem der Folge $(\frac{1}{n})_{n\in\mathbb{N}}$ vergleichbar ist. Formal

$$a_n = O(\tfrac{1}{n})\ (n \to \infty) \ :\Longleftrightarrow\ \text{Es existiert eine Konstante } c \text{ mit } a_n \cdot n \to c\ (n \to \infty).$$

Ist $c = 0$, so schreibt man $a_n = o(\frac{1}{n})$.

Beweis Satz 4.2.1: Brockwell / Davis (1991).

4.2.2 Konsistente Schätzer für die Spektraldichte

Neben der (zumindest asymptotischen) Erwartungstreue ist die Konsistenz (s. Anhang B.1) eine Minimalanforderung an einen Schätzer. Wie steht es diesbezüglich mit dem Periodogramm? Ist das modifizierte Periodogramm ein konsistenter Schätzer für die nicht-normierte Spektraldichte?

Nach Satz 4.2.1 besitzt das Periodogramm die Eigenschaft, dass die Korrelationen zwischen zwei verschiedenen Frequenzkomponenten gegen null gehen ($N \to \infty$), während die Varianz nicht verschwinden kann. Dies hat zur Folge, dass das Periodogramm, auch für sehr große N, stark um die tatsächlichen Werte von g schwankt, wie Abbildung 4.6 veranschaulicht.

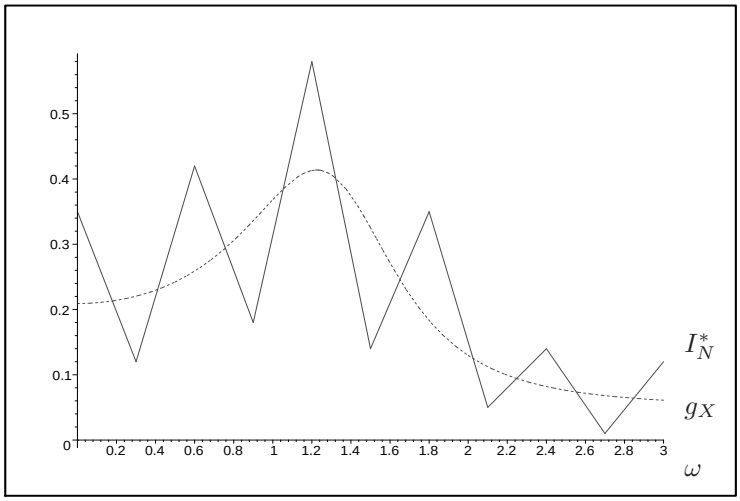

Abbildung 4.6: Das stark schwankende Verhalten des Periodogramms

Damit kann man mit Sicherheit ein $\epsilon > 0$ finden, so dass

$$P(|I_N^*(\omega) - g(\omega)| \geq \epsilon) \not\longrightarrow 0 \ (N \to \infty).$$

$I_N^*(\omega)$ ist also **kein** konsistenter Schätzer für $g(\omega)$.

Um einen konsistenten Schätzer zu erhalten ist es daher notwendig, I_N^* auf eine sinnvolle Weise zu glätten. Man erreicht dies durch Einführung einer Gewichtsfunktion λ auf dem Zeitparameter-Bereich (λ wird auch *Lag-Window* genannt) und setzt

$$\hat{g}(\omega) := \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \lambda(h)\hat{\gamma}(h)e^{-i\omega h}.$$

Die glättende Eigenschaft der Gewichtsfunktion λ wird deutlich, wenn man $\hat{g}$ als gewichtetes Integral über I_N^* betrachtet. Hierzu beachte man, dass analog zu (4.3) für die empirische Autokovarianzfunktion gilt

$$\hat{\gamma}(h) = \int_{-\pi}^{\pi} I_N^*(\theta)e^{i\theta h}d\theta.$$

Damit ist

$$\begin{aligned} \hat{g}(\omega) &= \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \lambda(h) \int_{-\pi}^{\pi} I_N^*(\theta)e^{i(\theta-\omega)h}d\theta \\ &= \int_{-\pi}^{\pi} I_N^*(\theta) \underbrace{\left(\frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \lambda(h)e^{-i(\omega-\theta)h} \right)}_{=:W(\omega-\theta)} d\theta. \end{aligned} \tag{4.14}$$

Man nennt W mit

$$W(\theta) = \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \lambda(h)e^{-i\theta h}$$

das zu λ korrespondierende *Frequenz-Window*. Das Frequenz-Window entspricht einer Gewichtsfunktion auf dem Frequenzbereich. In Gleichung (4.14) werden die Werte des Periodogramms entsprechend dieser Gewichtsfunktion gemittelt, d.h. $\hat{g}$ ist ein (gemäß dieser Gewichtsfunktion) geglättetes Periodogramm. Die Abbildungen 4.7 bis 4.9 zeigen einige Lag-Windows (links) mit den dazugehörigen Frequenz-Windows (rechts).

Das einfachste Lag-Window, das *truncated Periodogramm* (λ_1) verkürzt lediglich den relevanten Zeitparameterbereich auf ein Intervall $[-M, M]$, ohne eine weitere Gewichtung durchzuführen. Dabei ist $M \in \mathbb{N}$ geeignet zu wählen (man vergleiche hierzu die Ausführungen in PRIESTLEY (1981)). Das zum truncated Periodogramm korrespondierende Frequenz-Window (W_1) wird *Dirichlet-Kern* genannt.

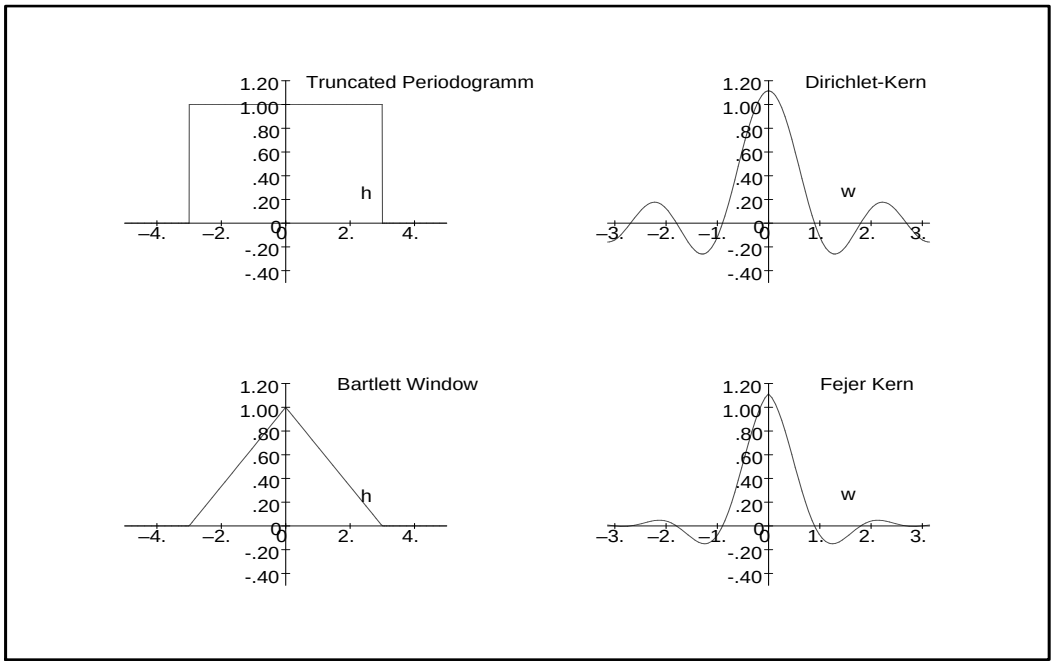

Abbildung 4.7: Truncated Periodogramm und Bartlett Window

Das Truncated Periodogramm bzw. der Dirichlet-Kern sind gegeben durch

$$\begin{aligned}
\lambda_1(h) &= \begin{cases} 1, & |h| \leq M, \\ 0, & |h| > M, \end{cases} \\
W_1(\theta) &= \frac{1}{2\pi} \sum_{m=-M}^{M} e^{-i\theta m} = \frac{1}{2\pi} \sum_{m=-M}^{M} (\cos(m\theta) + i \sin(m\theta)) = \frac{1}{2\pi} \sum_{m=-M}^{M} \cos(m\theta).
\end{aligned}$$

Das *Bartlett Window* (λ_2) gewichtet die zum Time Lag h näher gelegenen Zeitpunkte des Intervalls $[-M, M]$ stärker als die weiter entfernt gelegenen in einer linearen Weise. Das entsprechende Frequenz-Window (W_2) heißt *Fejer Kern*. Es ist

$$\begin{aligned}
\lambda_2(h) &= \begin{cases} 1 - \frac{|h|}{M}, & |h| \leq M, \\ 0, & |h| > M, \end{cases} \\
W_2(\theta) &= \frac{1}{2\pi} \sum_{m=-M}^{M} \left(1 - \frac{|m|}{M}\right) e^{-i\theta m} = \frac{1}{2\pi} \sum_{m=-M}^{M} \left(1 - \frac{|m|}{M}\right) \cos(m\theta).
\end{aligned}$$

Weitere wichtige Lag-Fenster sind das *Tukey-Hanning-Window* (λ_3)

$$\begin{aligned}
\lambda_3(h) &= \begin{cases} \frac{1}{2}\left(1 + \cos(\frac{h\pi}{M})\right), & |h| \leq M, \\ 0, & |h| > M, \end{cases} \\
W_3(\theta) &= \frac{1}{4} W_2\left(\theta - \frac{\pi}{M}\right) + \frac{1}{2} W_2\left(\theta\right) + \frac{1}{4} W_2\left(\theta + \frac{\pi}{M}\right),
\end{aligned}$$

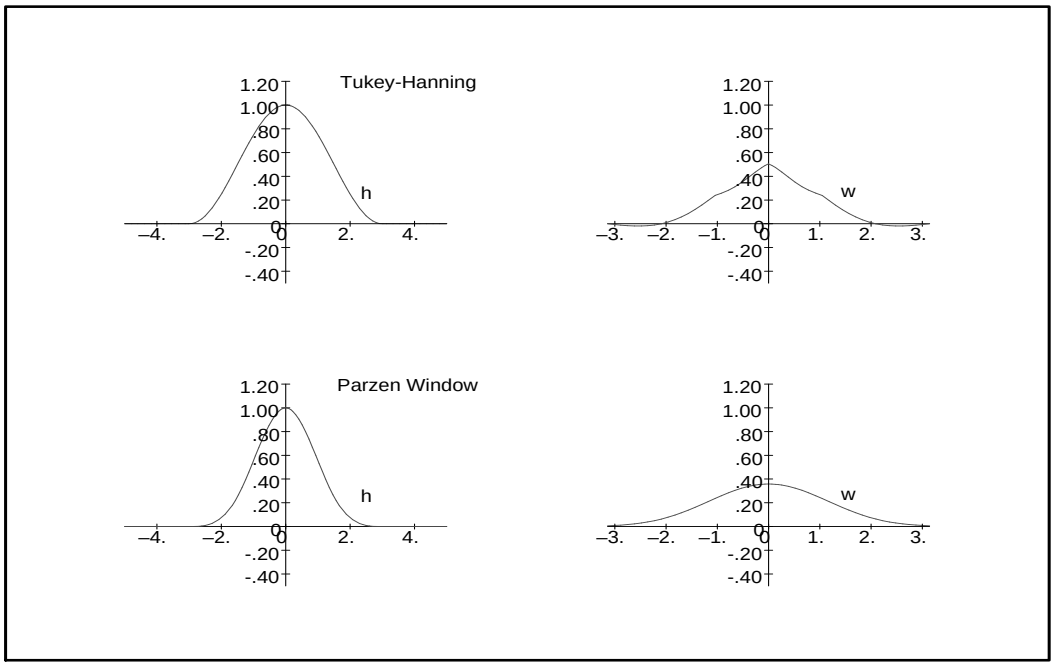

Abbildung 4.8: Tukey-Hanning und Parzen-Window

und das *Parzen Window* (λ_4)

$$\lambda_4(h) = \begin{cases} 1 - 6\left(\frac{h}{M}\right)^2 + 6\left(\frac{|h|}{M}\right)^3, & |h| \le \frac{M}{2}, \\ 2\left(1 - \frac{|h|}{M}\right)^3, & |h| \in [\frac{M}{2}, M], \\ 0, & |h| > M, \end{cases}$$

$$W_4(\theta) = \frac{3}{8\pi M^3}\left(\frac{\sin(\frac{M\theta}{4})}{\frac{1}{2}\sin(\frac{\theta}{2})}\right)^4 \cdot \left(1 - \frac{2}{3}\sin^2(\frac{\theta}{2})\right).$$

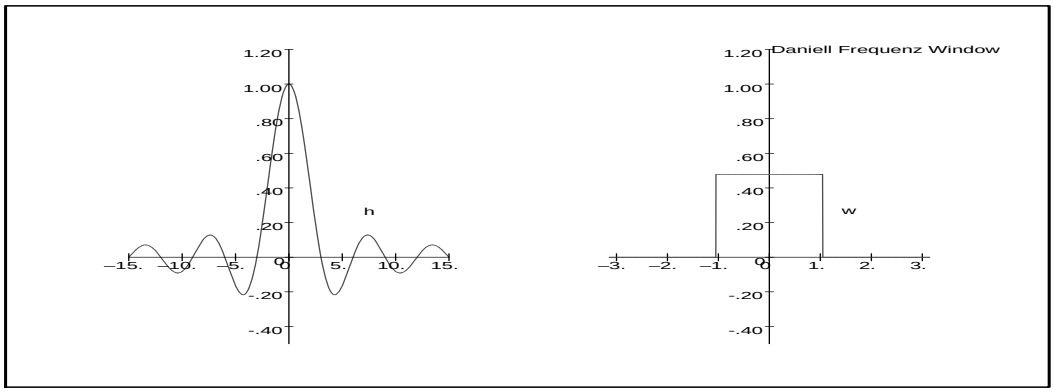

Abbildung 4.9: Daniell Frequenz Window

Von den oben aufgeführten Windows unterscheidet sich das *Daniell Frequenz Window* (W_5, vgl. Abb. 4.9)

dahingehend, dass es direkt den Frequenz-Bereich einschränkt. Es ist

$$\begin{aligned} W_5(\theta) &= \begin{cases} \frac{M}{2\pi}, & |\theta| \leq \frac{\pi}{M}, \\ 0, & |\theta| > \frac{\pi}{M}, \end{cases} \\ \lambda_5(h) &= \frac{\sin(\frac{\pi h}{M})}{\frac{\pi h}{M}}. \end{aligned}$$

In den Abbildungen 4.7 bis 4.9 ist durchgehend $M = 3$. Man beachte, dass Abbildung 4.9 einen größeren Zeitparameter-Bereich zeigt, als die Abbildungen 4.7 und 4.8.

Problematisch bei der Glättung des Periodogramms mittels Fenstern ist, dass der so erhaltene Schätzer für g stark von der Wahl des Windows und von M abhängt. Jeder dieser Schätzer ist jedoch konsistent (s. Priestley (1981[1])).

4.2.3 Die finite Fourier-Transformierte

Eine Berechnung des Periodogramms mittels der empirischen Autokorrelationsfunktion nach (4.11) ist rechentechnisch sehr aufwändig. Die meisten Software-Programme verwenden daher die *finite Fourier-Transformierte*

$$\zeta(\omega) := \frac{1}{\sqrt{2\pi N}} \sum_{t=1}^{N} (X_t - \overline{X}) e^{-i\omega t}, \ \omega \in [-\pi, \pi],$$

die eine direkte Berechnung des Periodogramms erlaubt und eine zuvorige Berechnung der Autokovarianzfunktion unnötig macht.

4.2.2 Lemma: *Es bezeichne $\overline{\zeta(\omega)}$ die zu $\zeta(\omega)$ konjugiert komplexe Zahl und $\overline{X}$ das arithmetische Mittel von $(X_t)_{t=1}^N$. Mit diesen Bezeichnungen gilt für die reellwertige Zeitreihe (X_t)*

$$I_N^*(\omega) = \zeta(\omega)\overline{\zeta(\omega)} = \frac{1}{2\pi N} \left| \sum_{t=1}^{N} (X_t - \overline{X}) e^{-i\omega t} \right|^2, \ \omega \in [-\pi, \pi].$$

Beweis:

$$\begin{aligned} & \frac{1}{2\pi N} \left(\sum_{t=1}^{N} (X_t - \overline{X}) e^{-i\omega t} \right) \left(\sum_{t=1}^{N} (X_t - \overline{X}) e^{i\omega t} \right) \\ = & \frac{1}{2\pi N} \sum_{t=1}^{N} \sum_{s=1}^{N} (X_t - \overline{X})(X_s - \overline{X}) e^{-i\omega(t-s)} \\ = & \frac{1}{2\pi N} \sum_{t=1}^{N} \sum_{h=-(t-1)}^{N-t} (X_t - \overline{X})(X_{t+h} - \overline{X}) e^{-i\omega h}. \end{aligned}$$

Eine Änderung der Summationsreihenfolge (das Gitter $\{(t,s) : \ t,s = 1,\ldots,N\}$ wird nicht mehr zeilenweise, sondern in diagonaler Richtung aufaddiert) ergibt

$$\begin{aligned} & \frac{1}{2\pi N} \left(\sum_{t=1}^{N} (X_t - \overline{X}) e^{-i\omega t} \right) \left(\sum_{t=1}^{N} (X_t - \overline{X}) e^{i\omega t} \right) \\ = & \frac{1}{2\pi N} \sum_{h=-(N-1)}^{N-1} \sum_{t=1}^{N-|h|} (X_{t+|h|} - \overline{X})(X_t - \overline{X}) e^{-i\omega h} \\ = & \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) e^{-i\omega h}. \end{aligned}$$

□

4.2.4 Ein Schätzer für das integrierte Spektrum

So wie bei Wahrscheinlichkeitsverteilungen die Schätzung der Verteilungsfunktion sehr viel weniger Schwierigkeiten bereitet als die Schätzung der Verteilungsdichte, so ist auch die Schätzung des *integrierten Spektrums* $G(\omega) := \int_{-\pi}^{\omega} g(\theta)d\theta,\ \omega \in [-\pi, \pi]$, sehr viel weniger problematisch als die Schätzung der Spektraldichte. Ist (X_t) ein allgemeiner linearer Prozess der Form (3.10) mit

(*i*) $\mathrm{E}(Z_t^4) < \infty$,

(*ii*) es existiert ein $k > 2$ so, dass $\psi_j = O(\frac{1}{|j|^k})$ $(j \to \infty)$,

so ist

$$\hat{G}(\omega) = \int_{-\pi}^{\omega} I_N^*(\theta)d\theta$$

ein konsistenter Schätzer für $G(\omega)$ (vgl. Priestley (1981[1])).

Kapitel 5. Lineare Modelle

5.1 Lineare Modelle

5.1.1 Lineare Modelle mit Kovarianzmatrix $\sigma^2 I_n$

Es sei $(Y_t)_{t\in T}$ ein stochastischer Prozess, der zu verschiedenen Zeitpunkten $t_{n1},\dots,t_{nn},\ n\in\mathbb{N}$, beobachtet wird. Diese Zeitpunkte sollen in einem Intervall $[a,b]$, genannt *Versuchsbereich*, liegen. Ein Tupel $(t_{n1},\dots,t_{nn})\in[a,b]^n$ heißt *Design* für n Versuche. Aufgrund der Beobachtungen $Y_{t_{n1}},\dots,Y_{t_{nn}}$ soll nun ein den Daten zugrundeliegender funktionaler Zusammenhang gefunden werden. Gesucht ist also eine Aussage der Form

$$Y_{ni} := Y_{t_{ni}} = \beta_1 f_1(t_{ni}) + \cdots + \beta_m f_m(t_{ni}) + \epsilon(t_{ni}),\ i=1,\dots n,\ n\in\mathbb{N}, \tag{5.1}$$

für geeignet gewählte Funktionen $f_1,\dots,f_m$ $(m\le n)$ und zufällige Fehler $\epsilon(t_{ni}) =: \epsilon_{ni}$. Für den Fehlervektor $\epsilon_n := (\epsilon_{n1},\dots,\epsilon_{nn})^\top$ gelte $\mathrm{E}(\epsilon_n)=0$ und

$$\mathrm{Cov}(\epsilon_n) = \sigma^2 I_n. \tag{5.2}$$

Die Funktionen $f_1,\dots f_m$ heißen *Regressionsfunktionen*, die $n\times m$-Matrix X_n mit

$$X_n := \begin{pmatrix} f_1(t_{n1}) & \dots & f_m(t_{n1}) \\ \vdots & & \vdots \\ f_1(t_{nn}) & \dots & f_m(t_{nn}) \end{pmatrix}$$

wird *Designmatrix* genannt.

Mit $\beta := (\beta_1,\dots,\beta_m)^\top$ und $Y_n := (Y_{n1},\dots,Y_{nn})^\top$ lässt sich (5.1) schreiben in der Form

$$Y_n = X_n\beta + \epsilon_n,\ \mathrm{E}(\epsilon_n)=0,\ n\in\mathbb{N}. \tag{5.3}$$

Man nennt (5.3) ein *lineares (Regressions-) Modell*. Wohlgemerkt, (5.3) ist linear in dem unbekannten Parametervektor β, der ja gerade von Interesse ist. Es gilt $\mathrm{E}(Y_n) = X_n\beta$.

5.1.1 Beispiel: (Geraden-Regression) Gegeben seien die Regressionsfunktionen $f_1\equiv 1$, $f_2(t)=t,\ t\in\mathbb{R}$, und es sei $Y=(Y_{n1},\dots,Y_{nn})^\top,\ \epsilon_n=(\epsilon_{n1},\dots,\epsilon_{nn})^\top$ mit $\mathrm{E}(\epsilon_n)=0$ sowie $\mathrm{Cov}(\epsilon_n)=\sigma^2 I_n$. Das lineare Modell der Form

$$Y_{ni} = \beta_1 + \beta_2 t_{ni} + \epsilon_{ni},\ i=1,\dots,n,$$

heißt *Geraden-Regression*, wobei β gegeben ist durch $\beta=(\beta_1,\beta_2)^\top\in\mathbb{R}^2$.

5.1.2 Beispiel: Es seien $Y_{n1},\dots,Y_{nn}$ unabhängige Zufallsvariablen mit gleicher Varianz $\mathrm{Var}(Y_{ni})=\sigma^2$. Soll an $Y_{n1},\dots,Y_{nn}$ ein Polynom zweiten Grades angepasst werden, so wählt man $f_1\equiv 1$, $f_2(t)=t,\ t\in\mathbb{R}$, und $f_3(t)=t^2,\ t\in\mathbb{R}$. Das lineare Modell besitzt dann die Form

$$Y_{ni} = \beta_1 + \beta_2 t_{ni} + \beta_3 t_{ni}^2 + \epsilon_{ni},\ i=1,\dots,n,$$

mit $\mathrm{E}(\epsilon_n)=0$, $\mathrm{Cov}(\epsilon_n)=\sigma^2 I_n$ und $\beta=(\beta_1,\beta_2,\beta_3)^\top\in\mathbb{R}^3$.

Üblicherweise schätzt man den Erwartungswert $\mathrm{E}(Y_n)=X_n\beta$ durch den *Kleinste-Quadrate-Schätzer (Least Squares Estimator)* $X_n\hat\beta$:

5.1.3 Definition und Bemerkung: Der Schätzer $X_n\hat\beta$ für $X_n\beta$, der die Bedingung

$$(Y_n - X_n\hat\beta)^\top(Y_n - X_n\hat\beta) = \min_\beta (Y_n - X_n\beta)^\top(Y_n - X_n\beta) \tag{5.4}$$

erfüllt, heißt *Kleinste-Quadrate-Schätzer (Least Squares Estimator)* für $X_n\beta$.

Nach dem Projektionssatz (s. Anhang B.3) wird (5.4) minimiert durch die Orthogonalprojektion von Y_n auf den Bildraum $C(X_n)$ von X_n.

5.1.4 Satz: *Besitzt die Matrix $X_n \in \mathbb{R}^{n\times m},\ m \leq n$, den Rang $Rg(X_n) = m$, so ist $X_n^\top X_n$ invertierbar und $M_n := X_n(X_n^\top X_n)^{-1}X_n^\top$ ist die Abbildungsmatrix der Orthogonalprojektion von Y_n auf den Bildraum $C(X_n)$ von X_n.*

Beweis: Es ist (i) und (ii) aus Definition B.3.7 zu zeigen:

(i) Es sei $y \in C(X_n)$, d.h. $y = X_n\tilde{y}$ für ein $\tilde{y} \in \mathbb{R}^n$. Dann ist

$$M_n y = X_n(X_n^\top X_n)^{-1}X_n^\top X_n\tilde{y} = X_n\tilde{y} = y.$$

(ii) Es sei $y \perp C(X_n)$, d.h. $\langle y, X_n\tilde{y}\rangle = 0$ für alle $\tilde{y} \in \mathbb{R}^n$. Dann ist auch $\langle M_n y, X_n\tilde{y}\rangle = \langle y, M_n X_n\tilde{y}\rangle = \langle y, X_n\tilde{y}\rangle = 0$ für alle $\tilde{y} \in \mathbb{R}^n$, also $M_n y \perp C(X_n)$. Andererseits ist $M_n y = X_n(X_n^\top X_n)^{-1}X_n^\top y$ ein Element aus $C(X_n)$ und somit orthogonal zu sich selbst. Das einzige Element, das diese Eigenschaft besitzt, ist der Nullvektor. □

5.1.5 Bemerkung: Als Abbildungsmatrix einer Orthogonalprojektion ist M_n symmetrisch und idempotent (s.Anhang B.3).

5.1.6 Bemerkung: Ist $X_n^\top X_n$ singulär, verwendet man anstelle von $(X_n^\top X_n)^{-1}$ eine *verallgemeinerte Inverse* $(X_n^\top X_n)^-$ von $X_n^\top X_n$, z.B. die Pseudoinverse.

5.1.7 Definition: Die Pseudoinverse einer Matrix A ist eine Matrix A^- für die gilt

$AA^-A = A$
$A^-AA^- = A^-$
AA^- und A^-A sind symmetrisch.

5.1.8 Satz: *Für jede Matrix $A \in \mathbb{R}^{m\times n}$ existiert eine eindeutig bestimmte Pseudoinverse $A^- \in \mathbb{R}^{n\times m}$.*
Ist $A \in \mathbb{R}^{m\times n}$ mit $Rg(A) = m$, so ist $AA^\top$ regulär und $A^- = A^\top(AA^\top)^{-1}$.
Ist $A \in \mathbb{R}^{m\times n}$ mit $Rg(A) = n$, so ist $A^\top A$ regulär und $A^- = (A^\top A)^{-1}A^\top$.
Ist $A \in \mathbb{R}^{n\times n}$ regulär, so ist $A^- = A^{-1}$.

Beweis: Graybill (1976).

Die Berechnung der Pseudoinversen ist z.B. in Graybill (1976) ausführlich beschrieben. In Mathematik- oder Statistik-Softwareprogrammen ist sie in der Regel fertig implementiert.

Im Folgenden bezeichnet A^{-1} die Inverse von A, falls A invertierbar ist und ansonsten die Pseudoinverse. Mit dieser Bezeichnung ist der Kleinste-Quadrate-Schätzer für $X_n\beta$ gegeben durch

$$X_n\hat{\beta} = M_n Y_n = X_n(X_n^\top X_n)^{-1}X_n^\top Y_n. \tag{5.5}$$

5.1.9 Bemerkung: Der Least Squares (LS) Estimator ist erwartungstreu ("unbiased"), d.h.

$$\mathrm{E}(X_n\hat{\beta}) = \mathrm{E}(X_n(X_n^\top X_n)^{-1}X_n^\top(X_n\beta + \epsilon_n)) = X_n\beta,$$

und wegen $X_n\hat{\beta} = M_n Y_n$ linear.

5.1.10 Satz und Definition: *Der Least Squares Estimator besitzt in der Klasse der linearen erwartungstreuen Schätzer die kleinste Varianz und ist in diesem Sinne optimal ("best"). Einen solchen Schätzer nennt man Best Linear Unbiased Estimator (BLUE).*

Beweis: Christensen (1987).

5.1.2 Lineare Modelle mit positiv definiter Kovarianzmatrix

5.1.11 Definition und Bemerkung: Eine Matrix $A \in \mathbb{R}^{n\times n}$ heißt *nichtnegativ definit*, falls A symmetrisch ist und für alle $y \in \mathbb{R}^n$ mit $y \neq 0$ gilt $y^\top Ay \geq 0$. Gilt sogar $y^\top Ay > 0\ \forall y \neq 0$, so heißt A *positiv definit*. Eine positiv definite Matrix ist invertierbar und ihre Inverse ist ebenfalls positiv definit.

5.1.12 Beispiel: Für eine beliebige Matrix M sind $M^\top M$ und $MM^\top$ nichtnegativ definit.

5.1.13 Satz und Definition: *Ist A positiv definit, so existieren eine positiv definite Matrix B und eine reguläre obere Dreiecksmatrix D mit*

$$\begin{aligned} A &= B^2, \\ A &= D^\top D. \end{aligned} \tag{5.6}$$

Man nennt B die Quadratwurzel von A und schreibt $B = A^{1/2}$. Die Darstellung (5.6) heißt *Cholesky-Zerlegung* von A.

Beweis: Horn / Johnsen (1988).

In Abschnitt 5.1.1 wurden lineare Modelle mit Fehler-Kovarianzmatrix $\sigma^2 I_n$ behandelt. In diesem Abschnitt soll der Fall einer positiv definiten Fehler-Kovarianzmatrix betrachtet werden. Es sei also

$$Y_n = X_n\beta + \epsilon_n \text{ mit } \mathrm{Cov}(\epsilon_n) = \sigma^2\Sigma_n,\ \Sigma_n \text{ positiv definit.} \tag{5.7}$$

Da Σ_n positiv definit ist, existiert sowohl die Inverse Σ_n^{-1}, als auch eine symmetrische Matrix $\Sigma_n^{-\frac{1}{2}}$ mit der Eigenschaft $\Sigma_n^{-\frac{1}{2}} \cdot \Sigma_n^{-\frac{1}{2}} = \Sigma_n^{-1}$.

Bezeichnet nun $\tilde{Y}_n = \Sigma_n^{-\frac{1}{2}} Y_n,\ \tilde{X}_n = \Sigma_n^{-\frac{1}{2}} X_n$ und $\tilde{\epsilon}_n = \Sigma_n^{-\frac{1}{2}}\epsilon_n$, so besitzt das transformierte Modell

$$\tilde{Y}_n = \tilde{X}_n\beta + \tilde{\epsilon}_n,\ \mathrm{E}(\tilde{\epsilon}_n) = 0, \tag{5.8}$$

die Fehler-Kovarianzmatrix $\sigma^2 I_n$. Mit Abschnitt 5.1.1 erhält man den Least Squares Schätzer für $\tilde{X}_n\beta$

$$\begin{aligned} \tilde{X}_n\hat{\beta}_G &= \tilde{X}_n(\tilde{X}_n^\top\tilde{X}_n)^{-1}\tilde{X}_n^\top\tilde{Y}_n \\ &= \Sigma_n^{-\frac{1}{2}} X_n(X_n^\top\Sigma_n^{-1}X_n)^{-1}X_n^\top\Sigma_n^{-1}Y_n. \end{aligned} \tag{5.9}$$

Da der Parametervektor durch die Tansformation des linearen Modells (5.7) in das lineare Modell (5.8) nicht beeinflusst wird, ist $\hat{\beta}_G$ aus (5.9) BLUE für den Parametervektor des Modells (5.7). Man bezeichnet $\hat{\beta}_G$ auch als *verallgemeinerten* (*generalized*) LS-Schätzer.. Es gilt

$$X_n\hat{\beta}_G = \Sigma_n^{\frac{1}{2}}\tilde{X}_n\hat{\beta}_G = M_G Y_n,\ \text{mit } M_G = X_n(X_n^\top\Sigma_n^{-1}X_n)^{-1}X_n^\top\Sigma_n^{-1}. \tag{5.10}$$

5.2 Residuen

5.2.1 Least Squares Residuen

5.2.1 Definition und Bemerkung: Es sei $\hat{\beta}$ der LS-Schätzer aus (5.5). Dann heißt der Vektor

$$r_n := Y_n - X_n\hat{\beta} = (I_n - M_n)Y_n \tag{5.11}$$

Least Squares (LS) Residuen-Vektor, seine Komponenten $r_{ni},\ i = 1,\ldots,n$, werden *(LS) Residuen* genannt. Unter Vorliegen des Modells (5.3) gilt

$$\mathrm{E}(r_n) = (I_n - M_n)(X_n\beta + \mathrm{E}(\epsilon_n)) = X_n\beta - \underbrace{M_nX_n\beta}_{=X_n\beta} = 0,$$

und unter der Annahme (5.2) ist

$$\mathrm{Cov}(r_n) = (I_n - M_n)\mathrm{Cov}(Y_n)(I_n - M_n)^\top = \sigma^2(I_n - M_n). \tag{5.12}$$

5.2.2 Bemerkung: Die Matrix $\tilde{M}_n := (I_n - M_n)$ ist symmetrisch und idempotent. Sie ist die Abbildungsmatrix der Orthogonalprojektion auf das orthogonale Komplement $(C(X_n))^\perp$ von $C(X_n)$.

Der LS-Schätzer $X_n\hat{\beta}$ für $X_n\beta$ ist zwar optimal im Sinne von Satz 5.1.10, besitzt aber den Nachteil, dass die LS-Residuen korreliert sind und eine schwankende Varianz aufweisen. Dies gilt selbst im Falle unabhängiger und identisch verteilter Fehler $\epsilon_{ni}, i = 1, \ldots, n$, wie man in der Kovarianzmatrix (5.12) des Residuenvektors sehen kann. Für manche Zwecke, z.B. das Testen auf Normalverteilung, wird daher in der Literatur eine *Standardisierung* der Residuen empfohlen. Dieses Verfahren gewichtet jedes Residuum mit der Inversen seiner geschätzten Standardabweichung. Nach (5.12) ist der Vektor der standardisierten Residuen $\tilde{r}_n = (\tilde{r}_{n1}, \ldots, \tilde{r}_{nn})^\top$ also gegeben durch

$$\tilde{r}_{ni} = \frac{r_{ni}}{\sqrt{\widehat{\sigma^2}(1 - M_{ii})}},$$

wobei (M_{ii}) das i-te Diagonalelement der Matrix M_n bezeichnet und $\widehat{\sigma^2} := \frac{1}{n-m}\sum_{i=1}^n r_{ni}^2$ die Fehlervarianz σ^2 schätzt (m ist die Anzahl der geschätzten Parameter $\hat{\beta}_1, \ldots, \hat{\beta}_m$). Die standardisierten Residuen besitzen somit unter der Annahme (5.2) die konstante Varianz $\mathrm{Var}(\tilde{r}_{ni}) = 1,\ i = 1, \ldots, n$. Sie sind jedoch, wie die nicht-standardisierten LS-Residuen, korreliert. Genauer gilt

$$\mathrm{Cov}(\tilde{r}_i, \tilde{r}_j) = \frac{-M_{ij}}{\sqrt{(1 - M_{ii})(1 - M_{jj})}},\ (i \neq j).$$

Für einige Zwecke, z.B. das Testen von Unkorreliertheit, bilden die standardisierten Residuen also keine befriedigende Basis. Man hat daher nach Möglichkeiten gesucht, Residuen so zu definieren, dass sie unter der Modellannahme (5.2) unkorreliert sind.

5.2.2 Unkorrelierte Residuen

Gegeben sei das lineare Modell (5.1) mit Fehlerkovarianzmatrix (5.2). Ein *linearer, unkorrelierter Residuenvektor* ist ein $w \in \mathbb{R}^{n-m}$ mit

$$w = U_n \cdot Y_n, \tag{5.13}$$

wobei $U_n \in \mathbb{R}^{(n-m)\times n}$ so beschaffen ist, dass gilt

$$U_n X_n = 0, \tag{5.14}$$
$$U_n U_n^\top = I_{n-m}. \tag{5.15}$$

Nach (5.14) ist $w = U_n\epsilon_n$ und somit $\mathrm{E}(w) = U_n\mathrm{E}(\epsilon_n) = 0$. Man bezeichnet daher w als erwartungstreuen (unbiased) "Schätzer" für ϵ_n. (Man beachte, dass diese Bezeichnung nicht ganz korrekt ist, da es sich bei ϵ_n nicht um einen zu schätzenden Parameter, sondern um einen Zufallsvektor handelt. Außerdem besitzen die Vektoren ϵ_n und w unterschiedlichen Dimensionen.) Wegen (5.15) gilt $\mathrm{Cov}(w) = U_n\mathrm{Cov}(\epsilon_n)U_n^\top = \sigma^2 I_{n-m}$. Man sagt in diesem Zusammenhang, w besitze eine "skalare Kovarianzmatrix". Schließlich geht w nach (5.13) mittels einer linearen Abbildung aus Y_n hervor. Unkorrelierte Residuen werden daher auch als *LUS- (Linear Unbiased with Scalar covariance) Residuen* bezeichnet.

Ist r_n der LS-Residuenvektor, so gilt wegen (5.14)

$$U_n r_n = (U_n - U_n X_n (X_n^\top X_n)^{-1} X_n^\top) Y_n = U_n Y_n = U_n\epsilon_n = w.$$

Ein unkorrelierter Residuenvektor w kann also aus dem Vektor r_n der LS-Residuen berechnet werden. Eine Matrix U_n kann zum Beispiel wie folgt erhalten werden:

Man unterteilt X_n in die ersten m und die letzten $n - m$ Zeilen und definiert $X_m,\ X_*$ und Z so, dass

$$X_n = \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} = \begin{pmatrix} I_m \\ \cdots \\ Z \end{pmatrix} X_m. \tag{5.16}$$

Dabei wird $\mathrm{Rg}(X_m) = m$ vorausgesetzt, so dass $Z = X_* X_m^{-1}$. (Ist X_m singulär, so müssen die Beobachtungen Y_i und die Zeilen von X_n so umgeordnet werden, dass die ersten m Zeilen der Designmatrix linear unabhängig sind, s. Theil (1965). Ist X_m einmal geeignet gewählt, so wird die Unterteilung gemäß (5.16) als fest betrachtet.)

Mit $J := (-Z \vdots I_{n-m}) \in \mathbb{R}^{(n-m)\times n}$ gilt

$$JX_n = (-Z \vdots I_{n-m}) \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} = -ZX_m + X_* = 0 \tag{5.17}$$

und somit

$$v := Jr_n = JY_n - JX_n\hat{\beta} = JY_n = JX_n\beta + J\epsilon_n = J\epsilon_n. \tag{5.18}$$

5.2.3 Bemerkung: Die zur Matrix J korrespondierende lineare Abbildung transformiert den n-dimensionalen Residuenvektor r_n in den $(n-m)$-dimensionalen Residuenvektor v.

5.2.4 Lemma: *Der Vektor v besitzt den Erwartungswertvektor 0 und die reguläre Kovarianzmatrix $\sigma^2 JJ^\top$.*

Beweis: Es gilt $\mathrm{Cov}(v) = J\mathrm{Cov}(\epsilon)J^\top = \sigma^2 JJ^\top = \sigma^2(I_{n-m} + ZZ^\top)$. Die symmetrische Matrix $ZZ^\top$ ist nichtnegativ definit, d.h. für alle $y \in \mathbb{R}^{n-m}$ ist $y^\top(ZZ^\top)y \geq 0$. Für die ebenfalls symmetrische Matrix $JJ^\top = I_{n-m} + ZZ^\top$ folgt daher für jedes $y \in \mathbb{R}^{n-m}$, $y \neq 0$,

$$y^\top(JJ^\top)y = \underbrace{y^\top y}_{>0} + \underbrace{y^\top(ZZ^\top)y}_{\geq 0} > 0.$$

Also ist $JJ^\top$ nach Definition positiv definit und somit regulär. □

5.2.5 Satz: *Für $\tilde{M}_* := (JJ^\top)^{-1} \in \mathbb{R}^{(n-m)\times(n-m)}$ gilt*

(i) $\tilde{M}_* = I_{n-m} - X_*(X_n^\top X_n)^{-1}X_*^\top$,

(ii) *$\tilde{M}_*$ entspricht der rechten unteren Teilmatrix von $\tilde{M}_n = I_n - M_n$,*

(iii) *$\tilde{M}_*$ ist positiv definit, d.h.*
$\tilde{M}_*$ ist symmetrisch und für alle $y \in \mathbb{R}^{n-m}$ mit $y \neq 0$ ist $y^\top \tilde{M}_* y > 0$.

Beweis:

(i) Wegen

$$X_n^\top X_n = \left(X_m^\top \vdots X_*^\top\right) \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} = X_m^\top X_m + X_*^\top X_*$$

ist

$$\begin{aligned} & (I_{n-m} - X_*(X_n^\top X_n)^{-1}X_*^\top)(I_{n-m} + X_*(X_m^\top X_m)^{-1}X_*^\top) \qquad (5.19) \\ = \; & I_{n-m} - X_*(X_n^\top X_n)^{-1}X_*^\top + X_*(X_m^\top X_m)^{-1}X_*^\top \\ & -X_*(X_n^\top X_n)^{-1} \underbrace{X_*^\top X_*}_{=X_n^\top X_n - X_m^\top X_m} (X_m^\top X_m)^{-1}X_*^\top \\ = \; & I_{n-m}. \qquad (5.20) \end{aligned}$$

Somit gilt (man beachte die Symmetrie in (5.19))

$$\begin{aligned} (I_{n-m} - X_*(X_n^\top X_n)^{-1}X_*^\top)^{-1} &= I_{n-m} + X_*(X_m^\top X_m)^{-1}X_*^\top \\ &= I_{n-m} + ZZ^\top = JJ^\top. \end{aligned}$$

(ii) Für $M_n = X_n(X_n^\top X_n)^{-1}X_n^\top$ gilt

$$\begin{aligned} X_n(X_n^\top X_n)^{-1}X_n^\top &= \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} (X_n^\top X_n)^{-1} \left(X_m^\top \vdots X_*^\top \right) \\ &= \begin{pmatrix} X_m(X_n^\top X_n)^{-1}X_m^\top & X_m(X_n^\top X_n)^{-1}X_*^\top \\ X_*(X_n^\top X_n)^{-1}X_m^\top & X_*(X_n^\top X_n)^{-1}X_*^\top \end{pmatrix}. \end{aligned}$$

Somit ergibt sich mit (i) die Behauptung.

(iii) $\tilde{M}_*$ ist als Inverse der positiv definiten Matrix $JJ^\top$ ebenfalls positiv definit.

Nach den Sätzen 5.2.5 und 5.1.13 existiert also eine Quadratwurzel von $\tilde{M}_*$. Für die Matrix $U_n := \tilde{M}_*^{1/2}J$ gilt:

$$\begin{aligned} U_nX_n &= \tilde{M}_*^{1/2}\underbrace{JX_n}_{=0} = 0, \\ \text{sowie } U_nU_n^\top &= \tilde{M}_*^{1/2}\underbrace{JJ^\top}_{=\tilde{M}_*^{-1}}M_*^{1/2} = I_{n-m}. \end{aligned}$$

Somit ist $w = U_nY_n = U_nr_n$ ein Vektor unkorrelierter Residuen.

Es stellt sich nun die Frage, ob es noch weitere unkorrelierte Residuenvektoren gibt.

5.2.6 Definition: Eine Matrix P heißt *orthogonal*, falls $P^\top = P^{-1}$.

5.2.7 Satz: *Für eine gegebene Unterteilung* $X_n = \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix}$ *von* X_n *ist die Klasse aller* $U_n \in \mathbb{R}^{(n-m)\times n}$*, die (5.14) und (5.15) erfüllen, gegeben durch*

$$U_n = \underbrace{P\tilde{M}_*^{1/2}}_{=:U_*}J,$$

wobei P *eine* $(n-m)\times(n-m)$*-Orthogonalmatrix ist.*

Beweis: Wegen $JX_n = 0$ ist $U_nX_n = P\tilde{M}_*^{1/2}JX_n = 0$ für alle P. Weiter gilt

$$\begin{aligned} U_nU_n^\top = P\tilde{M}_*^{1/2}J(P\tilde{M}_*^{1/2}J)^\top &= P\tilde{M}_*^{1/2}JJ^\top\tilde{M}_*^{1/2}P^\top \\ &= PP^\top. \end{aligned}$$

Schließlich ist

$$PP^\top = I_{n-m} \iff P \text{ orthogonal.}$$

□

Ein Vektor $w \in \mathbb{R}^{n-m}$ ist also genau dann ein Vektor unkorrelierter Residuen, wenn $w = P\tilde{M}_*^{1/2}JY_n$ gilt mit einer orthogonalen Matrix P.

5.2.8 Bemerkung: Man beachte, dass $\tilde{M}_* = (JJ^\top)^{-1} = (I_{n-m} + ZZ^\top)^{-1}$ lediglich von $Z \in \mathbb{R}^{(n-m)\times m}$ aus Gleichung (5.16) abhängt. Theil (1965) hebt die Bedeutung der Matrix Z hervor, die besonders für $m = 1$ deutlich wird: In diesem Falle ist $X_m =: c \neq 0$ ein Skalar, $X_* =: (x_1, \ldots, x_{n-1})^\top$ ein Vektor. $Z = X_*X_m^{-1}$ ist also der skalierte Vektor $(x_1/c, \ldots, x_{n-1}/c)^\top$, X_m die "Skalierungsgrundlage". Ist z.B. das jährliche Einkommen in Euro die einzige erklärende Variable, so wird das erste Jahr als Grundlage der Skalierung verwendet (Jahr 1 = 100%), alle folgenden Jahre werden im Verhältnis zu diesem Jahr betrachtet. Für $m > 1$ erfolgt die Skalierung in Bezug auf mehrere erklärende Variablen.

Für das Modell (5.2) mit $\mathrm{Rg}(X_n) = m$ erhält man mit Satz 5.2.7 einen $(n-m)$-dimensionalen Vektor unkorrelierter Residuen. Das folgende Lemma sagt, dass es keine höherdimensionalen Vektoren unkorrelierter Residuen gibt.

5.2.9 Lemma: *Im linearen Modell (5.1) mit $Rg(X_n) = m$ kann man höchstens $n - m$ unkorrelierte Residuen erhalten.*

Beweis: Es sei

$$X_n = \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} \text{ mit } \mathrm{Rg}(X_n) = \mathrm{Rg}(X_m) = m.$$

Dann ist jede Zeile aus X_* eine Linearkombination von Zeilenvektoren von X_m, d.h. es existiert ein $A \in \mathbb{R}^{(n-m)\times m}$ mit $X_* = AX_m$.

Weiter sei $U_n \in \mathbb{R}^{p\times n}$ mit $p \leq n$ und U_m bezeichne die Matrix der ersten m Spalten von U_n. Schreibe $U_n = (U_m \vdots U_*)$. Mit diesen Bezeichnungen ist nach (5.14)

$$\begin{aligned} 0 = U_n X_n &= (U_m \vdots U_*) \begin{pmatrix} X_m \\ \cdots \\ X_* \end{pmatrix} \\ &= U_m X_m + U_* X_* \\ &= U_m X_m + U_* A X_m. \end{aligned}$$

Da X_m invertierbar ist, gilt also $U_m = -U_* A$, d.h. die m Spalten von U_m sind Linearkombinationen der $n - m$ Spalten von U_*. Mit (5.15) ist somit $p = \mathrm{Rg}(U_n U_n^\top) = \mathrm{Rg}(U_n) \leq n - m$. □

5.2.10 Bemerkung: Der Hintergrund von Lemma 5.2.9 ist die Tatsache, dass $\hat{\beta}_j$ für jedes $j = 1, \ldots, m$ eine Linearkombination der Beobachtungen $Y_{n1}, \ldots, Y_{nn}$ ist. Auch im Vektor der LS-Residuen $r_n = Y_n - X_n\hat{\beta}$ sind also m der n Residuen durch lineare Abhängigkeiten in Folge der Schätzung des Parametervektors β "gebunden". Man spricht in diesem Zusammenhang von einem "Verlust von m Freiheitsgraden".

5.2.11 Bemerkung: Ist der Fehlervektor ϵ_n normalverteilt, so besitzt auch ein Vektor unkorrelierter Residuen zum Modell (5.1) (wie auch der LS-Residuenvektor) als lineare Transformation von ϵ_n eine Normalverteilung.

BLUS-Residuen

In der Klasse aller unkorrelierten Residuen zum linearen Modell (5.1) mit der Designmatrix (5.16) ist

$$w = \tilde{M}_*^{1/2} J\epsilon_n = \tilde{M}_*^{1/2} J r_n$$

im folgenden Sinne optimal:

w minimiert in der Klasse $\mathcal{R}$ der gemäß Satz 5.2.7 gegebenen unkorrelierten Residuen den euklidischen Abstand zum $(n - m)$-dimensionalen Residuenvektor $v = Jr_n = J\epsilon_n$ aus (5.18), d.h.

$$\mathrm{E}\left[(w - J\epsilon_n)^\top (w - J\epsilon_n)\right] = \min_{z\in\mathcal{R}} \mathrm{E}\left[(z - J\epsilon_n)^\top (z - J\epsilon_n)\right],$$

(s. Godolphin / de Tullio (1978)). Man nennt $w = \tilde{M}_*^{1/2} J r_n$ daher *BLUS-Residuenvektor* (Best Linear Unbiased with Scalar covariance).

Der BLUS-Residuenvektor besitzt jedoch den Nachteil, dass Informationen, die in der Reihenfolge der Residuen enthalten sind (z.B. über Heteroskedastizität), "verwischt" werden können. Um ein "Verwischen" derartiger Informationen soweit als möglich zu verhindern, kann man auf einen anderen unkorrelierten Residuenvektor zurückgreifen, nämlich auf den Vektor *rekursiver Residuen*:

Rekursive Residuen

Im Folgenden bezeichne X_{j-1} die Matrix, die aus den $j-1$ ersten Zeilen von X_n besteht ($j > m$), sowie Y_{j-1} den Vektor der ersten $j - 1$ Komponenten von Y_n. Dann kann für $j = m+1, \ldots, n$ der unbekannte Parametervektor β geschätzt werden durch

$$\hat{\beta}_{j-1} := \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top Y_{j-1}.$$

Bezeichnet $x_j^\top$ die j-te Zeile von X_n, so erhält man eine "Vorhersage von Y_{nj}" durch

$$\hat{Y}_{nj} = x_j^\top \hat{\beta}_{j-1}.$$

Man beachte, dass im Falle stochastisch unabhängiger Fehler $\epsilon_{n1}, \ldots, \epsilon_{nn}$ die Vorhersage $\hat{Y}_{nj}$ stochastisch unabhängig von Y_{nj} ist. Somit besitzt der "Vorhersagefehler" $Y_{nj} - \hat{Y}_{nj}$ die Varianz

$$\begin{aligned}
\mathrm{Var}(Y_{nj} - \hat{Y}_{nj}) &= \mathrm{Var}\left(Y_{nj} - x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top Y_{j-1}\right) \\
&= \sigma^2 + \left(x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top\right)\left(x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top\right)^\top \sigma^2 \\
&= \sigma^2 \left(1 + x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} x_j\right). \qquad (5.21)
\end{aligned}$$

Man definiert nun den Vektor w rekursiver Residuen durch

$$w_j = \frac{Y_{nj} - \hat{Y}_{nj}}{\sqrt{1 + x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} x_j}},\ j = m+1, \ldots, n.$$

Unter der Voraussetzung

$$\mathrm{E}(\epsilon_n) = 0 \text{ und } \mathrm{Cov}(\epsilon_n) = \sigma^2 I_n \qquad (5.22)$$

gilt für $j = 1, \ldots, n-m$

$$\begin{aligned}
\sqrt{1 + x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} x_j} \cdot \mathrm{E}(w_j) &= \mathrm{E}(Y_{nj}) - x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top \mathrm{E}(Y_{j-1}) \\
&= x_j^\top \beta - x_j^\top \left(X_{j-1}^\top X_{j-1}\right)^{-1} X_{j-1}^\top X_{j-1} \beta \\
&= 0,
\end{aligned}$$

sowie

$$\mathrm{Var}(w_j) = \sigma^2\ .$$

Außerdem ist

$$\begin{aligned}
\mathrm{Cov}(w_i, w_j) &= \mathrm{E}(Y_{ni} Y_{nj}^\top) - \mathrm{E}(Y_{ni} \hat{Y}_{nj}^\top) - \mathrm{E}(\hat{Y}_{ni} Y_{nj}^\top) + \mathrm{E}(\hat{Y}_{ni} \hat{Y}_{nj}^\top) \\
&= x_i^\top \beta\beta^\top x_j - 2 x_i^\top \beta\beta^\top x_j + x_i^\top (X_i^\top X_i)^{-1} X_i^\top \mathrm{E}(Y_i Y_j^\top) X_j (X_j^\top X_j)^{-1} x_j \\
&= 0\ (i \neq j).
\end{aligned}$$

Ein Vektor rekursiver Residuen besitzt also unter der Modellannahme (5.22) den Erwartungswert 0 und die Kovarianzmatrix $\sigma^2 I_{n-m}$.

5.2.12 Beispiel: Zum Vergleich der vorgestellten Residuen-Definitionen wurden zwei einfache lineare Modelle simuliert und zwar

(A) Geraden-Regression: $Y_{nt} = \frac{1}{2}t + \epsilon_{nt},\ t = \frac{1}{n}, \frac{2}{n}, \ldots, 1,\ n = 100,$
mit $\mathrm{Cov}(\epsilon_n) = \mathrm{diag}(\sigma_t^2,\ t = \frac{1}{n}, \frac{2}{n}, \ldots, 1)$, wobei $\sigma_t^2 = (\frac{1}{1+t})^2$.

(B) Polynom 2. Grades: $Y_{nt} = \frac{1}{2}t + \frac{1}{2}t^2 + \epsilon_{nt},\ t = \frac{1}{n}, \frac{2}{n}, \ldots, 1,\ n = 100,$
mit $\mathrm{Cov}(\epsilon_n)$ wie bei Modell (A).

An diese simulierten Daten wurde jeweils ein Regressionsmodell

$$Y_{nt} = \beta_1 + \beta_2 t + \epsilon_{nt},\ \mathrm{Cov}(\epsilon_n) = I_n, \qquad (5.23)$$

angepasst. Damit entspricht das Modell (5.23) für die nach (A) simulierten Daten einer korrekt spezifizierten Geraden-Regression, jedoch unter Vorliegen einer nicht modellierten Inhomogenität der Varianz (Heteroskedastizität). Für die nach (B) simulierten Daten entspricht (5.23) einem falschspezifizierten Modell mit nicht modellierter Heteroskedastizität.

Zu den nach (A) und (B) simulierten Daten wurden, nach Anpassung des Regressionsmodells (5.23), jeweils die LS-, BLUS- und rekursiven Residuen berechnet und dahingehend miteinander verglichen, welcher Residuenvektor sich am ehesten zur Schätzung der Varianzfunktion eignet (siehe Abschnitt 7.2.1). Dabei wurde angenommen, dass die funktionale Form $(\frac{1}{1+t})^p$ der Varianzfunktion bekannt und lediglich der Parameter p zu schätzen sei.

Zur Durchführung des Vergleichs wurden die (simulierten) Beobachtungsvektoren Y_n der Modelle (A) und (B) jeweils 1000-mal realisiert, die verschiedenen Residuenvektoren berechnet und jeweils p geschätzt. Als Mittelwert und empirische Varianz von $\hat{p}$ erhielt man unter Vorliegen von Modell (A) die Werte

- Mean=2.0969 und Var=0.1818 bei Verwendung der LS-Residuen,
- 1.9867 (Mean) bzw. 0.1728 (Varianz) für BLUS-Residuen und
- 1.9970 bzw. 0.1692 mit rekursiven Residuen.

Für Modell (B) ergaben sich die Werte

- 2.0885 bzw. 0.1805 (LS),
- 1.9771 bzw. 0.1725 (BLUS) und
- 1.9867 bzw. 0.1688 (rekursive).

Somit liegen für beide Modelle die mittels rekursiver Residuen erhaltenen Schätzer für p am nächsten beim wahren Wert $p = 2$. Die mittels LS-Residuen erhaltenen Schätzer schneiden am schlechtesten ab. Außerdem besitzen die durch rekursive Residuen erhaltenen Schätzer in beiden Fällen die geringste Varianz.

5.2.13 Bemerkung: In Bischoff et al. (2005) wird die Hypothese einer homogenen Varianz geodätischer GPS-Messreihen anhand der rekursiven Residuen nach linearer Modellbildung getestet und klar verworfen. (Tests auf Homogenität der Varianz s. Abschnitt 6.4.) In Bischoff et al. (2006) wird die Varianzfunktion geodätischer GPS-Messreihen mittels rekursiver Residuen und LS-Residuen geschätzt. Die Ergebnisse werden miteinander verglichen.

Ein ausführlicher Überblick über verschiedene Residuen-Definitionen befindet sich auch in Cook und Weisberg (1982).

5.2.3 Residuen in Modellen mit positiv definiter Kovarianzmatrix

Liegt das lineare Modell (5.7) vor und ist $\hat{\beta}_G$ der Generalized Least Squares (GLS) Schätzer aus (5.9), so heißt der Vektor

$$r_n = Y_n - X_n\hat{\beta}_G$$

der *verallgemeinerte kleinste Quadrate* oder *Generalized Least Squares* (GLS) Residuenvektor. Auch für den GLS Residuenvektor r_n gilt $\mathrm{E}(r_n) = 0$. Die übrigen Aussagen aus Abschnitt 5.2.1 gelten für den GLS Residuenvektor in Bezug auf das gewichtete Modell (5.8) mit M_G wie in Gleichnung (5.10).

McGilchrist und Sandland (1979) haben die Definition rekursiver Residuen auf den Fall abhängiger Fehler mit bekannter Kovarianzmatrix erweitert.

5.2.4 Pseudo-Residuen

Zur Bildung der Residuen aus Abschnitt 5.2.1, 5.2.2 oder 5.2.3 muss das lineare Modell (5.1) bzw. (5.7) bekannt sein. Ist es jedoch unbekannt oder falsch spezifiziert, so kann es notwendig oder gar von Vorteil sein, den Fehlervektor ϵ_n mit nichtparametrischen Methoden zu "schätzen". Man spricht dann von *Pseudo-* oder *Quasi-Residuen*.

Differenzenbildung

Das einfachste Beispiel von Pseudo-Residuen ist die Differenzenbildung zweier benachbarter Beobachtungen

$$R_j = Y_j - Y_{j-1},\ j = 2, \ldots, n. \tag{5.24}$$

5.2.14 Beispiel: Nach linearer Modellbildung weisen die Residuen einer geodätischen GPS-Messreihe aus der Region der Antarktischen Halbinsel (vgl. Bischoff et al. (2006)) ein trendartiges Verhalten auf, wie Abbildung 5.1 (mit eingezeichneten Satelliten-Elevationskurven) zeigt. Nach Differenzierung gemäß (5.24) kann der Trend als nahezu eliminiert betrachtet werden, s. Abbildung 5.2.

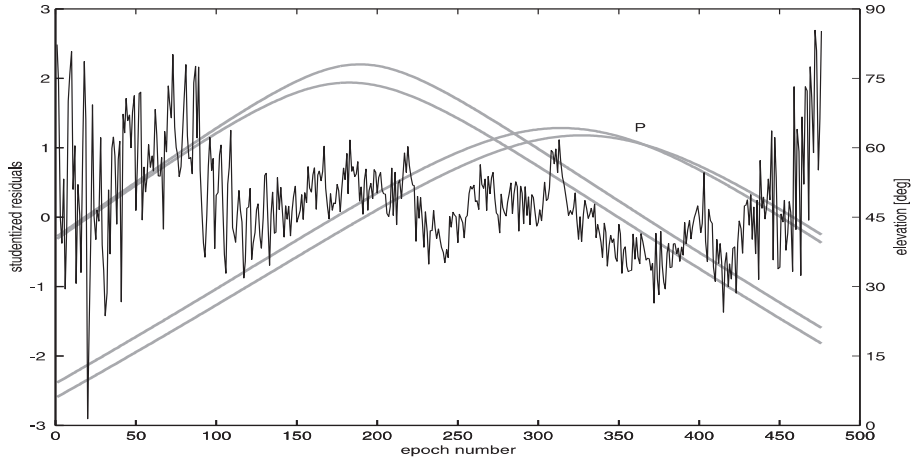

Abbildung 5.1: LS-Residuen einer GPS-Messreihe mit Satelliten-Elevationskurven

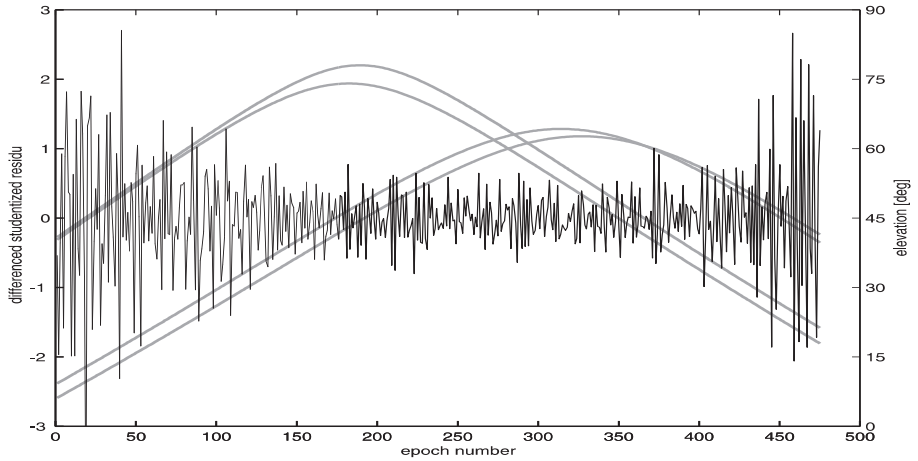

Abbildung 5.2: Die Zeitreihe aus Abb. 5.1 nach Differenzenbildung

Die Differenzenbildung aus Gleichung (5.24) entspricht der Anwendung des Differenzenoperators ∇ aus Abschnitt 3.1.3 auf die Zeitreihe $(Y_j)_{j=1}^n$. Verzichtet man bei der Differenzenbildung in (5.24) auf jedes zweite R_j, so erhält man im Falle unkorrelierter Fehler $\epsilon_{ni},\ i = 1, \ldots, n$, eine Folge von Differenzen $R_{2k},\ k = 1, \ldots, [\frac{n}{2}]$, mit derselben Eigenschaft. Die Eignung der Differenzen als "Schätzer" für die Fehler $\epsilon_{n2}, \ldots, \epsilon_{nn}$ hängt jedoch stark von den einzusetzenden statistischen Verfahren ab, was anhand eines einfachen Beispiels veranschaulicht werden soll:

5.2.15 Beispiel: Gegeben sei das lineare Modell

$$Y_{ni} = \beta_1 + \beta_2 t_{ni} + \epsilon_{ni},\ i = 1, \ldots, n, \tag{5.25}$$

mit $\mathrm{E}(\epsilon) = 0$ und $\mathrm{Cov}(\epsilon) = \sigma^2 I_n$. Dann gilt für $i = 2, \ldots, n$:

$$\mathrm{E}(R_i) = \mathrm{E}(Y_{ni} - Y_{n,i-1}) = \beta_2(t_{ni} - t_{n,i-1})$$

und

$$\mathrm{Cov}(R_i, R_j) = \mathrm{Cov}(\epsilon_{ni} - \epsilon_{n,i-1}, \epsilon_{nj} - \epsilon_{n,j-1}) = \begin{cases} 2\sigma^2, & i = j, \\ -\sigma^2, & i = j-1, \\ 0, & |i-j| > 1. \end{cases} \tag{5.26}$$

Im Falle äquidistanter Versuchspunkte, d.h. $t_{ni} - t_{n,i-1} = \frac{1}{n}$ $(i = 2, \ldots, n)$, ist der Vektor

$$d := (d_1, \ldots, d_{[\frac{n}{2}]})^\top := \frac{1}{\sqrt{2}}(R_2, R_4, \ldots, R_{2[\frac{n}{2}]})^\top$$

ein Vektor unkorrelierter Zufallsvariablen mit konstantem Erwartungswert $\frac{\beta_2}{n\sqrt{2}}$ und $\mathrm{Var}(d_k) = \sigma^2,\ k = 1, \ldots, [\frac{n}{2}]$. Wegen

$$\mathrm{E}(d) = \frac{\beta_2}{n\sqrt{2}} \cdot (1, \ldots, 1)^\top \neq (0, \ldots, 0)^\top = \mathrm{E}(\epsilon_2, \epsilon_4, \ldots, \epsilon_{2[\frac{n}{2}]})^\top$$

kann man für das Modell (5.25) die Verwendung des Differenzenvektors d als Vektor von Pseudo-Residuen nur eingeschränkt empfehlen. Der Differenzenvektor könnte z.B. zu Testzwecken eingesetzt werden, wenn $\mathrm{E}(d)$ bei der Bildung der Teststatistik nahezu eliminiert wird, wie dies bei der Dette-Munk-Teststatistik aus Abschnitt 6.4.6 der Fall ist.

Allgemeine Definition der Pseudo-Residuen

In der allgemeinen Form werden Pseudo-Residuen als gewichtete Summe von Beobachtungen

$$R_i := \sum_{j=k_l}^{k_r} w_{i+j} Y_{i+j},\ i = 1 - k_l, \ldots, n - k_r\ , \tag{5.27}$$

definiert, wobei für die Gewichte w_{i+j} gilt

$$\sum_{j=k_l}^{k_r} w_{i+j} = 0,\ i = 1 - k_l, \ldots, n - k_r\ .$$

5.2.16 Beispiel: Ist $(R_2, \ldots, R_n)^\top$ ein Differenzenvektor, so ist $k_l = -1,\ k_r = 0,\ w_{i-1} = -1$ sowie $w_i = 1$.

In der Regel besitzen die in der allgemeinen Form (5.27) definierten Quasi-Residuen einen betragsmäßig kleineren Erwartungswert als die Differenzen (5.24). Will man jedoch unkorrelierte Pseudo-Residuen, so muss ein erheblich höherer Anteil der R_i gestrichen werden, als dies bei Differenzenbildung notwendig ist.

5.3 Residuen-Partialsummenprozesse

Betrachten wir wieder das lineare Modell (5.1) und setzen voraus, dass für alle $n \in \mathbb{N}$ die Fehler $\epsilon_{ni} := \epsilon(t_{ni}),\ i = 1, \ldots, n$, stochastisch unabhängig und identisch verteilt sind. Weiter sei der Versuchsbereich gegeben durch das Intervall $[0, 1]$ (ggf. nach Transformation von $[a, b],\ a, b \in \mathbb{R}$ auf $[0, 1]$).

Der Einfachheit halber bezeichnet r_n im Folgenden einen Residuenvektor, unabhängig davon, ob es sich um LS-, BLUS-, rekursive oder Quasi-Residuen handelt. Die Dimension dieses Residuenvektors wird mit n bezeichnet. Man beachte, dass mit dieser Vereinbarung im Falle unkorrelierter Residuen $n = \dim(Y) - \mathrm{Rg}(X)$ gilt.

5.3.1 Definition und Bemerkung: Der Prozess

$$\frac{1}{\sigma\sqrt{n}}T_n(r_n)(z) := \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{[nz]} r_{ni} + (nz - [nz])r_{n,[nz]+1}\right) \tag{5.28}$$

heißt *Residuen-Partialsummenprozess*. Sind $t_{n1} = \frac{1}{n}, \ldots, t_{nn} = 1$ äquidistante Beobachtungszeitpunkte, so nimmt für $z = t_{nk},\ k \in \{1, \ldots, n\}$, (5.28) die Form

$$\frac{1}{\sigma\sqrt{n}}T_n(r_n)(z) = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{k} r_{ni}$$

an. In (5.28) hingegen verbindet die Komponente $(nz - [nz])r_{n,[nz]+1}$ die Punkte $(t_k, \frac{1}{\sigma\sqrt{n}}T_n(r_n)(t_k))$ und $(t_{k+1}, \frac{1}{\sigma\sqrt{n}}T_n(r_n)(t_{k+1}))$, $k = 1, \ldots, n-1$, durch jeweils eine Gerade.

5.3.2 Bemerkung: Sind die Residuen unabhängig und identisch verteilt, so lässt sich direkt der Satz von Donsker auf den Residuenvektor anwenden, und der Prozess (5.28) konvergiert nach Verteilung gegen die reelle, normale Brown'sche Bewegung, also

$$\frac{1}{\sigma\sqrt{n}}T_n(r_n) \xrightarrow{\mathcal{D}} B.$$

Ist die Annahme der Unabhängigkeit oder der identischen Verteilung jedoch verletzt, so wirkt sich dies auf die Grenzverteilung des Residuen-Partialsummenprozesses aus. Der Grenzprozess ist dann im Allgemeinen keine Brown'sche Bewegung. Man vergleiche hierzu Bischoff (1998) zum Partialsummenprozess der LS-Residuen.

Kapitel 6. Testverfahren

6.1 Statistische Tests

Häufig interessiert uns an beobachteten Daten eine ganz bestimmte Eigenschaft. Wir wollen z.B. wissen, ob zwei unabhängige Stichproben $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ und $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{N}(0, \tau^2)$ die gleich Varianz aufweisen. Wir formulieren daher eine entsprechende Hypothese H_0 und eine geeignete Alternative H_1. In unserem Beispiel ist $H_0 : \sigma^2 = \tau^2$ und als Alternative könnte $H_1 : \sigma^2 \neq \tau^2$ gewählt werden, oder, falls $\sigma^2 < \tau^2$ ausgeschlossen werden kann,

$$H_1' : \sigma^2 > \tau^2 \ (\Leftrightarrow \ \theta := \frac{\sigma^2}{\tau^2} > 1). \tag{6.1}$$

Nun sind wir an einem Kriterium interessiert, das uns erlaubt, aufgrund der Beobachtungsvektoren $x = (x_1, \ldots, x_n)^\top$ und $y = (y_1, \ldots, y_m)^\top$ die Hypothese H_0 als plausibel oder als nicht plausibel zu betrachten. Ein *statistischer Test* ist ein solches Kriterium. Genauer gesagt, ist er eine Entscheidungsvorschrift δ, die den realisierten Zufallsvektoren entweder die Entscheidung

d_1 : verwirf H_0 (die Hypothese ist nicht plausibel)

oder die Entscheidung

d_0 : verwirf H_0 nicht (die Beobachtungen widersprechen H_0 nicht)

zuordnet. In der angewandten Statistik ist ein statistischer Test i.d. Regel von der Form

$$\delta = \delta(x, y) = \begin{cases} d_1, & \text{falls } T(x, y) > c \\ d_0, & \text{falls } T(x, y) \leq c \end{cases}, \ c \in \mathbb{R}, \tag{6.2}$$

wobei $T = T(x, y)$ eine geeignete *Teststatistik* ist. In obigem Beispiel könnte die Teststatistik z.B. gegeben sein durch

$$T(x, y) = \frac{\sum_{i=1}^n x_i^2}{\sum_{i=1}^m y_i^2}.$$

Bei der Anwendung eines statistischen Tests sind zwei Arten von Fehlern möglich. Zum Einen kann der Test die Hypothese verwerfen, obwohl sie vorliegt. Man nennt eine solche Fehlentscheidung einen *Fehler erster Art*. Zum Anderen ist es möglich, dass der Test eine vorliegende Alternative nicht erkennt. Man spricht dann von einem *Fehler zweiter Art*.

Selbstverständlich ist man daran interessiert, dass der gewählte Test die richtige Entscheidung trifft. Daher versucht man, die Wahrscheinlichkeiten für die Fehler erster und zweiter Art möglichst gering zu halten. Da man jedoch nicht beide Fehler gleichzeitig kontrollieren kann, lässt man zunächst die Wahrscheinlichkeit für den Fehler erster Art ein gewisses Niveau nicht überschreiten.

6.1.1 Definition und Bemerkung: Das *Niveau* α eines Tests beschreibt die (max.) Wahrscheinlichkeit, mit der unter Vorliegen der Nullhypothese irrtümlich die Alternative erkannt, also die Hypothese verworfen, wird. Das heißt, α ist die Wahrscheinlichkeit für den Fehler erster Art. Das Niveau wird vor der Durchführung eines Tests fest gewählt. In (6.2) z.B. entspricht die Wahl des Niveaus α der Wahl eines zu α korrespondierenden Wertes $c = c(\alpha)$. Üblicherweise setzt man $c(\alpha) = Q_{T;1-\alpha}$, wobei $Q_{T;1-\alpha}$ das $(1-\alpha)$-Quantil der H_0-Verteilung von T bezeichnet.

6.1.2 Definition: Ist X eine Zufallsvariable mit Verteilungsfunktion F_X, so haißt

$$Q_{X;p} := \inf\{x \in \mathbb{R} : F_X(x) \geq p\}$$

das *p-Quantil* von F_X (bzw. von X).

6.1.3 Bemerkung: Statistikpakete geben als Testresultat den sogenannten *p-Wert* aus. Er gibt das bestmögliche Niveau (also das kleinste α) an, zu dem der Test bei einem vorliegenden Datensatz die Hypothese gerade noch verwirft. Überschreitet der p-Wert das vorgegebene Niveau α, wird H_0 **nicht** verworfen, ansonsten verworfen.

Man bestimmt also zunächst das Niveau α eines Tests und kontrolliert auf diese Weise den Fehler erster Art. Anschließend versucht man, unter allen zur Verfügung stehenden Tests zum Niveau α einen Test mit einer möglichst geringen Wahrscheinlichkeit für einen Fehler zweiter Art zu wählen. Eine geringe Wahrscheinlichkeit für einen Fehler zweiter Art entspricht einer hohen *Güte* des Tests.

6.1.4 Definition: Die Wahrscheinlichkeit, mit der unter Vorliegen einer konkreten Alternative $\theta \in H_1$ (wie z.B. in (6.1)) die richtige Entscheidung getroffen, d.h. die Hypothese verworfen wird, heißt *Güte* des Tests unter Vorliegen der Alternative θ. Damit ist die Güte eines Tests gegeben durch $1 - \beta(\theta)$, wobei $\beta(\theta)$ die Wahrscheinlichkeit für den Fehler zweiter Art bei Vorliegen der Alternative θ bezeichnet.

Existiert in einer Menge Δ von Tests für H_0 gegen H_1 ein Test δ_0, dessen Güte für alle in Frage kommenden Werte $\theta \in H_1$ mindestens so hoch ist wie die Güte aller anderen Tests $\delta \in \Delta$, so nennt man δ_0 *gleichmäßig besten* (*uniformly most powerful*, kurz *UMP*) Test in Δ.

Verwirft ein Test δ die Hypothese H_0 zum Niveau α, so gilt die Alternative H_1 zum betreffenden Niveau als statistisch signifikant nachgewiesen. Verwirft der Test die Hypothese nicht, kann **keine** statistisch relevante Aussage getroffen werden, denn die Ursache für diese Testentscheidung kann darin liegen, dass H_0 tatsächlich vorliegt, oder auch darin, dass der Test eine zu geringe Güte besitzt, um eine vorliegende Alternative nachweisen zu können. (Z.B. kann der Stichprobenumfang zu gering sein.)

Statistisch signifikante Aussagen können also nur durch ein Verwerfen der Hypothese erhalten werden. Daher ist es wichtig, dass die statistisch nachzuweisende Eigenschaft (falls möglich) in der Alternative formuliert wird.

6.2 Tests auf *IID*-Verteilung

Hat man mittels linearer Regression oder durch Anpassung eines *ARMA*-Prozesses ein Modell an vorliegende Daten angepasst, so muss man noch testen, ob es geeignet ist, die den Daten zugrunde liegenden Phänomene zu beschreiben. Man spricht in diesem Zusammenhang von *Goodness-of-Fit-Tests* oder *Anpassungstests*. Ist ein Modell geeignet spezifiziert, so sind die entsprechenden Residuen annähernd White Noise oder sogar annähernd identisch und unabhängig verteilt. Dies gilt insbesondere für unkorrelierte Residuen eines linearen Modells (vgl. Abschnitt 5.2.2). Üblicherweise testet man daher die Anpassung eines Modells, indem man die entsprechenden Residuen einem Test auf *IID*-Verteilung unterzieht.

6.2.1 Bemerkung: Nach Anpassung an ein *ARMA*-Modell werden Residuen definiert durch

$$\hat{R}_t = \frac{X_t - \hat{X}_t(\hat{\phi}, \hat{\theta})}{\sqrt{\hat{w}_{t-1}(\hat{\phi}, \hat{\theta})}},$$

wobei $\hat{X}_1 := 0$, und die Vorhersagen $\hat{X}_t$, $t = 2, \ldots, n$, sowie die mittleren quadratischen Abweichungen $\hat{w}_{t-1}$, $t = 1, \ldots, n$, mit Hilfe des Innovationsalgorithmus berechnet werden (vgl. (3.26) und (3.27). Hierbei ist zu beachten, dass die Schätzung der Parameter $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ und w_{t-1} den Verlust von $p + q + 1$ Freiheitsgraden bedeutet. Man vergleiche auch Bemerkung 5.2.10 zum Verlust von Freiheitsgraden bei Residuenbildung in linearen Modellen.

6.2.1 Tests auf der Basis der empirischen Autokorrelationsfunktion

Um die Hypothese $H_0 : (X_t) \sim \mathit{IID}$ gegen korrelierte Alternativen zu verwerfen, genügt es zu zeigen, dass $\rho(h) \neq 0$ ist für ein (fest gewähltes) $h \geq 1$. In der Regel wählt man zu diesem Zweck $h = 1$, da man davon ausgehen kann, dass benachbarte Beabachtungen am stärksten korreliert sind. In manchen Fällen kann es aber auch sinnvoll sein, sich für ein anderes $h \in \mathbb{N}$ zu entscheiden, z.B. $h = 12$, wenn bei monatlichen Daten ein Jahreszyklus vermutet wird.

Ein Test für die Hypothese $\rho(h) = 0$

Nach Satz 3.2.10 ist unter $H_0 : (X_t) \sim \mathit{IID}$ der Schätzer $\hat{\rho}(h)$ für $h \geq 1$ asymptotisch standard-normalverteilt, d.h. $\sqrt{n}\hat{\rho}(h) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. Damit erhält man sofort den Signifikanztest zum Niveau α

$$\text{Verwirf } H_0 \iff \sqrt{n}|\hat{\rho}(h)| > \Phi_{1-\alpha/2} \ .$$

Dabei bezeichnet $\Phi_{1-\alpha/2}$ das $(1 - \frac{\alpha}{2})$-Quantil der Standard-Normalverteilung.

Von Neumann Ratio

Gegeben seien n Beobachtungen der Zeitreihe (X_t). Dann heißt die Größe

$$VNR := \frac{\frac{1}{n-1}\sum_{j=2}^{n}(X_j - X_{j-1})^2}{\frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2},$$

mit $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, der *von Neumann Ratio* (VNR) von $X_1, \ldots, X_n$. Wegen

$$X_j - X_{j-1} = (X_j - \overline{X}) - (X_{j-1} - \overline{X})$$

gilt

$$\begin{aligned}\frac{n-1}{n}\,\hat{\gamma}(0)\cdot VNR &= \frac{1}{n-1}\sum_{j=2}^{n}(X_j - \overline{X})^2 + \frac{1}{n-1}\sum_{j=1}^{n-1}(X_j - \overline{X})^2 \\ &\quad - \frac{2}{n-1}\sum_{j=1}^{n-1}(X_j - \overline{X})(X_{j+1} - \overline{X}).\end{aligned}$$

Somit ist

$$VNR \approx 2 - 2\hat{\rho}(1).$$

Ein White Noise Prozess besitzt also einen von Neumann Ratio mit Erwartungswert 2 (approximativ). Sind die Beobachtungen stark positiv (negativ) korreliert, gilt genauer $\rho(1) = 1$ (bzw. $\rho(1) = -1$), so besitzt der entsprechende von Neumann Ratio den approximativen Erwartungswert 0 (bzw. 4). Hart (1942) hat für die Stichprobenumfänge $4, \ldots, 60$ und verschiedene Signifikanzniveaus obere bzw. untere Konfidenzschranken für den von Neumann Ratio berechnet. Einige dieser Werte sind in Tabelle 6.1 aufgeführt ($\alpha = 0.05$). Man verwirft demnach die Hypothese $H_0 : \rho(1) = 0$ zugunsten der positiv korrelierten Alternative

$$H_1 : \rho(1) > 0$$

genau dann zum Niveau α, wenn der VNR die untere Konfidenzschranke unterschreitet. Entsprechend verwirft man die Hypothese zugunsten der negativ korrelierten Alternative

$$H_1' : \rho(1) < 0$$

genau dann zum Niveau α, wenn der VNR die obere Konfidenzschranke überschreitet.

Tabelle 6.1: Einige Konfidenzschranken für den von Neumann Ratio ($\alpha = 0.05$)

n	untere Schranke	obere Schranke	n	untere Schranke	obere Schranke
5	1.0255	3.9745	35	1.5014	2.6163
10	1.1803	3.2642	40	1.5304	2.5722
15	1.2914	2.9943	45	1.5552	2.5357
20	1.3680	2.8425	50	1.5752	2.5064
25	1.4241	2.7426	55	1.5923	2.4819
30	1.4672	2.6707	60	1.6082	2.4596

Portmanteau-Test von Ljung und Box

Es seien $X_1, \ldots, X_N$ Beobachtungen des Prozesses (X_t) und $\hat{\rho}(\cdot)$ die empirische Autokorrelationsfunktion von (X_t). Nach Satz 3.2.10 strebt der Vektor

$$\sqrt{N}\cdot\hat{\rho}_h := \sqrt{N}\cdot(\hat{\rho}(1), \ldots, \hat{\rho}(h))^\top$$

unter den dortigen Voraussetzungen und unter

$$H_0 : \ (X_t) \sim \mathit{IID}(0, \sigma^2)$$

gegen eine Normalverteilung mit Erwartungswertvektor 0 und Kovarianzmatrix I_h. Nach dem Abbildungssatz B.4.12 ist somit

$$Q := N \cdot \hat{\rho}_h^\top \hat{\rho}_h = N \sum_{j=1}^{h} \hat{\rho}^2(j)$$

unter H_0 asymptotisch χ_h^2-verteilt. Es ergibt sich unmittelbar der sogenannte *Portmanteau-Test* von Box und Pierce (1970)

$$\text{Verwirf } H_0 \iff Q > \chi^2_{h;1-\alpha}.$$

Dabei bezeichnet $\chi^2_{h;1-\alpha}$ das $(1-\alpha)-$Quantil der χ_h^2-Verteilung.

6.2.2 Bemerkung: Bei Anwendungen hat sich gezeigt, dass der Portmanteau-Test bei schlechtem Fit H_0 häufig nicht verwirft (vgl.Brockwell / Davis (1991) sowie Ljung und Box (1978)). Die Ursache hierfür ist, dass Satz 3.2.10 lediglich eine **asymptotische** Aussage macht. Bei endlichen, insbesondere bei kleineren Stichprobenumfängen gilt für die empirische Autokorrelationsfunktion $\hat{\rho}(\cdot)$ eines White Noise Prozesses (vgl. Box und Pierce (1970))

$$\text{Var}(\hat{\rho}(h)) = \frac{N-h}{N(N+2)},\ h \in \mathbb{N}.$$

Ljung und Box ersetzen daher in ihrem Artikel (1978) die Testgröße Q durch

$$\tilde{Q} := N \sum_{j=1}^{h} \frac{N+2}{N-j} \cdot \hat{\rho}^2(j).$$

Kritiker merken an, dass die Teststatistik $\tilde{Q}$ eine größere Varianz als eine χ_h^2-verteilte Zufallsvariable besitzt. Dennoch ist sie wegen besserer empirischer Ergebnisse der Testgröße Q vorzuziehen (s. Ljung / Box (1978)).

6.2.3 Bemerkung: Aufgrund der Korreliertheit der LS-Residuen sollten für die Tests in Abschnitt 6.2 unkorrelierte Residuen eines linearen Modells vorgezogen werden. Wendet man den Portmanteu-Test dennoch auf LS-Residuen an, sollte man bei der χ^2-Verteilung der Teststatistik den Verlust von Freiheitsgraden durch Schätzung der Parameter berücksichtigen. Entsprechendes gilt für die Residuen nach Anpassung eines $ARMA(p,q)$-Modells. Man vergleiche hierzu die Bemerkungen 5.2.10 und 6.2.1.

6.2.4 Beispiel: Die geodätische Zeitreihe aus Abbildung 3.13 und deren Residuen nach Anpassung an ein $ARMA(3,3)$-Modell (Abb. 3.16) werden dem Portmanteau-Test nach Ljung und Box unterzogen. Man erhält dabei einen p-Wert < 0.0001 bzw. 0.9693. Somit wird die Hypothese eines White Noise Prozesses für die Zeitreihe aus Abbildung 3.13 eindeutig verworfen. Der Annahme eines WN-Prozesses für die Residuen aus Abbildung 3.16 steht hingegen nichts im Wege. Das in Beispiel 3.4.10 für die Daten aus Abbildung 3.13 geschätzte $ARMA(3,3)$-Modell kann somit als geeignet betrachtet werden.

6.2.2 Tests auf der Basis der empirischen Spektraldichte

Nach den Bemerkungen 4.1.17 und 4.1.2 ist die stationäre Zeitreihe (X_t) genau dann ein White Noise Prozess, wenn ihre normierte Spektraldichte f konstant gleich $\frac{1}{2\pi}$ auf dem Intervall $[-\pi,\pi]$ ist. Zum Testen der Hypothese

$$H_0 :\ (X_t) \sim IID(0,\sigma^2)$$

gegen die Alternative

$$H_1 :\ (X_t) \textit{ ist kein IID-Prozess}$$

können wir uns also, alternativ zur empirischen Autokorrelationsfunktion, der empirischen Spektraldichte bedienen. Zu diesem Zweck definieren wir ein über den positiven Frequenzbereich integriertes Spektrum

$$F^+(\omega) := 2\int_0^\omega f(\lambda)d\lambda,\ \omega \in [0,\pi].$$

Man beachte, dass unter H_0 gilt

$$F^+(\omega) := 2\int_0^\omega f(\lambda)d\lambda = \frac{\omega}{\pi},\ \omega \in [0,\pi].$$

Ein Schätzer für die normierte Spektraldichte f ist (NB $e^{i\theta} = \cos(\theta) + i\sin(\theta)$)

$$\begin{aligned}
\hat{f}_N(\omega) &:= \frac{1}{\hat{\gamma}(0)} I_N^*(\omega) \\
&= \frac{1}{2\pi} \sum_{h=-(N-1)}^{N-1} \hat{\rho}(h) e^{-i\omega h} \\
&\overset{\hat{\rho}\ gerade}{=} \frac{1}{2\pi}\underbrace{\hat{\rho}(0)}_{=1} + \frac{1}{\pi}\sum_{h=1}^{N-1} \hat{\rho}(h)\cos(\omega h),\ \omega \in [-\pi,\pi]. \qquad (6.3)
\end{aligned}$$

Damit ist ein naheliegender Schätzer für $F^+(\omega)$

$$\begin{aligned}
\widehat{F_N^+}(\omega) &:= 2\int_0^\omega \hat{f}_N(\lambda)d\lambda \\
&\overset{(6.3)}{=} \frac{\omega}{\pi} + \frac{2}{\pi}\sum_{h=1}^{N-1} \hat{\rho}(h)\frac{\sin(\omega h)}{h}\ ,\ \omega \in [0,\pi]. \qquad (6.4)
\end{aligned}$$

Unter der Hypothese H_0 gilt wegen $\mathrm{E}(\hat{\rho}(h)) = \rho(h) = 0$ für $h \geq 1$

$$\begin{aligned}
\mathrm{E}(\widehat{F_N^+}(\omega)) &= \frac{\omega}{\pi} + \frac{2}{\pi}\sum_{h=1}^{N-1} \underbrace{\mathrm{E}(\hat{\rho}(h))}_{=0} \frac{\sin(\omega h)}{h} \\
&= \frac{\omega}{\pi}\ ,\ \omega \in [0,\pi],
\end{aligned}$$

bzw.

$$\mathrm{E}(\widehat{F_N^+}(\pi z)) = z \text{ für } z \in [0,1].$$

Definiere nun für $z \in [0,1]$

$$\begin{aligned}
B_N(z) &:= \sqrt{\frac{N}{2}}\left(\widehat{F_N^+}(\pi z) - z\right) \\
&\overset{(6.4)}{=} \sqrt{\frac{N}{2}}\left(z + \frac{2}{\pi}\sum_{h=1}^{N-1} \hat{\rho}(h)\frac{\sin(\pi z \cdot h)}{h} - z\right) \\
&= \sqrt{2N}\cdot\frac{1}{\pi}\sum_{h=1}^{N-1} \hat{\rho}(h)\frac{\sin(\pi z \cdot h)}{h}. \qquad (6.5)
\end{aligned}$$

Nach dem folgenden Satz 6.2.5 strebt der so definierte stochastische Prozess B_N nach Verteilung gegen eine Brown'sche Brücke Zur Definition der Verteilungskonvergenz s. Anhang B.4.

6.2.5 Satz: *Ist $(X_t) \sim IID(0,\sigma^2)$ mit $E(X_t^8) < \infty$, so gilt*

$$B_N \xrightarrow{\mathcal{D}} B_0 \text{ in } C[0,1], \qquad (6.6)$$

wobei B_0 die Brown'sche Brücke bezeichnet.

Beweis: Nach Voraussetzung ist (X_t) ein linearer Prozess $X_t = \sum_{j=-\infty}^{\infty}\psi_j Z_{t-j}$, $(Z_t) \sim IID(0,\sigma^2)$ mit

$$\psi_j = \begin{cases} 1, & j = 0, \\ 0, & j \in \mathbb{Z}/\{0\}. \end{cases}$$

Die Voraussetzungen von Satz 3.2.10 sind somit erfüllt und es ergibt sich

$$\sqrt{N}(\hat{\rho}(1),\ldots,\hat{\rho}(k))^\top \xrightarrow{\mathcal{D}} \mathcal{N}((0,\ldots,0)^\top, I_k). \tag{6.7}$$

Man schreibe nun $B_N(z) = B_N^k(z) + R_N^k(z)$ $(k \in \{1,\ldots,N-1\})$ mit

$$B_N^k(z) := \frac{\sqrt{2}}{\pi}\sum_{h=1}^{k}\sqrt{N}\hat{\rho}(h)\frac{\sin(\pi z\cdot h)}{h}$$

und

$$R_N^k(z) := \frac{\sqrt{2}}{\pi}\sum_{h=k+1}^{N-1}\sqrt{N}\hat{\rho}(h)\frac{\sin(\pi z\cdot h)}{h}.$$

Setzt man in Satz B.4.12 (Abbildungssatz) $S = \mathbb{R}^k$ sowie $S' = C[0,1]$ (dabei bezeichntet $C[0,1]$ den Raum aller auf dem Intervall $[0,1]$ stetigen Funktionen), so erhält man mit (6.7) für festes $k \in \mathbb{N}$

$$B_N^k \xrightarrow{\mathcal{D}} B_0^k \ (N\to\infty) \text{ in } C[0,1],$$

wobei

$$B_0^k(z) := \frac{\sqrt{2}}{\pi}\sum_{h=1}^{k} Y_h\frac{\sin(\pi z\cdot h)}{h},\ Y_1,\ldots,Y_k \overset{iid}{\sim} \mathcal{N}(0,1). \tag{6.8}$$

Für $k\to\infty$ erhält man in (6.8)

$$\frac{\sqrt{2}}{\pi}\sum_{h=1}^{\infty} Y_h\frac{\sin(\pi z\cdot h)}{h},\ Y_1,Y_2,\cdots \overset{iid}{\sim} \mathcal{N}(0,1),\ z\in[0,1], \tag{6.9}$$

was eine alternative Darstellung für die Brown'sche Brücke ist (vgl. Tanaka (1996)).

Für R_N^k gilt unter der Bedingung $\mathrm{E}(X_t^8) < \infty$ (vgl. Grenander / Rosenblatt (1957), S.189)

$$\max_{z\in[0,1]}|R_N^k(z)| \xrightarrow{P} 0,\ N,k\to\infty. \tag{6.10}$$

Aus (6.9) und (6.10) ergibt sich mit Satz B.4.15 (Appendix B, $S = C[0,1]$)

$$B_N \xrightarrow{\mathcal{D}} B_0 \text{ in } C[0,1].$$

□

Ein Kolmogorov-Smirnov-Test

Unter der Hypothese $H_0 : (X_t) \sim IID(0,\sigma^2)$ ist f konstant auf $[-\pi,\pi]$ und somit $F^+(\pi z) = z$. Damit beschreibt die Größe

$$\max_{z\in[0,1]}|B_N(z)| \tag{6.11}$$

unter H_0 den maximalen Abstand von $\widehat{F_N^+}$ zu $\mathrm{E}(\widehat{F_N^+}) = F^+$, multipliziert mit einem Faktor $\sqrt{\frac{N}{2}}$. Der Wert (6.11) sollte also unter H_0 "relativ klein" sein und liefert somit einen Anhaltspunkt, ob die Hypothese aufgrund der beobachteten Daten plausibel ist, oder ob nicht.

Aus Satz 6.2.5 und dem Abbildungssatz folgt unter H_0 unmittelbar

$$\max_{z\in[0,1]}|B_N(z)| = \max_{z\in[0,1]}\left|\sqrt{\frac{N}{2}}\left(\widehat{F_N^+}(\pi z) - z\right)\right| \xrightarrow{\mathcal{D}} \max_{z\in[0,1]}|B_0(z)|.$$

Die Verteilung der Zufallsvariablen

$$\max_{z\in[0,1]}|B_0(z)|$$

heißt *Kolmogorov-Verteilung*. Es gilt (vgl. FELLER (1948))

$$P\left(\max_{z\in[0,1]}|B_0(z)|\leq\lambda\right)=1-2\sum_{j=1}^{\infty}(-1)^{j-1}e^{-2j^2\lambda^2}=\sum_{j=-\infty}^{\infty}(-1)^j e^{-2j^2\lambda^2},\ \lambda>0.$$

Einen Signifikanztest zum Niveau $\alpha>0$ erhält man nun durch die Vorschrift

$$\text{Lehne } H_0:(X_t)\sim IID(0,\sigma^2) \text{ ab } \Longleftrightarrow T>K_{1-\alpha} \tag{6.12}$$

mit $T=\max_{z\in[0,1]}\left|\sqrt{\frac{N}{2}}\left(\widehat{F_N^+}(\pi z)-z\right)\right|$ und dem $(1-\alpha)$-Quantil der Kolmogorov-Verteilung $K_{1-\alpha}$. Einige Quantile der Kolmogorov Verteilung sind in Tabelle 6.2 aufgelistet.

Tabelle 6.2: Die wichtigsten $(1-\alpha)$-Quantile der Kolmogorov-Verteilung

α	0.05	0.025	0.01	0.005	0.001
$K_{1-\alpha}$	1.3581	1.4802	1.6276	1.7308	1.9495

Anmerkung: Dieser Test ist in Anlehnung an den *Kolmogorov-Smirnov-Test* für Wahrscheinlichkeits-Verteilungsfunktionen (vgl. Abschnitt 6.3) entstanden. Literatur: BARTLETT (1966), ANDERSON (1993).

Ein Cramér-von Mises-Test

Sind x und y Elemente eines normierten Raumes E, versehen mit einer Norm $\|\cdot\|$, so beschreibt $\|x-y\|$ den *Abstand* von x und y in E (vgl. HEUSER (1992)).

Nach der Definition der Maximumsnorm auf $C[0,1]$ entspricht die Kolmogorov-Smirnov-Teststatistik T aus Abschnitt 6.2.2 der Zufallsvariablen $\left\|\sqrt{\frac{N}{2}}\left(\widehat{F_N^+}(\pi z)-z\right)\right\|$, wobei $\|\cdot\|$ die Maximumsnorm auf $C[0,1]$ ist. Somit misst die Kolmogorov-Smirnov-Teststatistik den Abstand zwischen der Spektral-Verteilungsfunktion F^+ unter H_0 und der empirischen Spektral-Verteilungsfunktion $\widehat{F_N^+}$ bezüglich der Maximumsnorm. Als Funktionen, die auf einem abgeschlossenen Intervall definiert und dort stetig sind, sind F^+ und $\widehat{F_N^+}$ auch quadratisch integrierbar. Man kann also auch anhand ihres Abstandes in $L^2(0,1)$ entscheiden, ob die Daten der Hypothese H_0 widersprechen, oder ob nicht. Den Abstand in $L^2(0,1)$ misst die *Cramér-von Mises-Teststatistik*

$$T:=\frac{N}{2}\int_0^1\left(\widehat{F_N^+}(\pi z)-z\right)^2dz. \tag{6.13}$$

Die Verteilung der Teststatistik (6.13) konvergiert nach dem Abbildungssatz schwach gegen die Verteilung der Zufallsvariablen

$$\int_0^1 B_0^2(z)dz. \tag{6.14}$$

Einige Quantile der Verteilung (6.14) sind in Tabelle 6.3 aufgelistet. Die Werte in der Tabelle sind ANDERSON / DARLING (1952) entnommen.

Tabelle 6.3: Die wichtigsten $(1-\alpha)$-Quantile für den Cramér-von Mises-Test

α	0.1	0.05	0.02	0.01	0.001
$C_{1-\alpha}$	0.34730	0.46136	0.61981	0.74346	1.16786

Ein Test zum Niveau α ist also

$$\text{Lehne } H_0:(X_t)\sim IID(0,\sigma^2) \text{ ab } \Longleftrightarrow T>C_{1-\alpha}.$$

Anmerkung: Dieser Test entspricht dem *Cramér-von Mises-Test* für Wahrscheinlichkeits-Verteilungsfunktionen (vgl. Abschnitt 6.3).

Fisher-Test

An dieser Stelle ist noch ein Test von Fisher erwähnenswert, der bereits 1924 entwickelt wurde. Er testet die Hypothese

$$H_0 : (X_t) \textit{ ist ein Gaußverteilter White Noise Prozess}$$

gegen die Alternative

$$H_1 : (X_t) \textit{ enthält eine zyklische Komponente.}$$

Zu N Beobachtungen des Prozesses (X_t) liege das Periodogramm $\{I_N(\omega_i),\ i = 1, \ldots, n\}$ vor mit $n := [\frac{N-1}{2}]$. Dann misst die Testgröße

$$Y_n := \frac{\max_{i=1}^n I_N(\omega_i)}{\frac{1}{n}\sum_{i=1}^n I_N(\omega_i)}$$

das Verhältnis zwischen dem größten Peak und dem durchschnittlichen Wert des Periodogramms. Ist dieses Verhältnis "zu groß", so muss H_0 verworfen werden. Mit der Bezeichnung $x_+ := \max\{x, 0\}$, $x \in \mathbb{R}$, gilt unter H_0 (vgl. Brockwell / Davis (1991))

$$P(Y_n \leq a) = \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left[\left(1 - \frac{ja}{n} \right)_+ \right]^{n-1}, \ a > 0.$$

Bezeichnet y_n eine Realisierung der Statistik Y_n, so berechnet man

$$P(Y_n \geq y_n) = 1 - \sum_{j=0}^{n} (-1)^j \binom{n}{j} \left[\left(1 - \frac{jy_n}{n} \right)_+ \right]^{n-1}.$$

Ist diese Wahrscheinlichkeit kleiner als α, wird der Fisher-Test zum Niveau α verworfen.

6.2.3 Tests auf der Basis des Residuen-Partialsummenprozesses

Will man anhand des Residuen-Partialsummenprozesses aus Abschnitt 5.3 Rückschlüsse auf die Verteilung der Residuen ziehen, so besteht nach Bemerkung 5.3.2 die Möglichkeit, die Hypothese

$$H_0 : \textit{ der Residuen} - \textit{Partialsummenprozess strebt gegen eine Brown}'\textit{sche Bewegung}$$

gegen die Alternative

$H_1 :$ *H_0 gilt nicht*

zu testen. Wird H_0 dabei verworfen, so kann man auch die Hypothese *iid*-verteilter Residuen verwerfen.

Zur Berechnung des Residuen-Partialsummenprozesses wird die Bildung unkorrelierter Residuen empfohlen (vgl. Abschnitt 5.2.2). Außerdem muss die Standardabweichung σ aus (5.28) konsistent geschätzt werden. Sind die Residuen des Linearen Modells (5.3) *iid*-verteilt, gilt also H_0, so lässt sich die Varianz σ^2 des Fehlervektors konsistent schätzen durch

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (r_{ni} - \overline{r}_{ni})^2 \tag{6.15}$$

mit $\overline{r}_{ni} = \frac{1}{n}\sum_{i=1}^n r_{ni}$. Man setze nun $\hat{\sigma}_n = \sqrt{\hat{\sigma}_n^2}$. Dann folgt mit Lemma B.4.14

$$\frac{1}{\hat{\sigma}_n \sqrt{n}} T_n(r_n) \xrightarrow{\mathcal{D}} B, \tag{6.16}$$

d.h. der Residuen-Partialsummenprozess strebt nach Verteilung gegen eine Brown'sche Bewegung.

Zum Testen der Hypothese H_0 gibt es nun mehrere Möglichkeiten:

Ein Test basierend auf der Kolmogorov-Verteilung

Nach dem Satz von Donsker und dem Abbildungssatz (s. Anhang B) konvergiert der Prozess

$$M_n(z) := \frac{1}{\hat{\sigma}_n\sqrt{n}}\left(T_n(r_n)(z) - z \cdot T_n(r_n)(1)\right) \tag{6.17}$$

(mit $\hat{\sigma}_n$ wie oben) im Falle *iid*-verteilter Residuen schwach gegen eine Brown'sche Brücke $B_0(z)$. An den Stellen $t_{nk},\ k = 1, \ldots, n$, ist (6.17) gegeben durch

$$M_{n,k} := \frac{1}{\hat{\sigma}_n\sqrt{n}}\left(\sum_{i=1}^{[nt_{nk}]} r_{ni} - t_{nk} \cdot \sum_{i=1}^{n} r_{ni}\right).$$

Die Zufallsvariable

$$\sup_{z\in[0,1]} |B_0(z)|$$

besitzt eine Kolmogorov-Verteilung (vgl. Abschnitt 6.2.2). Somit ist ein Signifikanztest zum Niveau $\alpha > 0$ gegeben durch die Vorschrift

$$\text{Lehne } H_0 \text{ ab} \iff T_n > K_{1-\alpha}\ .$$

Dabei ist $T_n = \max_{k=1}^{n} |M_{n,k}|$ und $K_{1-\alpha}$ das $(1-\alpha)$-Quantil der Kolmogorov-Verteilung aus Tabelle 6.2.

Cramér-von Mises

Da der Prozess (6.17) nach Verteilung gegen eine Brown'sche Brücke konvergiert, erhält man mit dem Abbildungssatz

$$T_n := \int_0^1 (M_n(z))^2 dz \xrightarrow{\mathcal{D}} \int_0^1 B_0^2(z)dz. \tag{6.18}$$

Die Verteilung der Zufallsvariablen $\int_0^1 B_0^2(z)dz$ ist uns schon in Abschnitt 6.2.2 begegnet, als Grenzwert (bzgl. Verteilungskonvergenz) der Cramér-von Mises-Teststatistik. Wir erhalten daher sofort den α-Niveau-Test

$$\text{Lehne } H_0 \text{ ab} \iff T_n > C_{1-\alpha}\ ,$$

mit dem $(1-\alpha)$-Quantil $C_{1-\alpha}$ aus Tabelle 6.3.

Die Verteilung im Punkt $z = 1$

Der Kürze halber bezeichne im Folgenden $R_n(\cdot) := \frac{1}{\hat{\sigma}_n\sqrt{n}} T_n(r_n)(\cdot)$ den Residuen-Partialsummenprozess mit geschätzter Standardabweichung.

Wie bereits erwähnt wurde, strebt der Residuen-Partialsummenprozess im Falle *iid*-verteilter Residuen gegen eine Brown'sche Bewegung. Aus Lemma 3.5.9 folgt direkt, dass die Brown'sche Bewegung im Punkt $z = 1$ eine $\mathcal{N}(0,1)$-Verteilung besitzt. Sind nun $R_n^{(1)}(\cdot), \ldots, R_n^{(s)}(\cdot)$ s unabhängige Beobachtungen des Residuen-Partialsummenprozesses (z.B. Beobachtungen desselben Prozesses an s verschiedenen Tagen), dann strebt die Zufallsvariable

$$T := \left(R_n^{(1)}(1)\right)^2 + \cdots + \left(R_n^{(s)}(1)\right)^2$$

unter der Voraussetzung *iid*-verteilter Residuen schwach gegen eine χ^2-Verteilung mit s Freiheitsgraden (s. Anhang B.2). Man beachte, dass für $l = 1, \ldots, s$ gilt

$$R_n^{(l)}(1) = \frac{1}{\hat{\sigma}_n^{(l)}\sqrt{n}} T_n(r_n^{(l)})(1) = \frac{1}{\hat{\sigma}_n^{(l)}\sqrt{n}} \sum_{i=1}^{n} r_{ni}^{(l)}.$$

Mit den obigen Aussagen ergibt sich der folgende χ^2-Test zum Niveau α:

$$\text{Lehne } H_0 \text{ ab} \Longleftrightarrow T > \chi^2_{s,1-\alpha} \ .$$

Dabei bezeichnet $\chi^2_{s,1-\alpha}$ das $(1-\alpha)$-Quantil der χ^2-Verteilung mit s Freiheitsgraden.

Man beachte, dass für jede Beobachtung $R_n^{(l)}(1),\ l = 1,\dots s$, die Standardabweichung σ separat geschätzt werden muss, um stochastisch unabhängige $R_n^{(1)}(1),\dots,R_n^{(s)}(1)$ zu erhalten. Die stochastische Unabhängigkeit gewährleistet dann die asyptotische χ^2-Verteilung der Teststatistik T.

Die Verteilung des Integrals

Da die Funktion $f(z) = z$ auf dem Intervall $[0,1]$ von beschränkter Variation und die Brown'sche Bewegung f.s. stetig ist, existiert nach Definition B.4.18 das Integral $\int_0^1 B(z)dz$ (f.s.). Nach Lemma B.4.20 ist damit

$$\int_0^1 B(z)dz \sim \mathcal{N}(0,\frac{1}{3}).$$

Sind die Residuen $r_{ni},\ i = 1,\dots n,\ n \in \mathbb{N}$, unabhängig und identisch verteilt, strebt also der Residuen-Partialsummenprozess nach Verteilung gegen eine Brown'sche Bewegung, so gilt für den integrierten Partialsummenprozess nach dem Abbildungssatz B.4.12

$$\int_0^1 R_n(z)dz \xrightarrow{\mathcal{D}} \mathcal{N}(0,\frac{1}{3}).$$

Oder gleichbedeutend

$$\sqrt{3}\cdot\int_0^1 R_n(z)dz \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Wie oben ergibt sich mit dem konsistenten Schätzer $\hat{\sigma}_n$ aus (6.15) der α-Niveau-Test

$$\text{Lehne } H_0 \text{ ab} \Longleftrightarrow T > \chi^2_{s,1-\alpha} \ ,$$

mit

$$T = (M_n^{(1)})^2 + \dots + (M_n^{(s)})^2.$$

Dabei sind $M_n^{(1)},\dots,M_n^{(s)}$ Beobachtungen des Prozesses

$$\sqrt{3}\cdot\int_0^1 \frac{1}{\hat{\sigma}_n^{(l)}\sqrt{n}} T_n(r_n^{(l)})(z)dz,\ l = 1,\dots,s.$$

Auch hier ist zu beachten, dass σ für jede Beobachtung $T_n^{(l)},\ l = 1,\dots,s$, separat geschätzt werden sollte.

Die Verteilung des Maximums / Minimums

6.2.6 Satz: *Es sei (B_z) eine reelle, normale Brown'sche Bewegung auf $[0,1]$. Dann gilt*

$$\begin{aligned} P\left(\sup_{z\in[0,1]} B_z \le \alpha\right) &= \tfrac{2}{\sqrt{2\pi}}\textstyle\int_0^\alpha e^{-\frac{1}{2}u^2}du, \quad \alpha \ge 0, \\ \text{und } P\left(\sup_{z\in[0,1]} B_z \le \alpha\right) &= 0, \qquad\qquad \alpha < 0. \end{aligned} \tag{6.19}$$

Beweis: Es seien $\xi_i,\ i \in \mathbb{N}$, unabhängig und identisch verteilt mit

$$P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2},$$

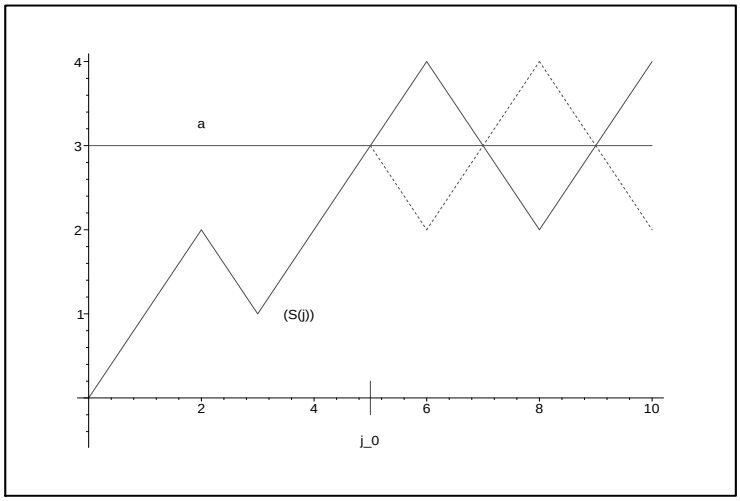

Abbildung 6.1: Das Spiegelungsprinzip (a=3)

sowie $a \in \mathbb{N}$. Weiter bezeichne $S_j = \sum_{i=1}^{j} \xi_i$ die j-te Partialsumme. Es ist

$$\begin{aligned} P(\max_{j=1}^{n} S_j \geq a) &= P(\max_{j=1}^{n} S_j \geq a, S_n < a) \\ &+ P(\max_{j=1}^{n} S_j \geq a, S_n = a) \\ &+ P(\max_{j=1}^{n} S_j \geq a, S_n > a). \end{aligned} \tag{6.20}$$

Man betrachte den Prozess $(S_j)_{j\in\mathbb{N}}$. Erreicht oder überschreitet dieser den Punkt $a \in \mathbb{N}$ mindestens einmal, so bezeichne j_0 den Zeitpunkt, an dem $(S_j)_{j\in\mathbb{N}}$ zum ersten Mal den Wert a erreicht. Sieht man nun den Prozess **bis** zum Zeitpunkt j_0 (einschließlich) als gegeben an, ab dem Zeitpunkt j_0 jedoch als zufällig, so ist die Wahrscheinlichkeit, dass S_n für $n > j_0$ einen Wert $> a$ annimmt, gleich der Wahrscheinlichkeit, dass S_n einen Wert $< a$ annimmt. Denn zu jedem Pfad mit $S_n > a$ gibt es genau einen Pfad mit $S_n < a$, der gewissermaßen die Spiegelung ist von $(S_j)_{j\in\mathbb{N}}$ an der Geraden, die zur Zeitachse parallel ist mit Abstand a (Spiegelungsprinzip, s. Abbildung 6.1).

Aus dem Spiegelungsprinzip folgt daher

$$\begin{aligned} P(\max_{j=1}^{n} S_j \geq a, S_n < a) &= P(\max_{j=1}^{n} S_j \geq a, S_n > a) \\ &= P(S_n > a). \end{aligned}$$

Das zweite Gleichheitszeichen gilt, da das Ereignis "$S_n > a$" nicht eintreten kann, falls das Ereignis "$\max_{j=1}^{n} S_j < a$" eintritt. Aus demselben Grund gilt

$$P(\max_{j=1}^{n} S_j \geq a, S_n = a) = P(S_n = a).$$

Also ergibt sich mit (6.20)

$$P(\max_{j=1}^{n} S_j \geq a) = 2P(S_n > a) + P(S_n = a). \tag{6.21}$$

Nun sei $\alpha \geq 0$ beliebig und a_n das größte Ganze **größer** oder gleich $\alpha\sqrt{n}$, d.h. $a_n = -[-\alpha\sqrt{n}]$. Nach dem Zentralen Grenzwertsatz (s. Anhang B) gilt

$$P(S_n > a_n) = P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i > \alpha\right) \longrightarrow P(N > \alpha) = P(N \geq \alpha),$$

wobei N eine $\mathcal{N}(0,1)$-verteilte Zufallsvariable ist. Da die Standard-Normalverteilung keine Masse im Punkt α besitzt, folgt

$$P(S_n = a_n) \longrightarrow 0.$$

Somit ergibt sich aus (6.21)

$$P\left(\max_{j=1}^{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_i \geq \alpha\right) \longrightarrow 2P(N \geq \alpha) = \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{1}{2}u^2} du, \ \alpha \geq 0.$$

Also gilt

$$\begin{aligned} P\left(\max_{j=1}^{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_i \leq \alpha\right) &\longrightarrow 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\alpha} e^{-\frac{1}{2}u^2} du - \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-\frac{1}{2}u^2} du \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\alpha} e^{-\frac{1}{2}u^2} du, \ \alpha \geq 0. \end{aligned} \tag{6.22}$$

Da der Prozess

$$\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_i\right)_{j \geq 0}$$

bei $j = 0$ den Wert 0 annimmt, ist

$$P\left(\max_{j=1}^{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_i \leq \alpha\right) = 0 \text{ für } \alpha < 0.$$

Andererseits gilt, da die $\xi_i, \ i = 1, \ldots, n$, die Voraussetzungen für den Satz von Donsker erfüllen,

$$X_n(z) := \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{[nz]} \xi_i + (nz - [nz])\xi_{[nz]+1}\right) \xrightarrow{\mathcal{D}} B.$$

Mit dem Abbildungssatz aus Anhang B ergibt sich

$$\sup_{z \in [0,1]} (X_n(z)) = \max_{j=1}^{n} \frac{1}{\sqrt{n}} \sum_{i=1}^{j} \xi_i \xrightarrow{\mathcal{D}} \sup_{z \in [0,1]} B_z. \tag{6.23}$$

Da der schwache Grenzwert eindeutig ist (s. Anhang B) folgt aus (6.22) und (6.23)

$$\begin{aligned} P\left(\sup_{z \in [0,1]} B_z \leq \alpha\right) &= \tfrac{2}{\sqrt{2\pi}} \textstyle\int_0^{\alpha} e^{-\frac{1}{2}u^2} du, \quad \alpha \geq 0, \\ \text{und } P\left(\sup_{z \in [0,1]} B_z \leq \alpha\right) &= 0, \qquad \alpha < 0. \end{aligned}$$

□

Betrachten wir nun den Partialsummenprozess *iid*-verteilter Residuen. Da auch dieser schwach gegen eine Brown'sche Bewegung konvergiert, erhält man (wiederum mit dem Abbildungssatz)

$$P\left(\sup_{z \in [0,1]} \frac{1}{\hat{\sigma}_n \sqrt{n}} T_n(r_n)(z) \leq \alpha\right) \longrightarrow P\left(\sup_{z \in [0,1]} B_z \leq \alpha\right) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\alpha} e^{-\frac{1}{2}u^2} du, \ \alpha \geq 0. \tag{6.24}$$

Dies entspricht der Verteilung einer Zufallsvariablen $|N|$ mit $N \sim \mathcal{N}(0,1)$, denn:

$$\begin{aligned} P(|N| \leq \alpha) &= P(0 \leq N \leq \alpha) + P(-\alpha \leq N \leq 0) \\ &= 2P(0 \leq N \leq \alpha) \\ &= \frac{2}{\sqrt{2\pi}} \int_{0}^{\alpha} e^{-\frac{1}{2}u^2} du \ (\alpha \geq 0). \end{aligned}$$

Wegen $|N|^2 \sim N^2$ lässt sich nun wie in den Abschnitten 6.2.3 und 6.2.3 ein χ^2-Test formulieren. Als Teststatistik wähle man $T := (M_n^{(1)})^2 + \ldots (M_n^{(s)})^2$, wobei die $M_n^{(l)},\ l = 1, \ldots, s$, Beobachtungen der Zufallsvariablen

$$M_n^{(l)} := \sup_{z\in[0,1]} \frac{1}{\hat\sigma_n^{(l)}\sqrt{n}} T_n(r_n^{(l)})(z) = \max_{k=1}^{n} \frac{1}{\hat\sigma_n^{(l)}\sqrt{n}} \sum_{i=1}^{[nt_{nk}]} r_{ni}^{(l)} \tag{6.25}$$

sind, mit $\hat\sigma_n$ wie oben.

Alternativ zum Maximum des Residuen-Partialsummenprozesses kann auch dessen Minimum betrachtet werden. Aufgrund der Symmetrieeigenschaften der Normalverteilung und wegen $\mathrm{E}(B_z) = 0\ \forall z$ ist

$$P\left(\inf_{z\in[0,1]} B(z) \geq -\alpha\right) = P\left(-\inf_{z\in[0,1]} B(z) \leq \alpha\right) = P\left(\sup_{z\in[0,1]} B(z) \leq \alpha\right) = P(|N| \leq \alpha).$$

Also können für die Teststatistik T die Beobachtungen (6.25) ersetzt werden durch Beobachtungen von

$$\inf_{z\in[0,1]} \frac{1}{\hat\sigma_n^{(l)}\sqrt{n}} T_n(r_n^{(l)})(z) = \min_{k=1}^{n} \frac{1}{\hat\sigma_n^{(l)}\sqrt{n}} \sum_{i=1}^{[nt_{nk}]} r_{ni}^{(l)}.$$

6.2.4 Tests auf der Basis von Rang- und Ordnungsstatistiken

Tests aufgrund von Rang- und Ordnungsstatistiken sind nichtparametrische statistische Verfahren, d.h. es werden (außer der Stetigkeit der Verteilungsfunktion) keine Verteilungsannahmen über die Verteilung der zu Grunde liegenden Zufallsvariablen getroffen.

6.2.7 Definition: Es sei $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ mit $x_i \neq x_j\ (i \neq j)$. Dann heißt

$$r_j := \sum_{i=1}^{n} \mathbf{1}\{x_i \leq x_j\} \text{ mit } \mathbf{1}\{x_i \leq x_j\} = \begin{cases} 1, & x_i \leq x_j, \\ 0, & \text{sonst}, \end{cases}$$

der *Rang* von x_j unter $x_1, \ldots, x_n$. Der Vektor

$$r(x) := (r_1, \ldots, r_n)^\top$$

heißt *Rangvektor* von x. Eine Statistik T heißt *Rangstatistik*, falls ihr Wert ausschließlich vom Rangvektor abhängig ist, d.h. falls gilt

$$r(x) = r(y) \Longrightarrow T(x) = T(y)$$

für alle x, y mit $x_i \neq x_j\ (i \neq j)$ und $y_i \neq y_j\ (i \neq j)$.

6.2.8 Bemerkung: Unter der Voraussetzung $X_1, \ldots, X_n \overset{iid}{\sim} F$ ist die Verteilung einer Rangstatistik T unabhängig von der Verteilung F der zu Grunde liegenden Zufallsvariablen, sofern F nur eine stetige Verteilungsfunktion besitzt.

Rangtest von Kendall

Es sei

$$A := \sum_{j=2}^{n} \sum_{\substack{i=1 \\ i<j}}^{n-1} \mathbf{1}\{X_i < X_j\}$$

die Anzahl der Paare (i, j) mit $i < j$ und $X_i < X_j\ (i \in \{1, \ldots, n-1\},\ j \in \{2, \ldots, n\})$. Es gibt $\frac{n(n-1)}{2} = \sum_{j=1}^{n-1} j$ solcher Paare. Demnach besitzt A unter der Hypothese

$$H_0 : (X_t) \sim WN(0, \sigma^2)$$

den Erwartungswert $\frac{n(n-1)}{4}$. Genauer gilt unter H_0 (s. KENDALL / STUART (1983))

$$T_n := \frac{A - \frac{1}{4}n(n-1)}{\sqrt{\frac{1}{72}n(n-1)(2n+5)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Man verwerfe daher H_0 zum Niveau α genau dann, wenn $|T_n| > \Phi_{1-\alpha/2}$. Dabei ist $\Phi_{1-\alpha/2}$ das $(1-\frac{\alpha}{2})-$Quantil der Standard-Normalverteilung.

6.2.9 Bemerkung: Die Größe

$$\tau := \frac{4A}{n(n-1)} - 1$$

heißt *Kendall's τ-Koeffizient*. Unter H_0 besitzt τ den Erwartungswert 0.

Ordnungs-Teststatistiken

Sind $x_1, \ldots, x_n \in \mathbb{R}$, so bezeichnet $x_{(k)}$ das k-kleinste Element aus $\{x_1, \ldots, x_n\}$, d.h. $x_{(1)}, \ldots, x_{(n)}$ ist eine Folge aufsteigender Zahlen aus $\mathbb{R}$.

6.2.10 Definition: Es sei $X = (X_1, \ldots, X_n)^\top$ ein Vektor reellwertiger Zufallsvariablen. Dann heißt $X_{(k)}$ die *k-te Ordnungsstatistik* von X, der Zufallsvektor $(X_{(1)}, \ldots, X_{(n)})^\top$ heißt *Ordnungsstatistik* von X.

Es seien $X_1, \ldots, X_n$ Beobachtungen des Prozesses (X_t) zu den Zeitpunkten $t_1 < \cdots < t_n$. Getestet werden soll die Hypothese

$$H_0 : \; X_1, \ldots, X_n \sim \mathit{IID}(0, \sigma^2). \tag{6.26}$$

Man transformiere nun die Beobachtungen X_t und setze $Y_t := \Phi(X_t)$, wobei die Transformation Φ in Abhängigkeit von der Alternative H_1 gewählt werden muss. Soll die Hypothese z.B. gegen die Alternative

$$H_1 : \; (X_t) \textit{ ist } AR(1)$$

getestet werden, wähle man $Y_t = \Phi(X_t) := X_{t+1}, \; t = 1, \ldots, n-1$. Näheres zur Wahl der Transformation s. KULPERGER / LOCKHART (1998).

Nun werden die Paare (X_t, Y_t) nach der Größe von X geordnet und mit $(X_{(1)}, Y_{(1)}), \ldots, (X_{(n-1)}, Y_{(n-1)})$ bezeichnet. Es gilt also $X_{(1)} < \cdots < X_{(n-1)}$. Für $k = 1, \ldots, n-1$ setzt man nun

$$S_k := \frac{1}{\sqrt{n-1}} \sum_{j=1}^{k} (Y_{(j)} - \overline{Y}) \tag{6.27}$$

mit $\overline{Y} = \frac{1}{n-1}\sum_{j=1}^{n-1} Y_j$. Der folgende Satz über die Grenzverteilung des Prozesses $(S_k)_{k=1}^{n-1}$ bildet die Grundlage für eine Reihe von Ordnungs-Teststatistiken:

6.2.11 Satz: *Der Prozess $(\frac{1}{\sigma}S_k)$ mit S_k wie in (6.27) und $\sigma^2 = Var(Y)$ konvergiert in $D[0,1]$ (s. Abschnitt 3.5.3) nach Verteilung gegen eine Brown'sche Brücke B_0.*

Beweis: KULPERGER / LOCKHART (1998).

Nach Schätzung von σ^2 durch $\widehat{\sigma^2} = \frac{1}{n-2}\sum_{j=1}^{n-1}(Y_j - \overline{Y})^2$ und $\hat{\sigma} := \sqrt{\widehat{\sigma^2}}$ kann man zum Testen der Hypothese (6.26) verschiedene Teststatistiken wählen, zum Beispiel

1) *Kolmogorov-Smirnov*: $K_{n-1} = \frac{1}{\hat{\sigma}} \max_{k=1}^{n-1} |S_k|$,

2) *Cramér-von Mises*: $C_{n-1} = \frac{1}{\widehat{\sigma^2}(n-1)} \sum_{k=1}^{n-1} S_k^2$.

Nach Satz 6.2.11 folgt mit dem Abbildungssatz sofort

$$K_{n-1} \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |B_0(t)| \text{ in } D[0,1], \tag{6.28}$$

$$C_{n-1} \xrightarrow{\mathcal{D}} \int_0^1 B_0^2(t)dt \text{ in } D[0,1]. \tag{6.29}$$

Wegen (6.28) und (6.29) lassen sich die Kolmogorov-Smirnov- bzw. Cramér-von Mises-Testvorschriften aus Abschnitt 6.2.2 unmittelbar auf die Teststatistiken K_{n-1} bzw. C_{n-1} anwenden.

6.2.5 Weitere nichtparametrische Tests

Turning Point Test

Es sei $x_1, \dots x_n$ eine Folge von Beobachtungen und für ein x_i ($i \in \{2, \dots, n-1\}$) gelte

$$\begin{aligned} & x_{i-1} < x_i \text{ und } x_{i+1} < x_i \\ \text{oder} \quad & x_{i-1} > x_i \text{ und } x_{i+1} > x_i \ , \end{aligned}$$

dann heißt x_i *Turning Point* der Folge $x_1, \dots, x_n$. Ist nun $X_1, X_2, \dots$ eine Folge *iid*-verteilter Zufallsvariabler und T_n die Anzahl der Turning Points der Folge $X_1, \dots, X_n$, so ist T_n asymptotisch normalverteilt mit Erwartungswert $\mu_n := \frac{2}{3}(n-2)$ und Varianz $\sigma_n^2 := \frac{1}{90}(16n-29)$. (Zum Beweis s. Kendall / Stuart (1983)). Man verwerfe daher die Hypothese $H_0 : X_1, \dots, X_n \sim iid$ genau dann, wenn

$$\frac{|T_n - \mu_n|}{\sigma_n} > \Phi_{1-\alpha/2}.$$

Wie üblich bezeichnet dabei $\Phi_{1-\alpha/2}$ das $(1-\frac{\alpha}{2})$-Quantil der Standard-Normalverteilung.

6.2.12 Beispiel: Werden die geodätische Zeitreihe aus Abbildung 3.13 und deren Residuen nach Anpassung an ein $ARMA(3,3)$-Modell (Abb. 3.16) dem Turning Point Test unterzogen, erhält man die p-Werte 0.0782 bzw. 0.7691. Die Hypothese eines White Noise Prozesses für die Zeitreihe aus Abbildung 3.13 kann somit zum Niveau 0.1 verworfen werden. Der Annahme eines WN-Prozesses für die Residuen aus Abbildung 3.16 widersprechen die Testergebnisse hingegen nicht. Man vergleiche die entsprechenden p-Werte in Beispiel 6.2.4 (Portmanteau-Test), die dieselben Schlussfolgerungen zulassen. Zu beachten sind jedoch die eindeutiger ausfallenden Ergebnisse in Beispiel 6.2.4, was u.a. auf die höhere Güte des Portmanteau-Tests zurückzuführen ist (Güte von Tests s. Abschnitt 6.5).

6.2.13 Bemerkung: Der Turning Point Test ist nicht geeignet, einen evtl. vorhandenen schwachen Trend zu entdecken.

Vorzeichen-Test

Mit den Bezeichnungen wie oben sei S_n die Anzahl der Zeitpunkte i mit $X_i > X_{i-1}$, $i = 2, \dots, n$. Sind die $X_1, X_2, \dots$ unabhängig und identisch verteilt, so ist S_n asymptotisch normalverteilt mit Erwartungswert $\mu_n := \frac{1}{2}(n-1)$ und Varianz $\sigma_n^2 := \frac{n+1}{12}$ (Brockwell / Davis (1991)). Also verwerfe man die Hypothese $H_0 : X_1, \dots, X_n \sim iid$ genau dann, wenn

$$\frac{|S_n - \mu_n|}{\sigma_n} > \Phi_{1-\alpha/2}.$$

6.2.14 Bemerkung: Dieser Test ist geeignet, wenn ein (v.a. linearer) Trend vermutet wird. Eine zyklische z.B. Komponente kann er nicht entdecken.

6.3 Anpassungstests

Anpassungstests (Goodness-of-Fit-Tests) dienen der Überprüfung einer bestimmten Modellannahme. In diesem Abschnitt werden Anpassungstests vorgestellt, die der Überprüfung einer Verteilungsannahme dienen. Die häufigste Verbreitung finden solche Tests zur Überprüfung der Normalverteilungsannahme.

6.3.1 Definition: Es seien $X_1, \ldots, X_n$ unabhängig und identisch verteilte Zufallsvariablen. Dann wird durch

$$\hat{F}_n(t) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{X_i \leq t\},\ t \in \mathbb{R},$$

die *empirische Verteilungsfunktion* der Zufallsvariablen $X_1, \ldots, X_n$ definiert.

Kolmogorow-Smirnow-Test

Gegeben seien $X_1, \ldots, X_n \overset{iid}{\sim} F$, F eine stetige (unbekannte) Verteilungsfunktion. Weiter sei F^* die vermutete und $\hat{F}_n$ die empirische Verteilungsfunktion der Zufallsvariablen $X_1, \ldots, X_n$. Zu testen sei die Hypothese

$$H_0 : F(t) = F^*(t) \text{ für alle } t \in \mathbb{R}.$$

Die Teststatistik

$$T := \sqrt{n} \cdot \sup_t |F^*(t) - \hat{F}_n(t)|$$

heißt *Kolmogorov-Smirnov-Teststatistik*. Die Testvorschrift lautet

$$\text{Verwirf } H_0 \iff T > K_{1-\alpha}\ .$$

Dabei ist $K_{1-\alpha}$ das $(1-\alpha)$-Quantil der Kolmogorov-Verteilung aus Tabelle 6.2.

6.3.2 Bemerkung: Die Werte aus Tabelle 6.2 sind approximativ und somit für größere Stichprobenumfänge geeignet. In Conover (1971) findet man genauere Quantile für Stichprobenumfänge $n \leq 40$.

6.3.3 Bemerkung: Mit $F^* = \Phi$ (Φ bezeichnet die Verteilungsfunktion der Standard-Normalverteilung) erhält man einen Test zur Überprüfung einer Normalverteilungsannahme. In der Regel müssen zu diesem Zweck die Zufallsvariablen *standardisiert* werden. Man schätzt hierzu den Erwartungswert μ und die Standardabweichung σ der Zufallsvariablen $X_1, \ldots, X_n$ und ersetzt $X_1, \ldots, X_n$ durch $Y_i := \frac{X_i - \hat{\mu}}{\hat{\sigma}},\ i = 1, \ldots, n$.

Cramér-von Mises-Test

Es seien $X_1, \ldots, X_n,\ F$ und F^* wie oben gegeben. Weiter bezeichne $Y_i := \frac{X_i - \hat{\mu}}{\hat{\sigma}},\ i = 1, \ldots, n$, die standardisierte Stichprobe und $\hat{G}_n(u) := \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{F^*(Y_i) \leq u\},\ 0 \leq u \leq 1$, die empirische Verteilungsfunktion der Zufallsvariablen $F^*(Y_i),\ i = 1, \ldots, n$. Dann heißt die Teststatistik

$$T := n \cdot \int_0^1 \left(\hat{G}_n(s) - s\right)^2 ds$$

Cramér-von Mises-Teststatistik. Man verwirft nun die Hypothese

$$H_0 : F(t) = F^*(t) \text{ für alle } t \in \mathbb{R}$$

genau dann, wenn

$$T > C_{1-\alpha}\ ,$$

mit $C_{1-\alpha}$ aus Tabelle 6.3 (vgl. Anderson / Darling (1952)).

6.3.4 Bemerkung: Zur Überprüfung einer Normalverteilungsannahme werden häufig auch graphische Methoden eingesetzt. In den gängigen Statistikpaketen sind z.B. Dichteschätzer sowie der Normal-QQ-Plot standardmäßig implementiert. Eine annähernd normalverteilte Stichprobe sollte dabei einen annähernd linearen Normal-QQ-Plot erzeugen und eine empirische Dichte, die jener der Normalverteilung ähnelt, s. Abbildung 7.3. Vergleicht man die empirische Dichte einer Stichprobe mit der Dichte der Standardnormalverteilung, ist auf eine Standardisierung der Zufallsvariablen (s. Bemerkung 6.3.3) zu achten.

6.4 Tests auf Homogenität der Varianz

Viele statistische Verfahren, z.B. jene aus Abschnitt 3.3, setzen eine konstante Varianzfunktion (*Homoskedastizität*) voraus. Ist diese Voraussetzung verletzt (*Heteroskedastizität*), so sind die Ergebnisse der Verfahren u.U. nicht zuverlässig. Es ist daher sinnvoll, die Annahme von Homoskedastizität mittels eines statistischen Tests zu überprüfen und ggf. durch Gewichtung oder Transformation der Beobachtungen Homoskedastizität zu erzeugen. In diesem Abschnitt werden verschiedene Tests zum Überprüfen der Hypothese homoskedastischer Beobachtungen vorgestellt.

Tests auf Homogenität der Varianz können grundsätzlich in vier Kategorien eingeteilt werden: Für einen *Zwei-Stichproben-Test* unterteilt man den Beobachtungsvektor in zwei Vektoren (Stichproben), für die jeweils homogene Varianzen σ_1^2 und σ_2^2 angenommen werden. Mit diesen Teilvektoren testet man die Hypothese $H_0 : \sigma_1^2 = \sigma_2^2$ oder $H_0' : \sigma_1^2 \leq \sigma_2^2$ gegen die Alternative $H_1 : \sigma_1^2 \neq \sigma_2^2$ bzw. $H_1' : \sigma_1^2 > \sigma_2^2$.

Bei einem *multiplen Test* verfährt man im Prinzip genauso, unterteilt jedoch den Beobachtungsvektor in $k > 2$ Vektoren (Stichproben) und testet $H_0 : \sigma_1^2 = \cdots = \sigma_k^2$ gegen $H_1 : Es\ existieren\ i, j \in \{1, \ldots, k\}\ mit\ \sigma_i^2 \neq \sigma_j^2$. Zwei-Stichproben-Tests und multiple Tests setzen neben homogenen Varianzen der einzelnen Stichproben auch konstante Stichproben-Mittel voraus.

Wird angenommen, dass den Daten ein lineares Modell mit bekannter Designmatrix zugrunde liegt, kann ein *Test auf Homoskedastizität in linearen Modellen* angewendet werden.

Schließlich gibt es noch *nichtparametrische Tests*, die keine Festlegung auf ein bestimmtes lineares Modell erfordern.

6.4.1 Zwei-Stichproben-Tests

Sind $X_{11}, \ldots X_{1n_1} \overset{iid}{\sim} \mathcal{N}(\mu_1, \sigma_1^2)$ und $X_{21}, \ldots X_{2n_2} \overset{iid}{\sim} \mathcal{N}(\mu_2, \sigma_2^2)$ zwei unabhängige Stichproben, so ist ein erwartungstreuer Schätzer für $\sigma_j^2,\ j = 1, 2$, gegeben durch

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ji} - \overline{X_{j\cdot}})^2, \text{ wobei } \overline{X_{j\cdot}} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji}\ ,\ j = 1, 2. \tag{6.30}$$

Für die Zufallsvariablen $\frac{n_j - 1}{\sigma_j^2}\ s_j^2$, $j = 1, 2$, gilt

$$\frac{n_j - 1}{\sigma_j^2}\ s_j^2 \sim \chi_{n_j - 1}^2\ . \tag{6.31}$$

F-Test

Nach (6.31) und Definition B.2.8 ist der Quotient $\frac{s_1^2 \sigma_2^2}{s_2^2 \sigma_1^2}$ unter Normalverteilungsannahme F-verteilt mit $n_1 - 1$ und $n_2 - 1$ Freiheitsgraden. Damit gilt unter $H_0 :\ \sigma_1^2 = \sigma_2^2$

$$F := \frac{s_1^2}{s_2^2} \sim F_{n_1 - 1, n_2 - 1}\ .$$

Entsprechend verwirft man die Hypothesen

- $H_0 : \sigma_1^2 = \sigma_2^2$ genau dann, wenn $F < F_{n_1 - 1, n_2 - 1; \alpha/2}$ oder $F > F_{n_1 - 1, n_2 - 1; 1 - \alpha/2}$,
- $H_0 : \sigma_1^2 \leq \sigma_2^2$ genau dann, wenn $F > F_{n_1 - 1, n_2 - 1; 1 - \alpha}$.

Dabei bezeichnet $F_{n,m;1-\alpha}$ das $(1 - \alpha)$-Quantil der F-Verteilung mit n und m Freiheitsgraden.

6.4.1 Bemerkung: Ist der Erwartungswert μ_j bekannt, $j = 1, 2$, so ist ein erwartungstreuer Schätzer für σ_j^2

$$s_j^2 = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ji} - \mu_j)^2$$

und $\frac{n_j}{\sigma_j^2} s_j^2$ ist χ^2-verteilt mit n_j Freiheitsgraden. Unter diesen Bedingungen besitzt $\frac{s_1^2}{s_2^2}$ eine F-Verteilung mit n_1 und n_2 Freiheitsgraden. Ist μ_j hingegen unbekannt, so bedeutet die Schätzung von μ_j durch $\overline{X_{j\cdot}}$, $j = 1, 2$, jeweils den Verlust eines Freiheitsgrades (s. Abschnitt 5.2.2).

β-Test

Sind s_1^2 und s_2^2 wie in (6.30) gegeben, so besitzt unter Normalverteilungsannahme und unter der Bedingung $\sigma_1^2 = \sigma_2^2$ die Zufallsvariable

$$B := \frac{s_1^2}{s_1^2 + s_2^2} \tag{6.32}$$

nach (6.31) und Definition B.2.9 eine β-Verteilung mit Parametern $\frac{n_1-1}{2}$ und $\frac{n_2-1}{2}$. Man verwirft daher die Hypothesen

- $H_0 : \sigma_1^2 = \sigma_2^2$ genau dann, wenn $B < \beta_{\frac{n_1-1}{2},\frac{n_2-1}{2};\alpha/2}$ oder $B > \beta_{\frac{n_1-1}{2},\frac{n_2-1}{2};1-\alpha/2}$,
- $H_0 : \sigma_1^2 \leq \sigma_2^2$ genau dann, wenn $B > \beta_{\frac{n_1-1}{2},\frac{n_2-1}{2};1-\alpha}$.

Dabei ist $\beta_{p,q;1-\alpha}$ das $(1-\alpha)$-Quantil der Beta-Verteilung mit Parametern p und q.

6.4.2 Bemerkung: Bermerkung 6.4.1 gilt sinngemäß.

6.4.2 Multiple Tests (eindimensionale Teststatistiken)

Maximum F-Ratio Test

Es seien

$$X_{j1}, \ldots, X_{jn} \overset{iid}{\sim} \mathcal{N}(\mu_j, \sigma_j^2),\ j = 1, \ldots k, \tag{6.33}$$

k voneinander unabhängige Stichproben mit gleichem Stichprobenumfang n. Für jede dieser k Stichproben wird die Varianz σ_j^2 erwartungstreu geschätzt durch

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ji} - \overline{X_{j\cdot}})^2,\ \text{wobei } \overline{X_{j\cdot}} = \frac{1}{n}\sum_{i=1}^{n} X_{ji},\ j = 1, \ldots, k.$$

Wie bereits in Abschnitt 6.4.1 angesprochen, besitzt die Zufallsvariable $\frac{n-1}{\sigma_j^2}s_j^2$ eine Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden (χ_{n-1}^2). Man beachte Bemerkung 6.4.1. Die χ_{n-1}^2-Verteilung ist ein Spezialfall einer Gamma-Verteilung. Genauer gilt $\chi_{n-1}^2 = \Gamma(\frac{n-1}{2}, 2)$, s. Johnson / Kotz (1994). Nach Johnson / Kotz (1994), Kap. 17, folgt

$$\begin{aligned} \tfrac{n-1}{\sigma_j^2}s_j^2 \sim \chi_{n-1}^2 \Longrightarrow \quad & \ln\left(\frac{n-1}{\sigma_j^2}s_j^2\right) \text{ ist approximativ normalverteilt} \\ & \text{mit } \operatorname{Var}\left(\ln(\tfrac{n-1}{\sigma_j^2}s_j^2)\right) \approx \tfrac{2}{n-2}\ . \end{aligned} \tag{6.34}$$

Für k Zufallsvariablen $X_1, \ldots, X_k$ bezeichne $X_{max} = \max\{X_1, \ldots, X_k\}$ und $X_{min} = \min\{X_1, \ldots, X_k\}$. Entsprechend ist $s_{max}^2 = \max\{s_1^2, \ldots, s_k^2\}$, etc.

Damit besitzt nach (6.34) die Zufallsvariable

$$\sqrt{\frac{n-2}{2}}\ln\left(\frac{s_{max}^2}{s_{min}^2}\right) = \sqrt{\frac{n-2}{2}}\left(\ln\left(\frac{n-1}{\sigma^2}s_{max}^2\right) - \ln\left(\frac{n-1}{\sigma^2}s_{min}^2\right)\right)$$

unter $H_0 :\ \sigma_1^2 = \cdots = \sigma_k^2 =: \sigma^2$ approximativ dieselbe Verteilung wie der *Range* von k standard-normalverteilten Zufallsvariablen.

6.4.3 Definition: Für k Zufallsvariablen $X_1, \ldots, X_k$ heißt $X_{max} - X_{min}$ der *Range* oder die *Spannweite* der Beobachtungsreihe $X_1, \ldots, X_k$.

Tabelle 6.4: Die (adjustierten) 95%-Quantile der Verteilung von $\frac{s^2_{max}}{s^2_{min}}$

$n\backslash k$	2	3	4	5	6	7	8	9	10	11	12
10	3.72	4.77	5.54	6.18	6.70	7.18	7.55	7.86	8.17	8.50	8.85
12	3.28	4.10	4.71	5.21	5.58	5.93	6.23	6.49	6.69	6.96	7.17
15	2.86	3.49	3.94	4.31	4.57	4.86	5.05	5.21	5.37	5.58	5.76
20	2.46	2.92	3.25	3.49	3.71	3.86	4.02	4.14	4.26	4.35	4.48
30	2.08	2.39	2.59	2.75	2.89	3.00	3.10	3.16	3.22	3.29	3.35
60	1.67	1.84	1.95	2.03	2.10	2.16	2.20	2.25	2.29	2.32	2.34

Die Verteilung $\tilde{F}$ der Spannweite von k *iid*-verteilten $\mathcal{N}(0,1)$-Variablen ist z.B. in Pearson / Hartley (1970) tabelliert (geeignet für große n). Hartley (1950) hat die Tabelle der Verteilung von $\frac{s^2_{max}}{s^2_{min}} = \exp(\tilde{F} \cdot \sqrt{\frac{2}{n-2}})$ so adjustiert, dass sie für kleinere n besser an die Verteilung von $\frac{s^2_{max}}{s^2_{min}}$ angepasst ist. Die entsprechenden 95%-Quantile sind auszugsweise in Tabelle 6.4 aufgelistet.

Ein Standard-Verfahren zum Testen der Hypothese $H_0: \ \sigma_1^2 = \cdots = \sigma_k^2$ ist der

Bartlett-Test

Es seien $X_{j1}, \ldots, X_{jn_j} \overset{iid}{\sim} \mathcal{N}(\mu_j, \sigma_j^2), \ j = 1, \ldots k, \ \ k$ voneinander unabhängige Stichproben mit jeweiligem Stichprobenumfang $n_j \geq 5$. Wie beim Maximum F-Ratio Test wird die Varianz der j-ten Stichprobe geschätzt durch

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_{ji} - \overline{X_{j\cdot}})^2. \tag{6.35}$$

Ein Schätzer für die Gesamtvarianz aller Stichproben ist mit $n = \sum_{j=1}^k n_j$ gegeben durch

$$\begin{aligned} s^2 &= \frac{1}{n-k} \sum_{j=1}^{k} (n_j - 1) s_j^2 \\ &= \frac{1}{n-k} \sum_{j=1}^{k} \sum_{i=1}^{n_j} (X_{ji} - \overline{X_{j\cdot}})^2. \end{aligned} \tag{6.36}$$

Bartlett (1937) hat für die Statistik

$$\begin{aligned} M &:= \ln \left(\prod_{j=1}^{k} \left(\frac{s^2}{s_j^2} \right)^{n_j - 1} \right) \\ &= (n-k) \ln s^2 - \sum_{j=1}^{k} (n_j - 1) \ln s_j^2 \end{aligned}$$

den Vorfaktor

$$C = 1 + \frac{1}{3(k-1)} \left(\sum_{j=1}^{k} \frac{1}{n_j - 1} - \frac{1}{n-k} \right)$$

so bestimmt, dass $\frac{M}{C}$ approximativ χ^2_{k-1}-verteilt ist. Die Testvorschrift lautet somit:

$$\text{Verwirf } \ H_0 \text{ zum Niveau } \alpha \iff \frac{M}{C} > \chi^2_{k-1;1-\alpha} \ .$$

6.4.4 Bemerkung: Sind die Erwartungswerte $\mu_1, \ldots, \mu_k$ bekannt (man beachte Bemerkung 6.4.1), so ist

$$\begin{aligned} s_j^2 &= \frac{1}{n_j}\sum_{i=1}^{n_j}(X_{ji}-\mu_j)^2, \\ s^2 &= \frac{1}{n}\sum_{j=1}^{k} n_j s_j^2 \ , \\ M &= n\ln s^2 - \sum_{j=1}^{k} n_j \ln s_j^2 \ , \\ C &= 1+\frac{1}{3(k-1)}\left(\sum_{j=1}^{k}\frac{1}{n_j}-\frac{1}{n}\right), \end{aligned}$$

und $\frac{M}{C}$ besitzt approximativ eine χ^2_{k-1}-Verteilung.

6.4.5 Bemerkung: In Bischoff et al. (2006) wurde der Bartlett-Test so modifiziert, dass er auf differenzierte Daten (Pseudo-Residuen, vgl. Abschnitt 5.2.4) angewendet werden kann. Der modifizierte Test wird dort als Test auf Homoskedastizität für geodätische GPS-Residuenzeitreihen verwendet, deren trendartiges Verhalten durch Differenzenbildung eliminiert wurde.

6.4.3 Multiple Tests (mehrdimensionale Teststatistiken)

In der Praxis interessiert häufig nicht nur die Frage, ob Heteroskedastizität vorliegt, sondern auch, welche Form diese gegebenenfalls annimmt. In diesem Falle können mehrdimensionale Testatistiken Anhalstspunkte liefern, ob eine Varianzfunktion z.B. eher fällt oder eher steigt. Es sollen nun zwei multiple Tests mit mehrdimensionalen Teststatistiken vorgestellt werden, die auf der *Dirichlet-Verteilung* beruhen. Bei der folgenden Definition beachte man, dass für $\nu \in \mathbb{N}$ und $Z_1, \ldots, Z_\nu \overset{iid}{\sim} \mathcal{N}(0,1)$ die Summe $\sum_{i=1}^{\nu} Z_i^2$ eine χ^2_ν-Verteilung besitzt (vgl. Definition B.2.7).

6.4.6 Definition: Es seien $X_j,\ j = 1, \ldots, k$, unabhängige Zufallsvariablen mit $X_j \sim \chi^2_{\nu_j},\ \nu_j > 0$ $(j = 1, \ldots, k)$ und

$$Y_j = \frac{X_j}{\sum_{i=1}^{k} X_i}. \tag{6.37}$$

Dann heißt die gemeinsame Verteilung der Zufallsvariablen $Y_1, \ldots, Y_k$ *Dirichlet-Verteilung* mit Parametern $\frac{\nu_1}{2}, \ldots, \frac{\nu_k}{2}$, i.Z.

$$(Y_1, \ldots, Y_k) \sim \mathcal{D}\left(\frac{\nu_1}{2}, \ldots, \frac{\nu_k}{2}\right). \tag{6.38}$$

Wegen $Y_k = 1 - \sum_{i=1}^{k-1} Y_i$ schreibt man auch

$$(Y_1, \ldots, Y_{k-1}) \sim \mathcal{D}\left(\frac{\nu_1}{2}, \ldots, \frac{\nu_{k-1}}{2}; \frac{\nu_k}{2}\right).$$

6.4.7 Bemerkung: Für $j = 1, \ldots, k$ ist Y_j $\beta(p,q)$-verteilt mit den Parametern $p = \frac{\nu_j}{2}$ und $q = \sum_{i=1}^{k} \frac{\nu_i}{2} - \frac{\nu_j}{2}$. Die Zufallsvariablen $Y_j,\ j = 1, \ldots, k$, sind korreliert mit (s. Johnson et al. (2000))

$$\mathrm{Cov}(Y_{j_1}, Y_{j_2}) = \frac{-\nu_{j_1}\nu_{j_2}}{\left(\sum_{i=1}^{k}\nu_i\right)^2\left(\frac{1}{2}\sum_{i=1}^{k}\nu_i + 1\right)}.$$

Exakte Konfidenzbereiche für Dirichletverteilte Zufallsvektoren sind schwer zu berechnen. Wludyka und Nelson (1997) geben approximative Konfidenzschranken für $k = 3, \ldots, 12$, $\nu_1 = \cdots = \nu_k =: \nu \in \{3, \ldots, 34\}$ und $\alpha \in \{0.01, 0.05, 0.1\}$ an. Damit kann im Modell (6.33) mit $\nu = n - 1$ die Hypothese $H_0 : \sigma_1^2 = \cdots = \sigma_k^2$ getestet werden.

In der Praxis sind die Parameter $\nu_1, \ldots, \nu_k$ häufig nicht frei wählbar und voneinander verschieden. Wir wenden daher einen Trick an, durch den wir anstelle eines Dirichletverteilten Zufallsvektors einen Vektor unabhängiger β-verteilter Zufallsvariablen als Teststatistik erhalten. Die Konfidenzschranken dieses Tests sind dann für beliebige Freiheitsgrade $\nu_j,\ j = 1, \ldots, k,\ k \in \mathbb{N}$, und beliebige α leicht zu berechnen. Dazu seien Y_j wie in (6.37), $j = 1, \ldots, k$, und

$$V_j := \frac{Y_j}{1 - \sum_{i=1}^{j-1} Y_i} = \frac{X_j}{\sum_{i=j}^{k} X_i},\ j = 1, \ldots, k. \tag{6.39}$$

Die Zufallsvariablen

$$\begin{aligned} 1 - V_{k-1} &= \frac{X_k}{X_{k-1} + X_k} \\ 1 - V_{k-2} &= \frac{X_{k-1} + X_k}{X_{k-2} + X_{k-1} + X_k} \\ &\vdots \\ 1 - V_1 &= \frac{X_2 + \cdots + X_{k-1} + X_k}{X_1 + X_2 + \cdots + X_{k-2} + X_{k-1} + X_k} \end{aligned} \tag{6.40}$$

sind stochastisch unabhängig (s. Johnson / Kotz / Balakrishnan (1995) oder Johnson et al. (2000)). Damit sind auch die $V_1, \ldots, V_{k-1}$ stochastisch unabhängig. Weiter gilt

$$V_j \sim \beta(\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}),\ j = 1, \ldots, k-1.$$

Ferner ist

$$F_j := \sum_{i=1}^{j} Y_i = 1 - \prod_{i=1}^{j} (1 - V_i), \tag{6.41}$$

denn:

- $j = 1$: $1 - (1 - V_1) = V_1 = Y_1$.
- $j \Rightarrow j + 1$: Es sei $\prod_{i=1}^{j}(1 - V_i) = 1 - \sum_{i=1}^{j} Y_i$. Dann ist

$$\begin{aligned} \prod_{i=1}^{j+1}(1 - V_i) &= (1 - V_{j+1})(1 - \sum_{i=1}^{j} Y_i) \\ &= 1 - \frac{Y_{j+1}}{1 - \sum_{i=1}^{j} Y_i} - \sum_{i=1}^{j} Y_i + \frac{Y_{j+1}}{1 - \sum_{i=1}^{j} Y_i} \sum_{i=1}^{j} Y_i \\ &= \frac{(1 - \sum_{i=1}^{j} Y_i)^2 - Y_{j+1}(1 - \sum_{i=1}^{j} Y_i)}{1 - \sum_{i=1}^{j} Y_i} \\ &= 1 - \sum_{i=1}^{j+1} Y_i. \end{aligned}$$

Es sei nun F_j wie in (6.41) und

$$\begin{aligned} H_j &:= -\ln(1 - F_j) = -\sum_{i=1}^{j} \ln(1 - V_i),\ F_j = 1 - \exp(-H_j), \\ h_j &:= H_j - H_{j-1} = -\ln(1 - V_j),\ j = 1, \ldots, k-1. \end{aligned}$$

Dann gilt

$$1 - \exp(-h_j) = V_j \sim \beta\left(\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}\right),\ j = 1, \ldots, k-1.$$

Wie aus der Definition von H_j bzw. h_j hervorgeht, ist (F_j) eindeutig bestimmt durch (H_j), (h_j) bzw. (V_j). Damit enthält (V_j) im Prinzip dieselben Informationen wie (F_j). Theoretisch sind also die folgenden Hypothesen äquivalent:

$$H_0 \quad : \quad (Y_1, \ldots, Y_{k-1}) \sim \mathcal{D}\left(\frac{\nu_1}{2}, \ldots, \frac{\nu_{k-1}}{2}; \frac{\nu_k}{2}\right), \tag{6.42}$$

$$H_0' \quad : \quad V_1, \ldots, V_{k-1} \text{ unabhängig}, \; V_j \sim \beta\left(\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}\right), \; j = 1, \ldots, k-1. \tag{6.43}$$

Ein simultaner Konfidenzbereich für $V_1, \ldots, V_{k-1}$ zum Niveau $(1-\alpha)$ ist somit

$$\mathcal{K} = [A_{1;\alpha'(k)}, B_{1;\alpha'(k)}] \times \cdots \times [A_{k-1;\alpha'(k)}, B_{k-1;\alpha'(k)}], \text{ wobei}$$

$$A_{j;\alpha'(k)} = \beta_{\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}; \frac{\alpha'(k)}{2}} \text{ und } B_{j;\alpha'(k)} = \beta_{\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}; 1-\frac{\alpha'(k)}{2}}, \; j = 1, \ldots, k-1,$$

mit $1-\alpha = (1-\alpha'(k))^{k-1}$. Wie in den vorhergehenden Abschnitten bezeichnet $\beta_{p,q;\alpha}$ das α-Quantil der Beta-Verteilung mit Parametern p und q.

Wird der Konfidenzbereich $\mathcal{K}$ an mindestens einer Stelle über- oder unterschritten, so kann man zum Niveau $\alpha = 1-(1-\alpha'(k))^{k-1}$ auf eine (an der betreffenden Stelle) signifikant steigende bzw. signifikant fallende Varianz schließen.

6.4.8 Bemerkung: Dieser Test eignet sich vor allem für sehr kleine Stichprobenumfänge, kleine k oder fallende Alternativen (vgl. Tabelle 6.5.2). Ist man ausschließlich an fallenden Alternativen interessiert, so sollte eine einseitige Variante des Tests verwendet werden mit $A_{j;\alpha'(k)} = 0$ und $B_{j;\alpha'(k)} = \beta_{\frac{\nu_j}{2}, \sum_{i=j+1}^{k} \frac{\nu_i}{2}; 1-\alpha'(k)}$, $j = 1, \ldots, k-1$.

6.4.4 CUSUMSQ- und MOSUMSQ-Tests

Weitere mehrdimensionale Teststatistiken, die Anhaltspunkte über die Form der Varianzfunktion liefern, sind die CUSUMSQ- (Cumulative Sum of Squares) und die MOSUMSQ- (Moving Sum of Squares) Teststatistiken. Die bekannteste der beiden ist wohl die CUSUMSQ-Teststatistik von Brown, Durbin und Evans (1975).

Die CUSUMSQ-Teststatistik

Unterteilt man einen auf Homoskedastizität zu testenden Vektor rekursiver Residuen $\mathbf{r_{n-m}} := (r_{m+1}, \ldots, r_n)^\top$ wiederholt in zwei Abschnitte und bildet jeweils die Beta-Teststatistik (6.32), wobei die Trennlinie zwischen den Abschnitten sukzessive verschoben wird, so erhält man die CUSUMSQ-Teststatistik zum Testen der Hypothese

$$H_0 : Var(r_j) = \sigma^2 \; \forall j. \tag{6.44}$$

6.4.9 Definition: Im linearen Modell (5.3) mit $\beta \in \mathbb{R}^m$, $Y_n \in \mathbb{R}^n$ und einem Vektor $\mathbf{r_{n-m}} = (r_{m+1}, \ldots, r_n)^\top$ rekursiver Residuen heißt

$$C_j := \frac{\sum_{i=m+1}^{j} r_i^2}{\sum_{i=m+1}^{n} r_i^2}, \; j = m+1, \ldots, n,$$

die *CUSUMSQ-Teststatistik*.

Nach Definition B.2.9 gilt im Falle normalverteilter Fehler und unter Vorliegen der Hypothese (6.44) $C_j \sim \beta(\frac{j-m}{2}, \frac{n-j}{2})$, $j = m+1, \ldots, n-1$. Der Vektor der Zuwächse $C_{j+1} - C_j$, $j = m+1, \ldots, n-1$, der CUSUMSQ-Teststatistik besitzt unter denselben Voraussetzungen nach Definition 6.4.6 eine Dirichlet-Verteilung mit den Parametern $\frac{\nu_i}{2} = \frac{1}{2}$, $i = 1, \ldots, n-m$.

In der Literatur gibt es verschiedene Ansätze, Konfidenzschranken für die CUSUMSQ-Teststatistik zu approximieren. Brown, Durbin und Evans passen in ihrem Artikel von 1975 lineare Konfidenzschranken $s_u(n-m) := \left\{\frac{i}{n-m} - c(n-m), \; i = 1, \ldots, n-m\right\}$ und $s_o(n-m) := \left\{\frac{i}{n-m} + c(n-m), \; i = 1, \ldots, n-m\right\}$

an die auf das Zeitintervall $[0,1]$ transformierte CUSUMSQ-Teststatistik an. Dabei ist $c(n-m)$ eine von $n-m$ und α abhängige Konstante. Die Hypothese einer konstanten Varianzfunktion wird genau dann verworfen, wenn die CUSUMSQ-Teststatistik an mindestens einer Stelle den Konfidenzbereich verlässt. Es ist bekannt, dass diese linearen Konfidenzschranken in den Randbereichen seltener und im mittleren Bereich häufiger überschritten werden, als es dem angegebenen Niveau entspricht. Tanizaki (1995) gibt einen Simulations-Algorithmus an, mit dem zu gegebenem $n-m$ und α genauere Konfidenzschranken für die CUSUMSQ-Teststatistik berechnet werden können.

In Tabelle 6.5 sind die Werte $c(n-m)$ nach Brown et al. (1975) (siehe auch Durbin (1969)) auszugsweise angegeben. In Klammern wurden simulierte Niveaus $\alpha'(n-m)$ hinzugefügt. Verwendet man diese, wird die Hypothese genau dann zum Niveau α verworfen, wenn die auf das Zeitintervall $[0,1]$ transformierte CUSUMSQ-Teststatistik an mindestens einer Stelle $\frac{i}{n-m}$ das $\frac{\alpha'(n-m)}{2}$- oder das $(1-\frac{\alpha'(n-m)}{2})$-Quantil der $\beta(\frac{i}{2}, \frac{n-m-i}{2})$-Verteilung unter- bzw. überschreitet ($i \in \{1,\dots,n-m-1\}$).

Tabelle 6.5: Die Werte $c(n-m)$ und $\alpha'(n-m)$ (in Klammern) für die CUSUMSQ-Teststatistik für verschiedene α und Differenzen $n-m$

$\alpha \setminus n-m$	40	60	80	100	120	200
0.1	0.23781	0.19985	0.17595	0.15911	0.14641	0.11549
	(0.0084)	(0.0060)	(0.0056)	(0.0051)	(0.0046)	(0.0040)
0.05	0.26685	0.22383	0.19684	0.17785	0.16355	0.12881
	(0.0037)	(0.0026)	(0.0024)	(0.0022)	(0.0021)	(0.0018)
0.01	0.32459	0.27168	0.23857	0.21534	0.19786	0.15555
	(0.0007)	(0.0005)	(0.0004)	(0.0004)	(0.0004)	(0.0003)

Die Abbildung 6.2 zeigt eine CUSUMSQ-Teststatistik mit $n-m=100$ und den simulierten Konfidenzbereich nach Tabelle 6.5. Die linearen Konfidenzschranken von Brown, Durbin und Evans sind gestrichelt eingezeichnet.

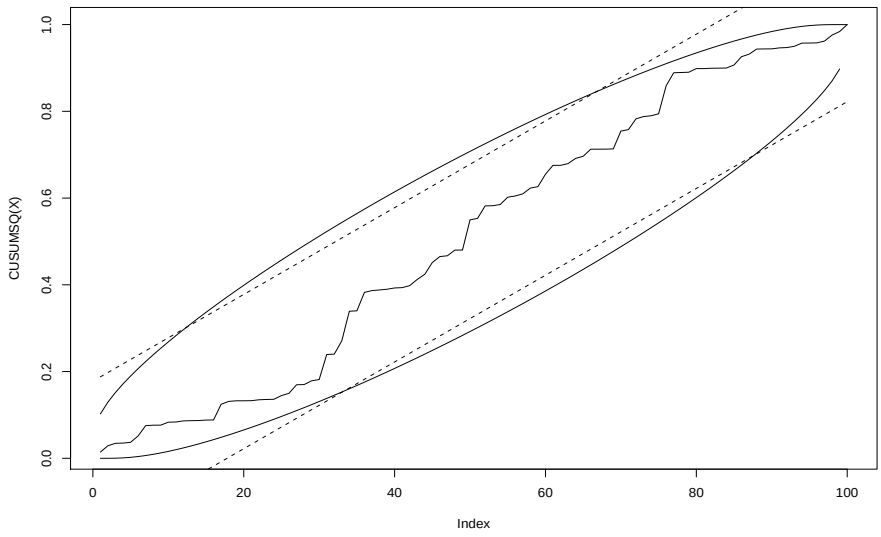

Abbildung 6.2: CUSUMSQ-Teststatistik mit Konfidenzbereichen

Die MOSUMSQ-Teststatistik

6.4.10 Definition: Im linearen Modell (5.3) mit $\beta \in \mathbb{R}^m$, $Y_n \in \mathbb{R}^n$ und einem Vektor $\mathbf{r_{n-m}} := (r_{m+1},\dots,r_n)^\top$ rekursiver Residuen heißt für $G \in \{1,\dots,n-m-1\}$ die Statistik

$$M_j(G) := \frac{\sum_{i=j-G+1}^{j} r_i^2}{\sum_{i=m+1}^{j-G} r_i^2 + \sum_{i=j+1}^{n} r_i^2} \cdot \frac{n-m-G}{G},\ j = m+G,\dots,n,$$

MOSUMSQ(G)-Teststatistik.

Die MOSUMSQ-Teststatistik besitzt bei normalverteilten Fehlern und unter $H_0 : Var(r_j) = \sigma^2 \; \forall j$, an jeder Stelle $j \in \{m+G, \ldots, n\}$ eine F-Verteilung mit G und $n-m-G$ Freiheitsgraden. HACKL (1980) wählt konstante Konfidenzschranken für die MOSUMSQ(G)-Teststatistik, die in Tabelle 6.6 eingetragen sind. Aufgrund von Korrelationen zwischen den F-verteilten Zufallsvariablen wird das Niveau des Tests mit diesen auf theoretischen Überlegungen beruhenden kritischen Werten nicht ausgeschöpft. D.h., das nominale Nivaeu α ist größer als die tatsächliche Wahrscheinlichkeit für einen Fehler erster Art. Als Alternative gibt HACKL (1980) simulierte Konfidenzschranken an. Diese sind in der Tabelle in Klammern angegeben.

Tabelle 6.6: Einige Konfidenzschranken für die MOSUMSQ(G)-Teststatistik

$n-m$	40	40	40	80	80	80	120	120	120
$\alpha \setminus G$	5	10	20	5	10	20	5	10	20
0.1	4.53	3.58	3.35	4.46	3.32	2.69	4.51	3.29	2.59
	(3.98)	(3.01)	(2.70)	(3.94)	(2.86)	(2.28)	(3.95)	(2.87)	(2.23)
0.01	6.38	4.92	4.80	5.86	4.19	3.33	5.82	4.05	3.10
	(5.68)	(4.39)	(4.05)	(5.33)	(3.74)	(2.95)	(5.12)	(3.17)	(2.77)

F-Test mit variabler Trennlinie

Ein weiterer Test auf Homoskedastizität für einen normalverteilten Residuenvektor ergibt sich, wenn wir im CUSUMSQ-Test an jeder Stelle $j = m+1, \ldots, n-1$ den β-Test durch einen F-Test mit der Teststatistik

$$F_j := \frac{\sum_{i=m+1}^{j} r_i^2}{\sum_{i=j+1}^{n} r_i^2} \cdot \frac{n-j}{j-m}$$

ersetzen. Berechnet man nach Realisierung des Residuenvektors an jeder Stelle $j = m+1, \ldots, n-1$ den p-Wert p_j der Teststatistik F_j, so lässt $P_{min} := \min_{j=m+1}^{n-1}(p_j)$ auf das Niveau α schließen, zu dem die Hypothese $H_0 : Var(r_j) = \sigma^2 \; \forall j$ verworfen werden kann. Dieser Test lässt sich auch für einseitige Alternativen formulieren. In diesem Fall nimmt die Hypothese die Form $H_0' : Var(r_i) \leq Var(r_j) \; für \; i < j$ an.

In Simulationen wurden für den Fall einseitiger Alternativen untere Schranken für P_{min} ermittelt. Man kann damit die Hypothese H_0' zum Niveau α verwerfen, falls P_{min} den entsprechenden Wert aus Tabelle 6.7 annimmt oder unterschreitet.

Tabelle 6.7: F-Test mit variabler Trennlinie (einseit. Alternative): untere Schranken für P_{min}

$\alpha \setminus n-m$	40	80	100	120
0.1	0.0091	0.0063	0.0053	0.0044
0.05	0.0037	0.0026	0.0022	0.0019
0.01	0.0006	0.0004	0.0004	0.0004

Für diesen Test kann man auch eine graphische Darstellung wählen. Man verwirft die Hypothese H_0' genau dann zum Niveau α, wenn die Teststatistik F_j an mindestens einer Stelle $j \in \{m+1, \ldots, n-1\}$ die Konfidenzschranke $F_{j-m,n-m-j;1-\alpha'(n)}$ überschreitet. Dabei ist $F_{j-m,n-m-j;1-\alpha'(n)}$, das $(1-\alpha'(n))$-Quantil der F-Verteilung mit $j-m$ und $n-m-j$ Freiheitsgraden und $\alpha'(n)$ bezeichnet den zum Stichprobenumfang $n-m$ und zum Niveau α korrespondierenden Eintrag in Tabelle 6.7. In Abbildung 6.3 ist eine solche graphische Darstellung des Tests zu sehen mit $n-m = 100, \; \alpha = 0.05$ und $\alpha'(n) = 0.0022$.

HSU (1977) wählt an Stelle von P_{min} das arithmetische Mittel

$$M_{n-m} := \frac{1}{n-m-1} \sum_{j=m+1}^{n-1} (1-p_j)$$

als Teststatistik und gibt (auf Simulationsstudien beruhende) kritische Werte für die Hypothese $H_0 : Var(r_j) = \sigma^2 \; \forall j$ an. Die Hypothese $H_0' : Var(r_i) \leq Var(r_j) \; für \; i < j$ kann (ebenfalls auf Basis von Simulationsstudien) verworfen werden, wenn M_{n-m} den entsprechenden Wert aus Tabelle 6.8 überschreitet.

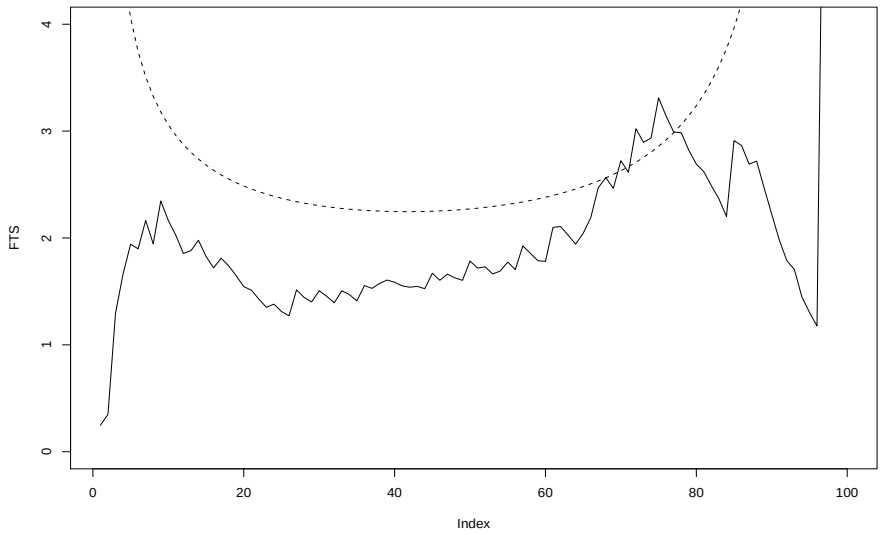

Abbildung 6.3: F-Teststatistik (variable Trennlinie) mit Konfidenzschranke

6.4.11 Beispiel: In Bischoff et al. (2005) wird mit dem F-Test von Hsu eine statistisch signifikante Abhängigkeit der Varianz doppeltdifferenzierter GPS-Trägerphasenbeobachtungen von den Elevationen der beteiligten Satelliten nachgewiesen.

Tabelle 6.8: F-Test mit variabler Trennlinie (einseit. Alternative): obere Schranken für M_{n-m}

$\alpha \setminus n-m$	40	80	100	120
0.1	0.765	0.765	0.765	0.760
0.05	0.820	0.820	0.820	0.815
0.01	0.900	0.900	0.900	0.900

6.4.5 Ein Test auf Homoskedastizität in linearen Modellen

Gegeben sei das lineare Modell

$$Y_n \quad = \quad X_n\beta + \epsilon_n$$

mit bekannter Designmatrix X_n und einem Vektor ϵ_n normalverteilter, unkorrelierter Fehler. Der nun folgende Goldfeld-Quandt-Test (Goldfeld / Quandt (1965)) eignet sich zum Testen der Hypothese einer konstanten Varianz gegen monotone Alternativen.

Goldfeld-Quandt-Test

Zum Testen der Hypothese

$$H_0 : \; Var(\epsilon_n) \equiv \; const$$

gegen die Alternative

$$H_1 : \; Var(\epsilon_n) \; ist \; monoton \; und \; \not\equiv \; const$$

mit dem Goldfeld-Quandt-Test unterteilt man den Beobachtungsvektor Y_n in drei Bereiche der Längen $\frac{n-k}{2}$, k und $\frac{n-k}{2}$. (Es wird empfohlen, die Länge k des mittleren Bereiches so zu wählen, dass alle drei so erhaltenen Vektoren in etwa die gleiche Dimension besitzen.) Anschließend führt man anhand der Vektoren $Y^{(1)} := (Y_{n1}, \ldots, Y_{n,\frac{n-k}{2}})^\top$ und $Y^{(2)} := (Y_{n,n-\frac{n-k}{2}+1}, \ldots, Y_{nn})^\top$ getrennt eine lineare Regression durch und bestimmt jeweils die Summe der quadrierten Residuen (*Residual Sum of Squares, RSS*). Werden zur Berechnung der Residual Sum of Squares unkorrelierte, normalverteilte Residuen verwendet, so gilt unter der Hypothese nach Definition B.2.7 und Lemma 5.2.9 für die *RSS* der Vektoren $Y^{(1)}$ und $Y^{(2)}$:

$$\frac{1}{\sigma^2} \cdot RSS_j \sim \chi^2_{\frac{n-k}{2}-m} \ , \ j = 1, 2, \ m = \dim(\beta).$$

Bei Verwendung der LS-Residuen gilt diese Verteilung approximativ. Man beachte in diesem Zusammenhang Bemerkung 5.2.10. Entsprechend ist die Teststatistik

$$GQ := \frac{RSS_1}{RSS_2}$$

unter H_0 F-verteilt (bzw. approximativ F-verteilt) mit $\frac{n-k}{2} - m$ und $\frac{n-k}{2} - m$ Freiheitsgraden (Definition B.2.8). Man verwirft H_0 genau dann zum Niveau α, wenn

$$GQ < F_{\frac{n-k}{2}-m, \frac{n-k}{2}-m; \alpha/2} \text{ oder } GQ > F_{\frac{n-k}{2}-m, \frac{n-k}{2}-m; 1-\alpha/2} \ .$$

Testet man H_0 gegen die Alternative $H_1 :$ $\mathrm{Var}(\epsilon_n)$ *ist monoton fallend und* $\not\equiv$ *const*, so wird H_0 genau dann verworfen, wenn

$$GQ > F_{\frac{n-k}{2}-m, \frac{n-k}{2}-m; 1-\alpha} \ .$$

6.4.6 Ein nichtparametrischer Test

Nichtparametrische Tests bieten den Vorteil, dass sie auf eine Modellspezifikation (z.B. auf die Kenntnis der Regressionsfunktionen eines linearen Modells oder der Verteilung der Fehlervektoren) verzichten und daher Fehler aufgrund falscher Modellannahmen nicht oder in geringerem Umfang auftreten können. In der Regel zahlt man dafür jedoch den Preis eines asymptotischen Verfahrens, das erst bei relativ hohen Stichprobenumfängen gute Ergebnisse liefert.

Dette-Munk-Test

Der Dette-Munk-Test wurde für unkorrelierte Zeitreihen mit deterministischem Trend entwickelt und kann daher direkt auf einen unkorrelierten Beobachtungsvektor Y_n angewendet werden. Er basiert auf der Überlegung, dass im Falle einer konstanten Varianzfunktion $\sigma^2(\cdot)$ ein $c \in \mathbb{R}$ existiert, so dass $\sigma^2(t) - c = 0$ für alle t aus dem Versuchsbereich $[a, b]$ gilt. Die Teststatistik des Dette-Munk-Tests schätzt approximativ den $L^2(a, b)$-Abstand zwischen der Varianzfunktion $\sigma^2(\cdot)$ und einer geeigneten konstanten Funktion $f(t) \equiv const$, indem sie den Ausdruck

$$T^2 := \frac{1}{n} \sum_{i=1}^{n} \left(\sigma^2(t_{ni}) - c\right)^2$$

über c minimiert. T^2 kann als Heteroskedastizitätsmaß betrachtet werden und verschwindet genau dann, wenn die Hypothese $H_0 :$ $\sigma^2 \equiv const$ vorliegt. Man beachte, dass unter H_0 gilt

$$c = \frac{1}{n} \sum_{j=1}^{n} \sigma^2(t_{nj}).$$

Damit kann eine geeignete Teststatistik entwickelt werden, die die Größe

$$\begin{aligned} T_n^2 &:= \frac{1}{n} \sum_{i=1}^{n} \left(\sigma^2(t_{ni}) - \frac{1}{n} \sum_{i=1}^{n} \sigma^2(t_{ni}) \right)^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \sigma^4(t_{ni}) - \left(\frac{1}{n} \sum_{i=1}^{n} \sigma^2(t_{ni}) \right)^2 \end{aligned} \tag{6.45}$$

schätzt. Gegeben sei nun das lineare Modell

$$Y_{ni} = g(t_{ni}) + \sigma^2(t_{ni})\epsilon_{ni},\ i = 1, \ldots, n, \tag{6.46}$$

mit Erwartungswertfunktion g und Varianzfunktion σ^2. Dabei sei für jedes $n \in \mathbb{N}$ der Fehlervektor $(\epsilon_{n1}, \ldots, \epsilon_{nn})^\top$ ein Vektor unabhängiger Zufallsvariablen mit $\mathrm{E}(\epsilon_{ni}) = 0$ und $\mathrm{Var}(\epsilon_{ni}) = 1,\ i = 1, \ldots, n$, und die Funktionen g und σ^2 seien auf $[a, b]$ *Lipschitzstetig* der Ordnung $\gamma > \frac{1}{2}$.

6.4.12 Definition: Eine auf dem Intervall $[a, b]$ definierte Funktion f heißt auf $[a, b]$ *Lipschitzstetig der Ordnung* γ, falls eine Konstante C existiert, so dass für alle $s, t \in [a, b]$ gilt

$$|f(s) - f(t)| \leq C \cdot |s - t|^\gamma.$$

Schließlich wird für das Design $a \leq t_{n1} \leq \ldots t_{nn} \leq b$ eine bestimmte "Glattheitseigenschaft" vorausgesetzt (s. Dette / Munk (1998)), die sicherstellt, dass die Verteilung der Punkte $t_{n1}, \ldots, t_{nn}$ auf dem Versuchsbereich $[a, b]$ nicht allzu unregelmäßig ist (insbesondere sollte sie keine Lücken aufweisen). Diese Glattheitseigenschaft ist z.B. im Falle eines äquidistanten Designs erfüllt.

Sind alle diese Voraussetzungen erfüllt und definiert man Pseudo-Residuen gemäß $R_j = Y_j - Y_{j-1}\ ,\ j = 2, \ldots, n$, so lässt sich T_n^2 aus (6.45) konsistent schätzen (s. Dette / Munk (1998)) durch

$$M_n^2 := \frac{1}{4(n-3)} \sum_{j=2}^{n-2} R_j^2 R_{j+2}^2 - \left(\frac{1}{2(n-1)} \sum_{j=2}^{n} R_j^2 \right)^2 . \tag{6.47}$$

Da R_i und R_j korreliert sind für $|i - j| \in \{0, 1\}$, enthält der zweite Term in (6.47) Produkte korrelierter Zufallsvariablen. Um dies zu vermeiden, kann die Teststatistik wie folgt modifiziert werden (vgl. Schoknecht (2001)):

$$\tilde{M}_n^2 := \frac{1}{4(n-3)} \sum_{j=2}^{n-2} R_j^2 R_{j+2}^2 - \frac{1}{4(n-3)(n-4)} \sum_{j=3}^{n-1} \left(R_j^2 \cdot \sum_{\substack{2 \leq i \leq n \\ i \notin \{j-1, j, j+1\}}} R_i^2 \right).$$

In Dette / Munk (1998) wird die asymptotische Normalverteilung der Statistik M_n^2 bewiesen. Der entsprechende Beweis für die Statistik $\tilde{M}_n^2$ funktioniert analog. Die Varianzen der Statistiken M_n^2 bzw. $\tilde{M}_n^2$ werden in den entsprechenden Literaturquellen unter allgemeinen Verteilungsannahmen für die Fehler ϵ_{ni} angegeben, insbesondere wird keine Normalverteilung der Fehler vorausgesetzt.

Die Testvorschrift für den Dette-Munk-Test lautet dann:

$$\text{Verwirf } H_0 \iff \frac{M_n^2}{\sqrt{\mathrm{Var}(M_n^2)}} < \Phi_{\frac{\alpha}{2}} \text{ oder } \frac{M_n^2}{\sqrt{\mathrm{Var}(M_n^2)}} > \Phi_{1-\frac{\alpha}{2}}\ ,$$

wobei Φ_α das α-Quantil der Standard-Normalverteilung bezeichnet, bzw.

$$\text{Verwirf } H_0 \iff \frac{\tilde{M}_n^2}{\sqrt{\mathrm{Var}(\tilde{M}_n^2)}} < \Phi_{\frac{\alpha}{2}} \text{ oder } \frac{\tilde{M}_n^2}{\sqrt{\mathrm{Var}(\tilde{M}_n^2)}} > \Phi_{1-\frac{\alpha}{2}}\ .$$

Spezialfall: Dette-Munk-Test für eine nach einer Funktion der Satellitenelevationen geordnete GPS-Residuen-Zeitreihe

Im konkreten Fall geodätischer GPS-Messdaten sollte nachgewiesen werden, dass sich deren Varianz in Abhängigkeit von der Satellitenelevationen monoton verändert. Zu diesem Zweck wurden doppeltdifferenzierte Messdaten aus der Region der antarktischen Halbinsel mit der Berner GPS-Software ausgewertet und die daraus erhaltenen Residuen entsprechend einer Funktion der Satellitenelevationen angeordnet. Abbildung 6.4 zeigt eine solche Residuen-Zeitreihe eines Satelliten-Empfänger-Paares in Abhängigkeit von der Elevation.

GPS-Residuen-Zeitreihen können einen z.T. deutlichen Trend aufweisen. Durch die Umordnung bezüglich der Satellitenelevationen erhält man dann eine Zeitreihe mit zwei "Ästen", wie Abbildung 6.4 beispielhaft zeigt.

Zum Testen solcher Zeitreihen ist daher eine weitere Modifikation der Statistik $\tilde{M}_n^2$ notwendig, die die Trends der beiden Äste (mittels Bildung der Pseudo-Residuen) getrennt eliminiert.

Hierzu bezeichne $a \leq s_{n1} \leq \cdots \leq s_{nn} \leq b$ das Design des längeren, $a \leq t_{n1} \leq \cdots \leq t_{nk} \leq c$ das Design des kürzeren Astes mit $k = k(n) \in \{0, \ldots, n\}$ und $c \in [a, b]$. Weiter seien

$$X_{ni} := Y(s_{ni}) = f(s_{ni}) + \sigma^2(s_{ni})\epsilon(s_{ni}),\ i = 1, \ldots, n, \tag{6.48}$$
$$Y_{nj} := Y(t_{nj}) = g(t_{nj}) + \sigma^2(t_{nj})\eta(t_{nj}),\ j = 1, \ldots, k, \tag{6.49}$$

die Beobachtungen und

$$R_{ni} := X_{ni} - X_{n,i-1},\ i = 2, \ldots, n,$$
$$S_{nj} := Y_{nj} - Y_{n,j-1},\ j = 2, \ldots, k,$$

die jeweiligen Pseudo-Residuen. Neben denselben Bedingungen, die an das Modell (6.46) gestellt wurden, wird für (6.48) und (6.49) zusätzlich gefordert, dass

- $s_{nj} - t_{nj} = O(\frac{1}{n}),\ j = 1, \ldots, k$, sowie
- $\mathrm{E}(\epsilon_{ni} \cdot \eta_{nj}) = 0$ für alle $i = 1, \ldots, n$ und $j = 1, \ldots, k$.

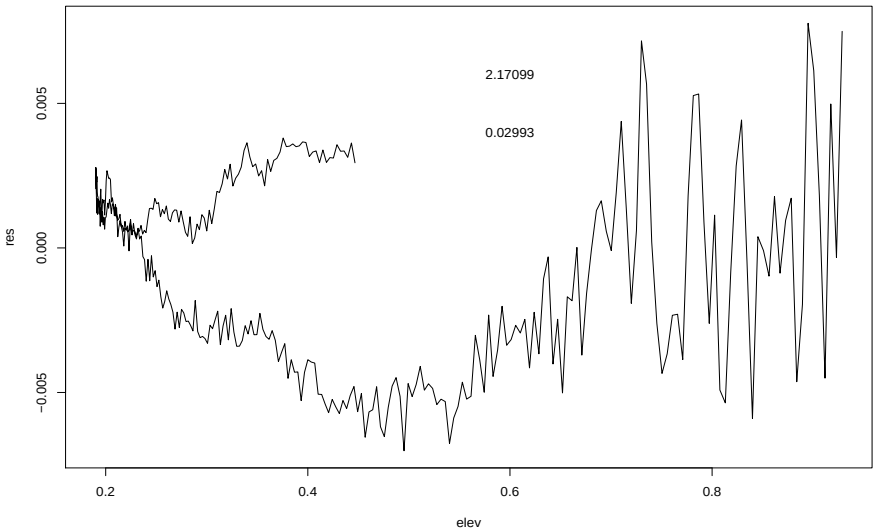

Abbildung 6.4: Eine GPS-Residuen-Zeitreihe in Abhängigkeit von der Elevation

Die neue Teststatistik lautet dann

$$\begin{aligned}
\hat{M}_n^2 := \quad & \frac{1}{4(n+k-8)} \left(\sum_{j=k+2}^{n-2} R_j^2 R_{j+2}^2 + \sum_{j=3}^{k} R_j^2 S_{j-1}^2 + \sum_{j=2}^{k-2} R_j^2 S_{j+2}^2 \right) \\
- \quad & \Bigg\{ \frac{1}{4(n+k-3)(n+k-4)} \sum_{j=3}^{n-1} \left(R_j^2 \cdot \sum_{\substack{2 \leq i \leq n \\ i \notin \{j-1, j, j+1\}}} R_i^2 \right) \\
& + 2 \cdot \left(\frac{1}{2(n+k-2)} \sum_{j=2}^{n-1} R_j^2 \right) \left(\frac{1}{2(n+k-2)} \sum_{j=3}^{k} S_j^2 \right) \\
& + \frac{1}{4(n+k-4)(n+k-3)} \sum_{j=3}^{k-1} \left(S_j^2 \cdot \sum_{\substack{2 \leq i \leq k \\ i \notin \{j-1, j, j+1\}}} S_i^2 \right) \Bigg\}.
\end{aligned}$$

Die Statistik $\hat{M}_n^2$ ist unter der Hypothese einer konstanten Varianzfunktion asymptotisch normalverteilt mit Erwartungswert 0. Im Falle normalverteilter Fehler besitzt $\hat{M}_n^2$ die Varianz

$$v_n^2 = \frac{9\ \sigma^8}{2(n+k-9)}.$$

Schätzt man $\sigma^8 = \left(\sigma^2\right)^4$ konsistent durch

$$\begin{aligned}\widehat{\sigma^8} \quad &:= \quad \frac{1}{16(n+k-14)} \\ &\quad \cdot \left(\sum_{j=k+1}^{n-6} R_j^2 R_{j+2}^2 R_{j+4}^2 R_{j+6}^2 + \sum_{j=3}^{k-2} R_{j-1}^2 R_{j+1}^2 S_{j-1}^2 S_{j+2}^2 + \sum_{j=3}^{k-2} R_{j-1}^2 R_{j+2}^2 S_{j-1}^2 S_{j+1}^2 \right),\end{aligned}$$

so gilt im Spezialfall normalverteilter Fehler

$$T := \frac{\sqrt{2(n+k-9)}}{3\sqrt{\widehat{\sigma^8}}} \cdot \hat{M}_n^2 \ \xrightarrow{\mathcal{D}}\ \mathcal{N}(0,1).$$

Man verwirft daher die Hypothese einer konstanten Varianzfunktion genau dann zum Niveau α, wenn der Wert der Teststatistik T das $(1-\frac{\alpha}{2})$-Niveau der Standard-Normalverteilung über- oder das $\frac{\alpha}{2}$-Niveau unterschreitet. In Abbildung 6.4 wird der Wert $T = 2.17099$ der Teststatistik und der zugehörige p-Wert (Definition s. Abschnitt 6.5), $p = 0.02993$, angegeben. Der Dette-Munk-Test verwirft also bei dieser Zeitreihe die Hypothese homoskedastischer Fehler zu einem 5%-Niveau.

6.4.13 Bemerkung: Im Falle $k = 0$ (1 Ast) reduziert sich T auf die in Schoknecht (2001) angegebene Teststatistik.

6.4.14 Bemerkung: Die Erfahrung hat gezeigt, dass der Dette-Munk-Test die Alternative erst bei relativ großen Stichprobenumfängen (ca. $n > 250$) einigermaßen zuverlässig erkennt.

6.5 Gütevergleich von Tests

Es sollen nun einige der in den Abschnitten 6.2 bzw. 6.4 eingeführten Tests bezüglich ihrer Güte verglichen werden.

6.5.1 Tests für *IID*-Prozesse

Es wurden jeweils drei Tests auf Basis der Autokorrelationsfunktion, der Spektraldichte bzw. des Partialsummenprozesses, sowie drei nichtparametrische Tests bezüglich verschiedener Alternativen auf ihre Güte hin untersucht. Zu jeder Alternative wurden 10 000 Zeitreihen mit unterschiedlichen Stichprobenumfängen ($n = 50, 100, 200$) simuliert und der Anteil der zum Niveau $\alpha = 0.05$ abgelehnten Hypothesen ermittelt.

Für den Gütevergleich wurden die folgenden Alternativen ausgewählt:

$$
\begin{aligned}
(X_t) &\sim AR(1) \quad \text{mit} \quad \phi_1 = 0.9, \\
(X_t) &\sim AR(1) \quad \text{mit} \quad \phi_1 = -0.9, \\
(X_t) &\sim MA(1) \quad \text{mit} \quad \phi_1 = 0.9, \\
(X_t) &\sim MA(1) \quad \text{mit} \quad \phi_1 = -0.9.
\end{aligned}
$$

Zusammenfassend lässt sich sagen, dass große Güteunterschiede, teilweise in Abhängigkeit der Alternative und / oder des Stichprobenumfangs, festgestellt werden konnten. Es folgt nun eine detailliertere Beschreibung der Simulationsergebnisse:

Tests aufgrund der empirischen Autokorrelationsfunktion

Als einziger Test dieser Kategorie erkannte der von Neumann Ratio (VNR) alle vorliegenden Alternativen zuverlässig. Man beachte dabei, dass hier mit dem VNR einseitige Alternativen getestet wurden, d.h. es wurde vorausgesetzt, dass bekannt ist, ob $\rho(1) > 0$ oder $\rho(1) < 0$. In der Praxis dürfte dies jedoch keine große Einschränkung darstellen, da $\rho(1)$ durch $\hat{\rho}(1)$ erwartungstreu und konsistent geschätzt werden kann.

Der auf der Teststatistik $\hat{\rho}(1)$ beruhende Test sowie der Portmanteau-Test zeigten sich gegen die $AR(1)$-Alternativen ebenfalls recht zuverlässig, schnitten jedoch bei den $MA(1)$-Prozessen relativ schlecht ab. In Tabelle 6.9 ist der Anteil der erkannten Alternativen zum Niveau $\alpha = 0.05$ von jeweils 10 000 simulierten Prozessen aufgelistet (gerundet).

Tabelle 6.9: Untersuchung der Güte der auf der empirischen ACF basierenden Tests

		$\hat{\rho}(1)$	VNR (einseitig)	Portmanteau
	H_0	$\rho(1) = 0$	$\rho(1) \leq 0$ $\rho(1) \geq 0$	$\rho(h) = 0\ \forall h \geq 1$
Simulation	$n \setminus H_1$	$\rho(1) \neq 0$	$\rho(1) > 0$ $\rho(1) < 0$	$\exists h > 0 : \rho(h) \neq 0$
$AR(1)$,	50	0.97	1.00 -	0.94
$\phi_1 = 0.9$	100	1.00	- -	1.00
	200	1.00	- -	1.00
$AR(1)$,	50	0.99	- 1.00	0.98
$\phi_1 = -0.9$	100	1.00	- -	1.00
	200	1.00	- -	1.00
$MA(1)$,	50	0.10	0.99 -	0.33
$\theta_1 = 0.9$	100	0.10	- -	0.35
	200	0.11	- -	0.37
$MA(1)$,	50	0.09	- 0.98	0.31
$\theta_1 = -0.9$	100	0.10	- -	0.35
	200	0.10	- -	0.37

Tests aufgrund der empirischen Spektraldichte

Von allen untersuchten Tests, die die Hypothese gegen die allgemeine Alternative

$$H_1 : es\ existiert\ ein\ h > 0 : \rho(h) \neq 0$$

testen, erreichte in den Simulationen der Cramér-von Mises-Test die besten Ergebnisse, direkt gefolgt vom Kolmogorov-Smirnov-Test. Der Test von Fisher war gegen die $AR(1)$-Alternativen ebenfalls sehr gut, bei $MA(1)$-Prozessen jedoch verhältnismäßig schwach. Tabelle 6.10 zeigt den Anteil der erkannten Alternativen zum Niveau $\alpha = 0.05$ (ebenfalls gerundet).

Tabelle 6.10: Gütevergleich: Tests aufgrund der empirischen Spektraldichte

		KS (Spektrald.)	CM (Spektrald.)	Fisher
	H_0	$\rho(h) = 0\ \forall h \geq 1$	$\rho(h) = 0\ \forall h \geq 1$	$\rho(h) = 0\ \forall h \geq 1$
Simulation	$n \setminus H_1$	$\exists h > 0 : \rho(h) \neq 0$	$\exists h > 0 : \rho(h) \neq 0$	$\exists h > 0 : \rho(h) \neq 0$
$AR(1)$,	50	1.00	1.00	0.96
$\phi_1 = 0.9$	100	1.00	1.00	1.00
	200	1.00	1.00	1.00
$AR(1)$,	50	1.00	1.00	0.98
$\phi_1 = -0.9$	100	1.00	1.00	1.00
	200	1.00	1.00	1.00
$MA(1)$,	50	0.85	0.97	0.33
$\theta_1 = 0.9$	100	1.00	1.00	0.40
	200	1.00	1.00	0.48
$MA(1)$,	50	0.88	0.98	0.33
$\theta_1 = -0.9$	100	1.00	1.00	0.41
	200	1.00	1.00	0.49

Auf dem Partialsummenprozess beruhende Tests

Die auf dem Partialsummenprozess (PSP) basierenden Tests sind offenbar nicht in der Lage, die negativ korrelierten Alternativen $AR(1)$ mit $\phi_1 < 0$ bzw. $MA(1)$ mit $\theta_1 < 0$ zu erkennen. Auch gegen den $MA(1)$-Prozess mit $\theta_1 > 0$ waren sie in den Simulationen äußerst schwach.

Lediglich die $AR(1)$-Alternative mit $\phi_1 > 0$ wurde von diesen Tests erkannt. Dabei waren die Ergebnisse des $B(1)$-Tests eher mäßig, jene von Cramér-von Mises- und Kolmogorov-Smirnov-Test recht gut, besonders für höhere Stichprobenumfänge.

Tabelle 6.11: Gütevergleich: Tests basierend auf dem Partialsummenprozess

		$B(1)$	KS (PSP)	CM (PSP)
	H_0	$\rho(h) = 0\ \forall h \geq 1$	$\rho(h) = 0\ \forall h \geq 1$	$\rho(h) = 0\ \forall h \geq 1$
Simulation	$n \setminus H_1$	$\exists h > 0 : \rho(h) \neq 0$	$\exists h > 0 : \rho(h) \neq 0$	$\exists h > 0 : \rho(h) \neq 0$
$AR(1)$,	50	0.68	0.87	0.88
$\phi_1 = 0.9$	100	0.67	0.95	0.94
	200	0.66	0.99	0.97
$AR(1)$,	50	0.00	0.00	0.00
$\phi_1 = -0.9$	100	0.00	0.00	0.00
	200	0.00	0.00	0.00
$MA(1)$,	50	0.17	0.18	0.20
$\theta_1 = 0.9$	100	0.17	0.22	0.21
	200	0.17	0.25	0.22
$MA(1)$,	50	0.00	0.00	0.00
$\theta_1 = -0.9$	100	0.00	0.00	0.00
	200	0.00	0.00	0.00

Nichtparametrische Tests

Von den nichtparametrischen Tests kann lediglich der Turning-Point-Test mit Einschränkung empfohlen werden. Dieser Test erzielte bei den meisten Alternativen ein für ein nichtparametrisches Verfahren überraschend gutes Ergebnis. Eine Ausnahme war der negativ korrelierter $MA(1)$-Prozess. Bei kleinen bis mittelgroßen Stichprobenumfängen war der Test dort sehr schwach. Bei sehr großen Stichprobenumfängen wurden die Ergebnisse etwas besser.

Kendall's Rangtest hatte in den Simulationen nur bei dem positiv korrelierten $AR(1)$-Prozess (mäßigen) Erfolg. Der Vorzeichen-Test erwies sich gegen die vorliegenden Alternativen als völlig unbrauchbar.

Tabelle 6.12: Gütevergleich: Nichtparametrische Tests auf iid-Verteilung

		Kendall	Turning Point	Vorzeichen
Simulation	H_0 $n \setminus H_1$	$\rho(h) = 0\ \forall h \geq 1$ $\exists h > 0 : \rho(h) \neq 0$	$\rho(h) = 0\ \forall h \geq 1$ $\exists h > 0 : \rho(h) \neq 0$	$\rho(h) = 0\ \forall h \geq 1$ $\exists h > 0 : \rho(h) \neq 0$
$AR(1)$, $\phi_1 = 0.9$	50	0.63	0.70	0.10
	100	0.64	0.92	0.07
	200	0.65	1.00	0.07
$AR(1)$, $\phi_1 = -0.9$	50	0.00	0.97	0.00
	100	0.00	1.00	0.00
	200	0.00	1.00	0.00
$MA(1)$, $\theta_1 = 0.9$	50	0.15	0.83	0.15
	100	0.16	0.99	0.03
	200	0.16	1.00	0.05
$MA(1)$, $\theta_1 = -0.9$	50	0.00	0.19	0.03
	100	0.00	0.32	0.02
	200	0.00	0.61	0.04

6.5.2 Tests auf Homogenität der Varianz

Zwei-Stichproben-Tests

Sind die Voraussetzungen an die Zufallsvariablen wie in Abschnitt 6.4.1 (insbesondere die Voraussetzung der Normalverteilung) erfüllt, so ist unter den Zwei-Stichproben-Tests zum Testen der Hypothese $H_0 : \sigma_1^2 \leq \sigma_2^2$ gegen die Alternative $H_1 : \sigma_1^2 > \sigma_2^2$ der F-Test gleichmäßig bester Signifikanztest zum Niveau α. Zum Testen der Hypothese $H_0 : \sigma_1^2 = \sigma_2^2$ gegen die Alternative $H_1 : \sigma_1^2 \neq \sigma_2^2$ hingegen ist der β-Test gleichmäßig optimal zum Niveau α (jeweils unter Annahme einer *Neyman-Pearson-Schadensfunktion*, vgl. Lehmann (1986)).

Multiple Tests

Hartley (1950) vergleicht seinen in derselben Arbeit erschienenen Maximum-F-Ratio-Test mit dem Bartlett-Test und stellt fest, dass letzterer unter den dort gewählten Alternativen eine etwas höhere Güte besitzt. Ein UMP Test existiert in der Klasse der multiplen Tests auf Homoskedastizität jedoch nicht (s. Lehmann (1986)).

In selbst durchgeführten Simulationsstudien wurde der Dirichlet-Test von Wludyka und Nelson (6.42) mit dem multiplen Beta-Test (6.43) und dem Bartlett-Test verglichen. Als Alternativen wurden die Varianzfunktionen

$$g_1(t) = \begin{cases} 1 & ,t \leq \frac{n}{2}, \\ \frac{1}{2} & ,t > \frac{n}{2}, \end{cases} \quad \text{sowie } g_2(t) = \begin{cases} \frac{1}{2} & ,t \leq \frac{n}{2}, \\ 1 & ,t > \frac{n}{2}, \end{cases}$$

betrachtet. Für die Anzahl der Teilvektoren k und die Anzahl der Freiheitsgrade ν wurden die Werte $k = 3, 4, 6, 10$ bzw. $\nu = 5, 10, 15, 35$ gewählt.

Bei diesen Simulationsstudien erzielte der Bartlett-Test eine höhere Güte als der Dirichlet-Test. Während der Unterschied bei kleinen Werten von k gering ausfällt, schneidet der Bartlett-Test für große k wesentlich besser ab als der Dirichlet-Test.

Die Güte des multivariaten Beta-Tests hängt deutlich von der Alternative ab. Bezüglich der fallenden Varianzfunktion g_1 ist die Güte des multivariaten Beta-Tests i.d.R. höher als die des Bartlett-Tests. Ausgenommen hiervon sind sehr große Stichprobenumfänge. Im Falle der steigenden Varianzfunktion g_2 schneidet der Bartlett-Test i.d.R. besser ab als der multivariate Beta-Test. Eine Ausnahme bilden hier sehr kleine Stichprobenumfänge.

Die Güte des Dirichlet-Tests überschreitet für keine der beiden Alternativen die Güte des multivariaten Beta-Tests.

Tabelle 6.13 zeigt den Anteil der erkannten Alternativen zum Niveau $\alpha = 0.05$ von jeweils 10 000 Simulationen.

Tabelle 6.13: Gütevergleich: Bartlett-, Dirichlet- und multivariater Beta-Test

		Bartlett	mult. Beta	Dirichlet
Sim.	$\nu \setminus k$	3 4 6 10	3 4 6 10	3 4 6 10
g_1	5	0.07 0.09 0.10 0.12	0.14 0.13 0.15 0.19	0.07 0.07 0.07 0.07
	10	0.13 0.18 0.22 0.28	0.14 0.22 0.26 0.32	0.12 0.14 0.15 0.14
	15	0.19 0.28 0.34 0.45	0.21 0.32 0.38 0.45	0.16 0.22 0.22 0.23
	25	0.29 0.49 0.58 0.75	0.32 0.49 0.57 0.68	0.28 0.38 0.38 0.42
	35	0.42 0.65 0.76 0.90	0.40 0.64 0.73 0.83	0.40 0.52 0.54 0.58
g_2	5	0.08 0.09 0.10 0.12	0.12 0.08 0.08 0.08	0.07 0.07 0.07 0.07
	10	0.13 0.18 0.21 0.29	0.12 0.15 0.14 0.14	0.12 0.14 0.14 0.15
	15	0.18 0.29 0.34 0.45	0.19 0.23 0.24 0.25	0.17 0.22 0.22 0.23
	25	0.30 0.48 0.59 0.74	0.28 0.40 0.43 0.46	0.29 0.37 0.39 0.42
	35	0.40 0.66 0.77 0.90	0.39 0.57 0.60 0.68	0.40 0.52 0.54 0.59

CUSUMSQ-, MOSUMSQ und Dette-Munk-Test

Vergleicht man CUSUMSQ-, MOSUMSQ- und Dette-Munk-Test bezüglich der Alternativen g_1 und g_2, so schneidet der CUSUMSQ-Test in allen Fällen deutlich besser ab als der MOSUMSQ-Test. Dabei sind die Ergebnisse bei fallender und steigender Varianz vergleichbar. Dasselbe gilt für die linearen Varianzfunktionen

$$g_3(t) = 2 - \frac{t}{n} \text{ sowie } g_4(t) = 1 + \frac{t}{n},\ t = 1, \ldots, n.$$

Der Dette-Munk-Test weist bei kleinen bis mittelgroßen Stichprobenumfängen sowohl eine hohe Wahrscheinlichkeit für einen Fehler erster Art auf, als auch - bei den vorliegenden Alternativen - eine sehr geringe Güte. Verhältnismäßig gering schwankende Varianzfunktionen wie die hier vorgestellten lassen sich also mit diesem Test quasi nicht von der Nullhypothese unterscheiden (vgl. Tabelle 6.14).

Tabelle 6.14: Gütevergleich: CUSUMSQ-, MOSUMSQ- und Dette-Munk-Test

		CUSUMSQ	MOSUMSQ	Dette-Munk
Simulation	$\nu \setminus \alpha$	0.10 0.05 0.01	0.10 0.01	0.10 0.05 0.01
H_0	40	0.11 0.06 0.01	0.07 0.01	0.27 0.14 0.10
	80	0.12 0.06 0.01	0.07 0.01	0.24 0.11 0.07
	120	0.11 0.06 0.01	0.07 0.01	0.23 0.10 0.06
	200	0.11 0.06 0.01	- -	0.22 0.08 0.05
g_1	40	0.41 0.28 0.11	0.23 0.06	0.27 0.14 0.10
	80	0.63 0.51 0.27	0.27 0.08	0.25 0.11 0.07
	120	0.78 0.69 0.44	0.31 0.08	0.25 0.09 0.06
	200	0.94 0.89 0.74	- -	0.25 0.09 0.05
g_3	40	0.49 0.36 0.16	0.24 0.07	0.27 0.14 0.10
	80	0.72 0.60 0.35	0.30 0.09	0.24 0.11 0.07
	120	0.87 0.78 0.55	0.36 0.12	0.23 0.09 0.06
	200	0.97 0.95 0.82	- -	0.22 0.08 0.04

Erfahrungsgemäß lässt sich der Dette-Munk-Test erst ab Stichprobenumfängen von ca. $n = 250$ bei stark ausgeprägten Alternativen (z.B. exponentiell fallende Varianz) sinnvoll einsetzen.

Multiple F-Tests gegen fallende Alternativen

Ein Vergleich der beiden multiplen F-Tests mit variabler Trennlinie (einseitige Alternativen, vgl. Abschnitt 6.4.4) zeigt bezüglich der fallenden Varianzfunktionen g_1 und g_3 eine in der Regel höhere Güte des von Hsu (1977) vorgeschlagenen Verfahrens. Lediglich bei großen Stichprobenumfängen erreicht die P_{min}-Teststatistik im Falle der Varianzfunktion g_1 eine durchschnittlich höhere Güte, siehe Tabelle 6.15.

Tabelle 6.15: Gütevergleich multipler F-Tests: P_{min} und Hsu-Teststatistik

		P_{min} (einseitig)	Hsu (einseitig)
Simulation	$\nu \setminus \alpha$	0.10 0.05 0.01	0.10 0.05 0.01
g_1	40	0.42 0.28 0.10	0.47 0.32 0.11
	80	0.61 0.47 0.23	0.65 0.49 0.20
	120	0.75 0.62 0.39	0.78 0.63 0.31
	200	0.91 0.85 0.69	0.91 0.82 0.49
g_3	40	0.27 0.16 0.05	0.33 0.21 0.07
	80	0.37 0.24 0.08	0.47 0.33 0.11
	120	0.44 0.30 0.12	0.56 0.42 0.17
	200	0.59 0.45 0.24	0.74 0.60 0.29

Kapitel 7. Erzeugung von Homoskedastizität

Verwirft ein Test auf Homoskedastizität die Null-Hypothese einer konstanten Varianz, so können in der Regel durch Transformation und/oder Gewichtung homoskedastische Beobachtungen erzeugt werden. Es folgt nun ein kurzer Überblick zu diesem Themenbereich. Weiterführende Informationen findet man z.B. in Carroll / Ruppert (1988) oder Seber / Wild (1989).

7.1 Transformation

Eine Transformation des Beobachtungsvektors $Y_n := (Y_{n1}, \ldots, Y_{nn})^\top$ empfiehlt sich in erster Linie bei nichtlinearen Modellen und zur Erzeugung annähernd normalverteilter Daten. Schwankt die Varianz eines nichtlinearen Modells proportioanal zur Erwartungswertfunktion, so kann durch Transformation unter Umständen Homoskedastizität erzeugt werden.

Transformation Both Sides (TBS)

Bei manchen Anwendungen ist ein funktionaler Zusammenhang des Beobachtungsvektors bekannt, z.B.

$$Y_{ni} \approx f(x_{ni}, \beta),\ i = 1, \ldots, n,$$

wobei β ein unbekannter Parametervektor ist. In diesen Fällen kann es sinnvoll sein, sowohl den Beobachtungsvektor als auch das funktionale Modell zu transformieren. Man sucht zu diesem Zweck eine Transformation h, so dass das transformierte Modell

$$h(Y_{ni}) = h(f(x_{ni}, \beta)) + \epsilon_{ni},\ i = 1, \ldots, n,$$

annähernd normalverteilte und homoskedastische Fehler besitzt.

Box-Cox-Transformationen

Eine gängige Familie von Transformationen sind die *Box-Cox-Transformationen* (siehe Box / Cox (1964))

$$Y_{ni}^{(\lambda_1,\lambda_2)} := \begin{cases} \frac{(Y_{ni}+\lambda_2)^{\lambda_1}-1}{\lambda_1}\ , & \lambda_1 \neq 0 \\ \ln(Y_{ni}+\lambda_2)\ , & \lambda_1 = 0 \end{cases},\ i = 1, \ldots, n.$$

Dabei sind die Parameter λ_1 und λ_2 so zu wählen, dass die Verteilung des transformierten Datenvektors $Y_n^{(\lambda_1,\lambda_2)} := (Y_{n1}^{(\lambda_1,\lambda_2)}, \ldots, Y_{nn}^{(\lambda_1,\lambda_2)})^\top$ einer Normalverteilung möglichst nahe kommt. In Seber / Wild (1989) wird ein Verfahren zur Bestimmung von evtl. existierenden optimalen Transformationsparametern angegeben. Häufig existieren solche optimalen Parameter jedoch nicht.

7.1.1 Beispiel: Die Logarithmus-Transformation

$$(Y_{n1}, \ldots, Y_{nn})^\top \longrightarrow (\ln Y_{n1}, \ldots, \ln Y_{nn})^\top$$

wird häufig bei Wachstumsdaten, wie z.B. Börsenindizes, eingesetzt. Sind $X_{n1}, \ldots, X_{nn}$ unabhängig und identisch verteilte Zufallsvariablen und existieren Skalare $\tau_{ni} > 0$ mit $Y_{ni} = \tau_{ni} X_{ni}, i = 1, \ldots, n$, so gilt

$$\mathrm{Var}(\ln Y_{ni}) = \mathrm{Var}(\ln \tau_{ni} + \ln X_{ni}) = \mathrm{Var}(\ln X_{ni}).$$

Damit besitzt der Log-transformierte Datenvektor $(\ln Y_{n1}, \ldots, \ln Y_{nn})^\top$ eine konstante Varianz.

7.1.2 Bemerkung: Die Box-Cox-Transformation unterscheidet sich von der oben beschriebenen TBS in erster Linie dadurch, dass lediglich der Beobachtungsvektor transformiert wird. Man geht bei einer Box-Cox-Transformation davon aus, dass der **transformierte** Datenvektor durch ein lineares Modell

$$Y_n^{(\lambda_1,\lambda_2)} = X_n \beta + \epsilon_n$$

mit annähernd homoskedastischen Fehlern $\epsilon_{ni},\ i = 1, \ldots, n$, beschrieben werden kann.

7.1.3 Bemerkung: Die Normalverteilungsannahme kann anhand der (Pseudo-) Residuen mit den Methoden aus Abschnitt 6.3 überprüft werden.

John-Draper-Transformation

Eine leicht abgewandelte Form der Box-Cox-Transformationen sind die *John-Draper-Transformationen*

$$Y_{ni}^{\lambda} := \begin{cases} \operatorname{sgn}(Y_{ni}) \cdot \frac{|Y_{ni}|^{\lambda}-1}{\lambda} , & \lambda \neq 0 \\ \operatorname{sgn}(Y_{ni}) \cdot \ln |Y_{ni}| , & \lambda = 0 \end{cases} , \; i = 1, \ldots, n,$$

wobei $\operatorname{sgn}(Y_{ni})$ das Vorzeichen von Y_{ni} bezeichnet. Auch hier wird der Parameter λ so gewählt, dass der transformierte Beobachtungsvektor annähernd normalverteilt ist.

7.2 Gewichtung

Wie bereits erwähnt, ist eine Transformation der Daten nur bei nichtlinearen Modellen empfehlenswert. Liegt bereits ein lineares Modell vor, oder ergibt eine Transformation zwar Linearität, nicht jedoch Homoskedastizität, so kann eine (evtl. zusätzliche) Gewichtung der Daten sinnvoll sein. Hierfür muss in der Regel zunächst die Varianzfunktion geschätzt werden.

7.2.1 Schätzung der Varianzfunktion

Häufig besteht bereits eine Vorstellung über die etwaige Form der Varianzfunktion. So kann z.B. eine Veränderung im Versuchsablauf (Wartung oder Austausch von Versuchsgeräten, etc.) zu einem Sprung in der Varianzfunktion führen. In anderen Fällen liegt aufgrund von theoretischen Überlegungen eine Vermutung über die Form der Heteroskedastizität vor. Wir wollen uns zunächst mit dem Spezialfall einer zweistufigen Varianzfunktion (Sprungfunktion) beschäftigen.

Zweistufige Varianzfunktion

Gegeben sei das lineare Modell $Y_n = X_n\beta + \epsilon_n$ mit einer zweistufigen Varianzfunktion, d.h.

$$\operatorname{Var}(Y_{ni}) = \sigma_1^2, \; i = 1, \ldots, t^*, \tag{7.1}$$
$$\operatorname{Var}(Y_{ni}) = \sigma_2^2, \; i = t^*+1, \ldots, n, \tag{7.2}$$

wobei $1 < t^* < n$ und $\sigma_1^2 \neq \sigma_2^2$. Ist t^* bekannt, so können σ_1^2 und σ_2^2 leicht mittels der (rekursiven) Residuen und des erwartungstreuen Schätzers (6.30) geschätzt werden. Ist t^* jedoch unbekannt, so können u.U. die mehrdimensionalen Teststatistiken aus den Abschnitten 6.4.3 und 6.4.4 bei der Bestimmung von t^* hilfreich sein. Diese Teststatistiken weisen in der Nähe von t^* in der Regel das folgende Verhalten auf:

- Die CUSUMSQ-Teststatistik überschreitet die obere (das bedeutet fallende Varianz) bzw. die untere Konfidenzschranke (bei steigender Varianz) um einen maximalen Wert.
- Die MOSUMSQ-Teststatistik fällt rapide ab (bei fallender Varianz) bzw. steigt rapide an (bei steigender Varianz).
- Die multiple Beta-Teststatistik überschreitet an einer zu t^* korrespondierenden Stelle (im Fall des Beispiels 7.2.1 in der Mitte) die obere (das bedeutet fallende Varianz) bzw. die untere Konfidenzschranke (bei steigender Varianz).
- Die F-Statistik mit variabler Trennlinie überschreitet die obere (d.h. fallende Varianz) bzw. die untere Konfidenzschranke (bei steigender Varianz) maximal.

7.2.1 Beispiel: Es sei $n = 120$ und

$$Z_t \overset{iid}{\sim} \mathcal{N}(0,1), \; t = 1, \ldots, 60,$$
$$Z_t \overset{iid}{\sim} \mathcal{N}(0, \frac{1}{2}), \; t = 61, \ldots, 120.$$

Abbildung 7.1 zeigt die CUSUMSQ-, MOSUMSQ-, multiple Beta- sowie die F-Statistik (F-Test mit variabler Trennlinie) für eine Realisierung von (Z_t). Für Beta- und F-Test wurden einseitige Konfidenzschranken gewählt. In diesem Beispiel erwiesen sich die CUSUMSQ-Teststatistik, die multiple Beta-Teststatistik (einseitige Alternativen) und die F-Statistik mit variabler Trennlinie (einseitige Alternativen) als relativ zuverlässig, die MOSUMSQ-Teststatistik lieferte im Vergleich häufiger irreführende Ergebnisse.

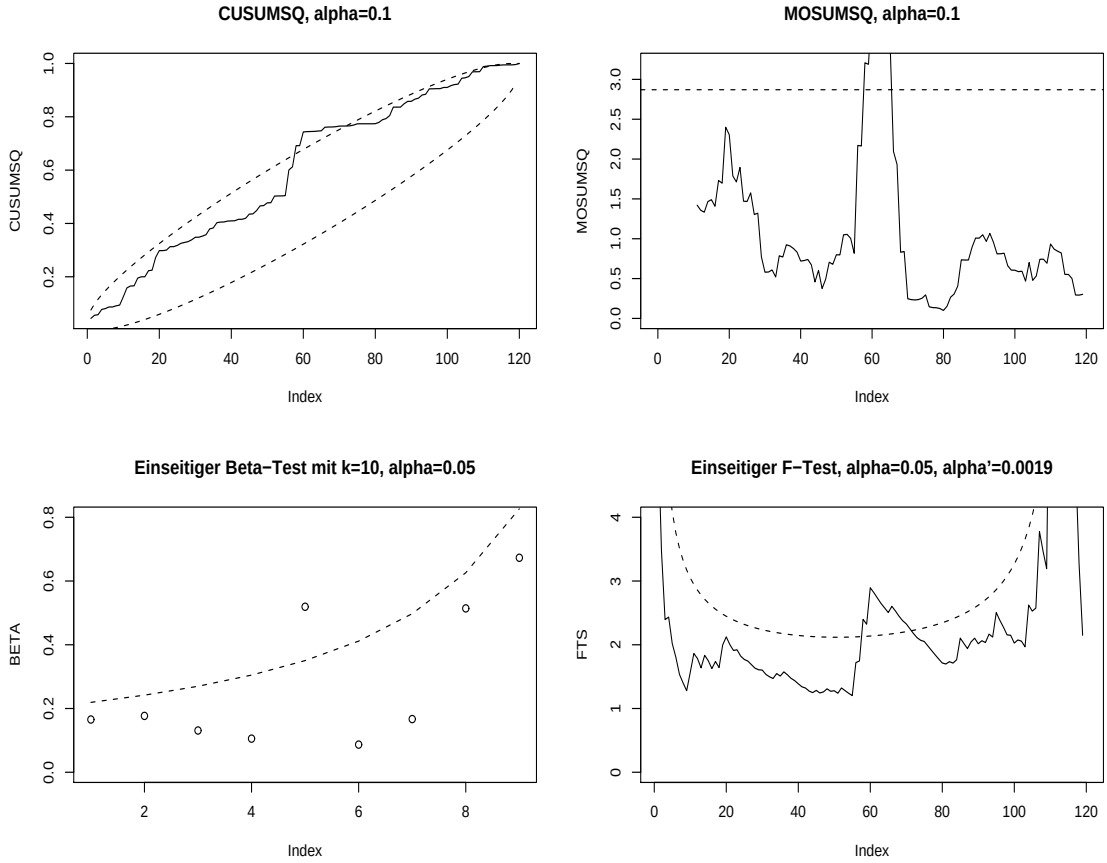

Abbildung 7.1: Die CUSUMSQ-, MOSUMSQ-, mult. Beta und mult. F-Statistik

Allgemeine Form einer Varianzfunktion

Ein heteroskedastisches lineares Modell kann man schreiben gemäß

$$Y_{ni} = (X_n(i))^\top \beta + \sigma_{ni}\epsilon_{ni},\ \sigma_{ni} > 0,\ \mathrm{E}(\epsilon_{ni}) = 0, \tag{7.3}$$
$$\mathrm{Cov}(\epsilon_{ni}, \epsilon_{nj}) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad (i = 1, \ldots, n),$$

wobei $(X_n(i))^\top$ die i-te Zeile von X_n bezeichnet. Im Allgemeinen geht man bei Vorliegen von Heteroskedastizität davon aus, dass sich die Varianzfunktion mit dem Erwartungswert $\mathrm{E}(Y_{ni})$ des Regressionsmodells verändert und von unbekannten Parametern abhängt. Außerdem sind weitere Einflussfaktoren denkbar. Beispielsweise nimmt man bei geodätischen GPS-Messungen eine Abhängigkeit der Varianz von den Satellitenelevationen an. Als allgemeine Form einer Varianzfunktion wählt man daher

$$\sigma_{ni}^2 = g^2(\mathrm{E}(Y_{ni}), z, \theta),\ i = 1, \ldots, n. \tag{7.4}$$

Dabei ist g^2 eine unbekannte, positive Funktion, θ ein unbekannter Parametervektor und z ein Vektor sämtlicher sonstiger Einflussfaktoren.

Für das lineare Modell (7.3) ist der gewichtete LS-Schätzer

$$\hat{\beta}_G = (X_n^\top \Sigma_n^{-1} X_n)^{-1} X_n^\top \Sigma_n^{-1} Y_n \ , \ \Sigma_n = \mathrm{diag}(\sigma_{n1}^2, \ldots, \sigma_{nn}^2),$$

der Best Linear Unbiased Estimator (BLUE). Doch die Fehlervarianzen $\sigma_{n1}^2, \ldots, \sigma_{nn}^2$ und somit Σ_n sind unbekannt, weshalb üblicherweise der herkömmliche LS-Schätzer

$$\hat{\beta} = (X_n^\top X_n)^{-1} X_n^\top Y_n \tag{7.5}$$

verwendet wird. Dieser Schätzer ist BLUE für das homoskedastische Regressionsmodell

$$\begin{aligned} &Y_{ni} = (X_n(i))^\top \beta + \sigma\epsilon_{ni},\ \sigma > 0,\ \mathrm{E}(\epsilon_{ni}) = 0, \\ &\mathrm{Cov}(\epsilon_{ni}, \epsilon_{nj}) = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad (i = 1, \ldots, n). \end{aligned}$$

Die mit dem LS-Schätzer (7.5) berechneten Residuen $r_{ni} = Y_{ni} - (X_n(i))^\top \hat{\beta},\ i = 1, \ldots, n$, spiegeln die bei der Berechnung von (7.5) nicht berücksichtigten Varianzen $\sigma_{n1}^2, \ldots, \sigma_{nn}^2$ des Modells (7.4) wider. Aus diesem Grund ist die Schätzung der Varianzfunktion durch Schätzen von

$$\mathrm{Var}(r_{ni}) = \mathrm{E}(r_{ni}^2)$$

eine naheliegende Vorgehensweise. Seber und Wild (1989) schlagen unter anderem vor, die Varianzfunktion durch Schätzen des Erwartungswertes der quadrierten Residuen (mittels **nicht**linearer Regression) zu schätzen. Alternativ kann versucht werden, durch eine (z.B. in Abschnitt 7.1 beschriebene) Transformation der quadrierten Residuen ein lineares Modell mit zumindest symmetrisch verteilten Fehlern zu erzeugen und dort eine **lineare** Regression durchzuführen. Etwas Aufmerksamkeit verdient in diesem Zusammenhang die Logarithmus-Transformation

$$T(r_{ni}^2) = \ln(r_{ni}^2). \tag{7.6}$$

Wie bereits in Abschnitt 7.1 erwähnt wurde, besitzen die Log-transformierten quadrierten Residuen eine konstante Varianz (siehe auch Carroll / Ruppert (1988)). Dies kann gerade zum Schätzen der unbekannten Varianzfunktion von Vorteil sein. Da ein Residuum nahe dem Wert 0 nach der Transformation (7.6) betragsmäßig sehr große Werte im negativen Bereich annehmen kann, sind die Log-transformierten quadrierten Residuen i.A. jedoch sehr unsymmetrisch verteilt. Abhilfe kann in diesem Fall die Box/Cox-Transformation

$$T(r_{ni}^2) = \ln(r_{ni}^2 + \lambda_2) \tag{7.7}$$

schaffen. Dabei gewährleistet $\lambda_2 > 0$ eine untere Schranke für die transformierten Daten und kann so gewählt werden, dass die transformierten Daten in etwa symmetrisch um ihren Median verteilt sind. Wählt man z.B. λ_2 gemäß

$$\lambda_2 := \frac{(\mathrm{med}(r^2))^2}{\max(r^2)}, \tag{7.8}$$

so erzielt man damit

$$\ln(\max(r^2)) - \ln(\mathrm{med}(r^2)) = \ln(\mathrm{med}(r^2)) - \ln(\lambda_2). \tag{7.9}$$

Sind die Residuen symmetrisch um 0 normalverteilt, kann man davon ausgehen, dass $\min(r^2)$ einen Wert nahe 0 annimmt. Ferner lässt sich λ_2 "groß" wählen im Vergleich zu $\min(r^2)$, aber "klein" im Verhältnis zu $\mathrm{med}(r^2)$ und $\max(r^2)$. Die Eigenschaft (7.9) lässt sich also ersetzen durch

$$\ln(\max(r^2) + \lambda_2) - \ln(\mathrm{med}(r^2) + \lambda_2) \approx \ln(\mathrm{med}(r^2) + \lambda_2) - \ln(\min(r^2) + \lambda_2). \tag{7.10}$$

Wegen

$$\max(T(r^2)) = T(\max(r^2)),\ \mathrm{med}(T(r^2)) = T(\mathrm{med}(r^2)),\ \text{etc.}$$

ist (7.10) gleichbedeutend mit

$$\max(T(r^2)) - \mathrm{med}(T(r^2)) \approx \mathrm{med}(T(r^2)) - \min(T(r^2)),$$

d.h. die transformierten Daten sind in etwa symmetrisch um ihren Median verteilt. In Experimenten wurden mit dieser einfachen Box/Cox-Transformation recht gute Erfahrungen gemacht. Sie wird daher in BISCHOFF ET AL. (2006) zur Schätzung der Varianzfunktion geodätischer GPS-Messreihen verwendet. Jedoch auch die Transformation

$$T(r_{ni}^2) = |r_{ni}|^{\frac{1}{2}}$$

erzielte bei der Anwendung auf quadrierte (normalverteilte) Residuen annähernd symmetrisch verteilte Daten. Einen Überblick über verschiedene Transformationen zur Schätzung der Varianzfunktion und deren Eigenschaften findet man z.B. in CARROLL UND RUPPERT (1988).

7.2.2 Beispiel: Wir wollen die GPS-Residuen-Zeitreihe aus Abbildung 6.4 aufgreifen. In Abschnitt 6.4.6 verwarf der Dette-Munk-Test die Hypothese, dass diese Zeitreihe homoskedastisch sei, zum Niveau 0.05. Wir wollen nun eine geeignete Gewichtsfunktion für diese Zeitreihe finden.

Da die Residuen-Zeitreihe aus Abbildung 6.4 einen starken Trend aufweist, eignen sich die Residuen in diesem Falle nicht für das oben beschriebene Vorgehen zur Schätzung der Varianzfunktion. Wir behandeln daher die Residuen-Zeitreihe wie eine herkömmliche Zeitreihe mit unbekanntem Trend und bilden wie in Abschnitt 6.4.6 auf jedem Ast gesondert die Pseudo-Residuen $R_{ni} = r_{ni} - r_{n,i-1},\ i = 2, \ldots, n$. Der Erwartungswert $\mathrm{E}(R_{ni})$ wird als "klein" angenommen und bei der Schätzung der Varianzfunktion vernachlässigt. Damit können wegen (5.26) die Größen $\frac{1}{2}R_{ni}^2,\ i = 2, 4, \ldots, 2[\frac{n}{2}]$, zum Schätzen der Varianzfunktion verwendet werden.

Bei geodätischen GPS-Messungen geht man von einer Abhängigkeit der Varianz von den Elevationen der Satelliten aus, nicht aber von einem Einfluss des Erwartungswertes $\mathrm{E}(Y_{ni})$. Da es sich in unserem Beispiel um Residuen doppeltdifferenzierter GPS-Beobachtungen handelt, muss man die Elevationen zweier Satelliten bzgl. jeweils zweier GPS-Empfänger berücksichtigen. Ist $el_{A_k}^{S_j}(t_{ni})$ die Elevation (in Grad) des Satelliten S_j bzgl. des Empfängers A_k zum Zeitpunkt $t_{ni},\ i = 1, \ldots, n,\ j, k = 1, 2$, so bezeichnet im Folgenden

$$zd_{A_k}^{S_j}(t_{ni}) := \frac{90 - el_{A_k}^{S_j}(t_{ni})}{90},\ i = 1, \ldots, n,$$

die entsprechende Zenitdistanz, auf das Intervall $[0, 1]$ transformiert $(j, k = 1, 2)$.

Zur Modellierung der Varianzfunktion sind verschiedene Ansätze denkbar. Zunächst entscheiden wir uns für eine Funktion der Satellitenelevationen, der Kürze halber "$ELEV$" genannt, und eine geeignete Transformation T.

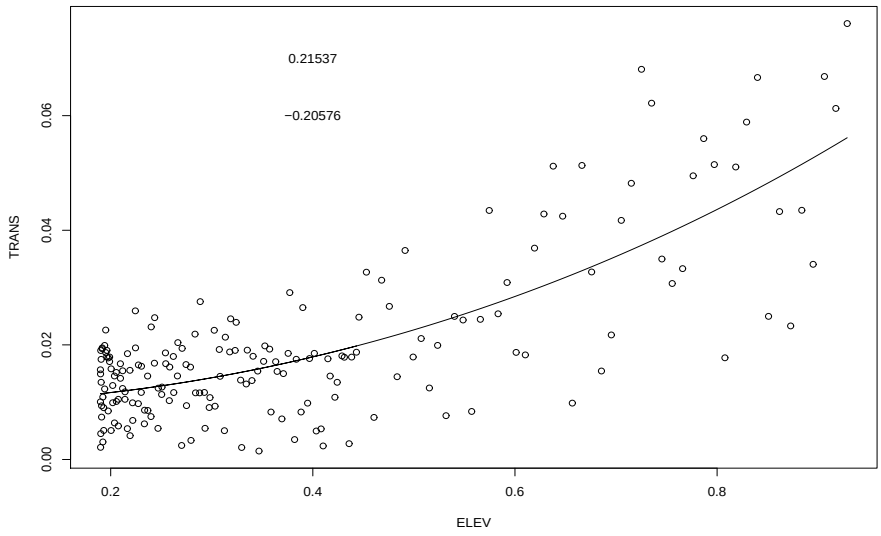

Abbildung 7.2: Transformierte Pseudo-Residuen der GPS-Zeitreihe aus Abb. 6.4

Dabei wählen wir $ELEV$ und T so, dass die nach der Funktion $ELEV$ angeordneten Daten $T(\frac{1}{2}R_{ni}^2)$ durch ein lineares Modell mit zumindest symmetrisch verteilten Fehlern beschrieben werden können. In diesem Beispiel

entscheiden wir uns für

$$ELEV(t_{ni}) := \frac{\pi}{8} \cdot \left((zd_{A_1}^{S_1}(t_{ni}))^2 + (zd_{A_1}^{S_2}(t_{ni}))^2 + (zd_{A_2}^{S_1}(t_{ni}))^2 + (zd_{A_2}^{S_2}(t_{ni}))^2 \right)$$

und $T(x) = (x)^{\frac{1}{4}},\ x > 0$. Als Regressionsfunktionen des linearen Modells wählen wir

$$f_1 \equiv 1 \text{ und } f_2(t_{ni}) = \cos^{\frac{1}{2}}(ELEV(t_{ni})).$$

Wir bestimmen nun für das lineare Modell

$$T\left(\frac{1}{2}R_{ni}^2\right) = \sqrt{\frac{|R_{ni}|}{\sqrt{2}}} \;=\; \theta_1 \;+\; \theta_2 \cdot \cos^{\frac{1}{2}}(ELEV(t_{ni})) \;+\; \delta_{ni}\ , \tag{7.11}$$

$$\mathrm{E}(\delta_{ni}) = 0,\ i = 2, 4, \ldots, 2[\frac{n}{2}]\ ,$$

die LS-Schätzer für die Parameter θ_1 bzw. θ_2 und erhalten $\hat{\theta}_1 = 0.21537$ und $\hat{\theta}_2 = -0.20576$. Abbildung 7.2 zeigt die transformierten Pseudoresiduen $T\left(\frac{1}{2}R_{ni}^2\right) = \left(\frac{1}{2}R_{ni}^2\right)^{\frac{1}{4}}$, $i = 2, 4, \ldots, 2[\frac{n}{2}]$, in Abhängigkeit von der Funktion $ELEV$. Der geschätzte Mean $\hat{M} := 0.21547 - 0.20576 \cdot \cos^{\frac{1}{2}}(ELEV)$ ist als Kurve eingezeichnet.

Wir wollen uns nun noch versichern, dass wir mit $T(x) = (x)^{\frac{1}{4}}$ eine geeignete Transformation und mit (7.11) ein brauchbares lineares Modell gewählt haben und betrachten zu diesem Zweck den Normal-QQ-Plot und die geschätzte Dichte der standardisierten Residuen des linearen Modells (7.11). Wie man in Abbildung 7.3 sehen kann, ist der Normal-QQ-Plot der standardisierten Residuen des Modells (7.11) nahezu linear, und die empirische Dichte stimmt relativ gut mit der (gestrichelt eingezeichneten) Dichte der Standard-Normalverteilung überein. Lediglich die augenscheinlich vorliegende Heteroskedastizität der quadrierten und transformierten Pseudo-Residuen (vgl. Abb. 7.2) ist eine weniger befriedigende Eigenschaft der transformierten Residuen $T(\frac{1}{2}R_{ni}^2)$ mit T wie in (7.7) und (7.8).

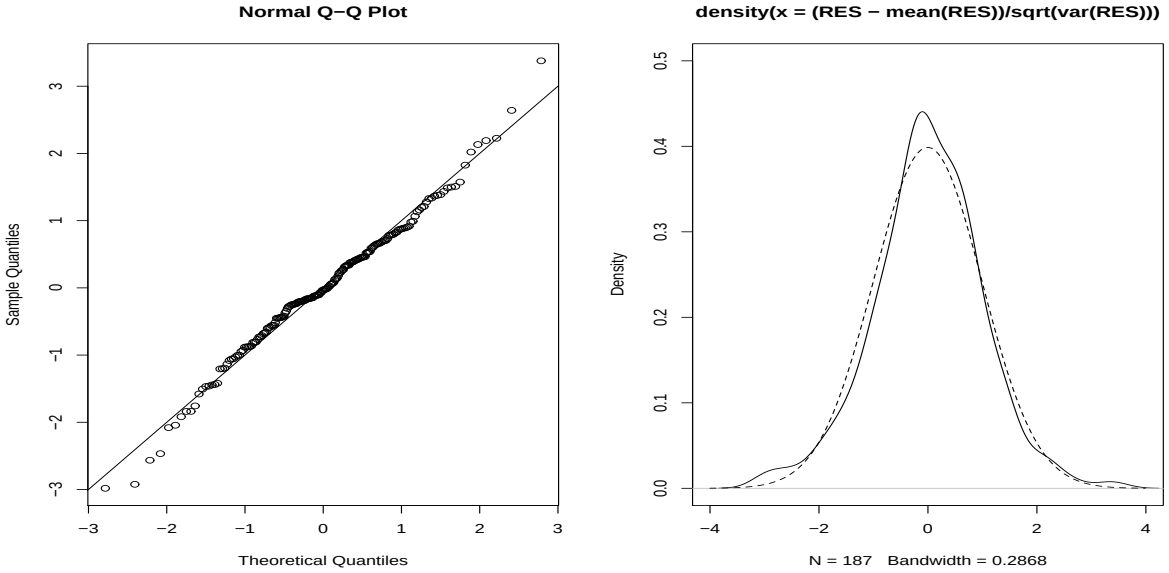

Abbildung 7.3: QQ-Plot und geschätzte Dichte der Residuen aus Modell (7.11)

Wendet man nun die Umkehrabbildung T^{-1} der Transformationsfunktion T auf den geschätzten Mean $\hat{M}$ des Modells (7.11) an, so erhält man einen Schätzer für die Varianzfunktion der Residuen-Zeitreihe durch

$$\widehat{\sigma^2}(t_{ni}) = T^{-1}(\hat{M}(t_{ni})) = (\hat{M}(t_{ni}))^4,\ i = 2, 4, \ldots, 2[\frac{n}{2}].$$

7.2.2 Gewichtung

Hat man wie in Abschnitt 7.2.1 die Varianzfunktion $\sigma^2(\,\cdot)$ geschätzt, gewichtet man die Beobachtungen in t_{ni} mit $w_{ni} := \left(\widehat{\sigma^2}(t_{ni})\right)^{-1/2},\ i = 1, \ldots, n$. Ist die Varianz korrekt geschätzt, zeigt der gewichtete Beobachtungsvektor $(w_{n1}Y_{n1}, \ldots, w_{nn}Y_{nn})^\top$ eine homogene Varianz. Nach erfolgter Gewichtung sollte daher erneut ein Test auf Homoskedastizität durchgeführt werden, um die Eignung der geschätzten Varianzfunktion zu überprüfen.

7.2.3 Bemerkung: Man beachte, dass sich durch eine Gewichtung des Beobachtungsvektors Y_n auch die Designmatrix des linearen Modells $Y_n = X_n\beta + \epsilon_n$ bzw. die Mean-Funktion $m(\,\cdot)$ der Zeitreihe $Y_t = m_t + Z_t,\ t = 1, \ldots, n$, verändert. Z.B. besitzt das gewichtete lineare Modell (7.3) die Form

$$\hat{\Sigma}_n^{-1/2} Y_n = \hat{\Sigma}_n^{-1/2} X_n \beta + \hat{\Sigma}_n^{-1/2} \epsilon_n \text{ mit } \hat{\Sigma}_n^{-1/2} = \operatorname{diag}(w_1, \ldots, w_n). \tag{7.12}$$

Der BLUE für das Modell (7.12) ist nun gegeben durch den gewichteten LS-Schätzer

$$\hat{\beta}_G = (X_n^\top \hat{\Sigma}_n^{-1} X_n)^{-1} X_n^\top \hat{\Sigma}_n^{-1} Y_n\ . \tag{7.13}$$

7.2.4 Beispiel: In Beispiel 7.2.2 ergab die Schätzung der Varianz der Residuen-Zeitreihe aus Abbildung 6.4 die Funktion $\widehat{\sigma^2}(t_{ni}) = (\hat{M}(t_{ni}))^4,\ i = 2, 4, \ldots, 2[\frac{n}{2}]$, mit $\hat{M}(t_{ni}) := 0.21547 - 0.20576 \cdot \cos^{\frac{1}{2}}(ELEV(t_{ni}))$. Daraus erhalten wir eine mögliche Gewichtsfunktion

$$w_{ni} := \begin{cases} 1/(\hat{M}(t_{ni}))^2, & i = 2, 4, \ldots, 2[\frac{n}{2}], \\ w_{n,i+1}, & i = 1, 3, \ldots, 2[\frac{n}{2}] - 1. \end{cases}$$

Abbildung 7.4 zeigt die mit $w_{ni},\ i = 1, \ldots, n$, gewichtete Zeitreihe aus Abb. 6.4, den Wert 0.43399 der Dette-Munk-Teststatistik und den entsprechenden p-Wert 0.6643. Im Gegensatz zur ungewichteten Zeitreihe aus Abb. 6.4 verwirft der Dette-Munk-Test im Falle der gewichteten Zeitreihe die Hypothese nicht mehr.

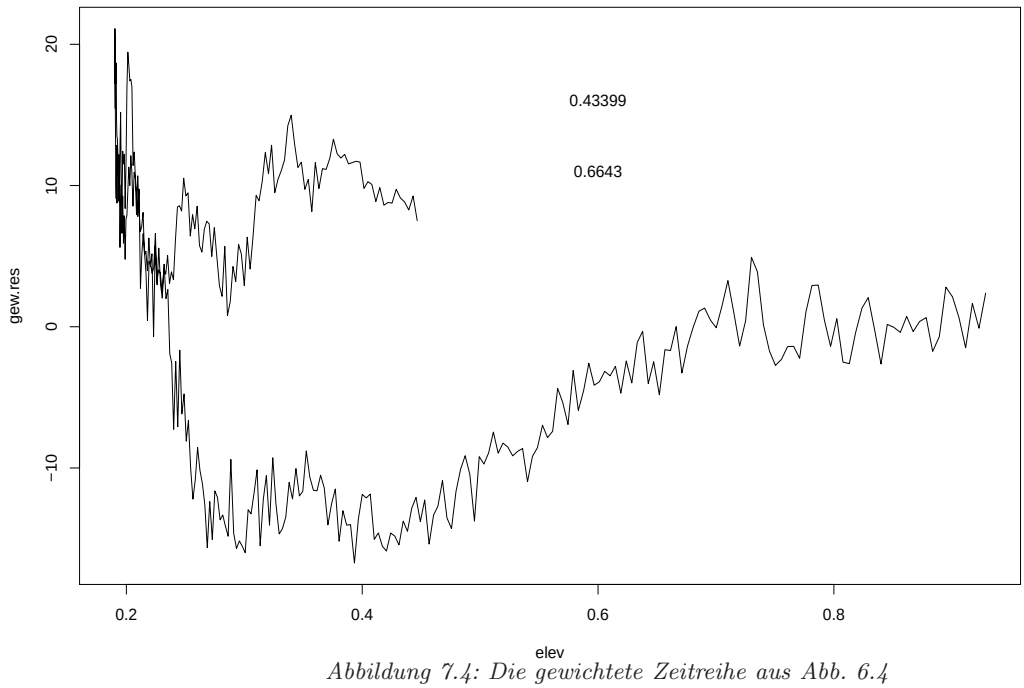

Abbildung 7.4: Die gewichtete Zeitreihe aus Abb. 6.4

7.2.3 Verallgemeinerte Kleinste Quadrate Schätzung

Um die Genauigkeit der Varianz- und Parameterschätzer zu erhöhen, kann man diese iterativ verfeinern, d.h. nach den folgenden Schritten vorgehen:

1.) Schätze die Varianzen $\sigma_1^2, \ldots, \sigma_n^2$ anhand der LS-Residuen $r = Y - X\hat{\beta}$, wobei $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ der herkömmliche LS-Schätzer ist.

2.) Gewichte das lineare Modell mit der Matrix $\hat{\Sigma}^{-1/2} = \mathrm{diag}(\hat{\sigma}_1^{-1}, \ldots, \hat{\sigma}_n^{-1})$, $\hat{\sigma}_i^{-1} := (\widehat{\sigma_i^2})^{-1/2}$, $i = 1, \ldots, n$, und berechne den gewichteten LS-Schätzer (7.13)

3.) Schätze die Varianzen $\sigma_1^2, \ldots, \sigma_n^2$ anhand der Residuen $r = Y - X\hat{\beta}_G$, wobei $\hat{\beta}_G$ der gewichtete LS-Schätzer aus dem zweiten Schritt ist.

4.) Wiederhole die Schritte 2.) und 3.) k-mal, wobei die Anzahl k der Iterationen vom Anwender festgelegt wird.

Dieses iterative Schätzverfahren findet man in der Literatur unter den Bezeichnungen *Generalized Least Squares* oder *Iteratively Reweighted Least Squares* (siehe z.B. Seber (1977) oder Carroll / Ruppert (1988).

7.2.5 Bemerkung: In Bischoff et al. (2006) wurden die Varianzfunktionen geodätischer GPS-Messreihen sowohl ohne Iteration (nur Schritt 1), als auch iterativ (Schritte 1 bis 4) geschätzt. Ferner wurden für die einfache Schätzung (keine Iteration) quadrierte LS-Residuen, für die iterative Schätzung hingegen rekursive Residuen verwendet. Anschließend wurden die gewichteten Zeitreihen jeweils auf Homogenität der Varianzfunktion getestet. Ein Vergleich zeigte in drei von 18 Fällen deutlich bessere Testergebnisse bei iterativer Schätzung. Die anderen Zeitreihen wiesen keine signifikanten Unterschiede in den Testergebnissen auf.

Kapitel 8. Ergänzung: Bispektralanalyse

8.1 Bispektralanalyse

8.1.1 Die Bispektraldichte

Als Fourier-Transformierte der Autokovarianzfunktion beinhaltet die Spektraldichte einer stationären Zeitreihe lediglich jene Informationen, welche in der Autokovarianzfunktion enthalten sind. Man spricht hierbei von Informationen zweiter Ordnung, da lediglich die Kovarianz zwischen jeweils zwei Zufallsvariablen X_s, X_t berücksichtigt wird. In manchen Fällen sind jedoch auch Informationen dritter Ordnung, d.h. jene, die die Kohärenz dreier Zufallsvariablen X_r, X_s, X_t berücksichtigen, von Bedeutung. Informationen dritter Ordnung sind in der *Bispektraldichte* enthalten.

8.1.1 Definition: Eine reellwertige Zeitreihe $(X_t),\ t \in \mathbb{Z}$, heißt *stationär bis zur Ordnung k* ($k \in \mathbb{N}$), falls

(*i*) $\mathrm{E}(X_t^k) < \infty$ für alle $t \in \mathbb{Z}$,

(*ii*) für alle $t_1, \ldots, t_n \in \mathbb{Z},\ n \in \mathbb{N}$, und für alle $k_1, \ldots, k_n \in \mathbb{N}_0$ mit $\sum_{j=1}^n k_j \leq k$ gilt:

$$\forall\, h \in \mathbb{Z}:\ \mathrm{E}\left(X_{t_1+h}^{k_1} \cdot\ \ldots\ \cdot X_{t_n+h}^{k_n}\right) = \mathrm{E}\left(X_{t_1}^{k_1} \cdot\ \ldots\ \cdot X_{t_n}^{k_n}\right).$$

Im Falle $k = 2$ entspricht dies der schwachen Stationarität aus Abschnitt 3.2.

8.1.2 Definition: Es sei (X_t) reellwertig und stationär bis zur Ordnung 3 mit $\mu := \mathrm{E}(X_t)$. Dann heißt

$$C(h_1, h_2) := \mathrm{E}(X_t - \mu)\mathrm{E}(X_{t+h_1} - \mu)\mathrm{E}(X_{t+h_2} - \mu)\ (h_1, h_2 \in \mathbb{Z})$$

die *Kumulante dritter Ordnung* für h_1 und h_2. Aufgrund der Stationarität bis zur Ordnung 3 ist $C(h_1, h_2)$ konstant in t. Ist die Kumulante dritter Ordnung absolut summierbar, d.h.

$$\sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} |C(h_1, h_2)| < \infty,$$

so wird für $\omega_1, \omega_2 \in [-\pi, \pi]$ durch

$$g(\omega_1, \omega_2) \quad := \quad \frac{1}{(2\pi)^2} \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} C(h_1, h_2) e^{-i\omega_1 h_1 - i\omega_2 h_2} \tag{8.1}$$

die *Bispektraldichte* von X_t definiert.

8.1.3 Bemerkung: Ist (X_t) ein reellwertiger Prozess und stationär bis zur Ordnung 3, so gilt

a) $C(h_1, h_2) = C(h_2, h_1) = C(-h_1, h_2 - h_1) = C(h_1 - h_2, -h_2)$.

b) Die Bispektraldichte ist eine in jedem Argument 2π-periodische, komplexwertige Funktion mit $\overline{g(-\omega_1, -\omega_2)} = g(\omega_1, \omega_2)$. Weiter folgt aus a) $g(\omega_1, \omega_2) = g(\omega_2, \omega_1) = g(-\omega_1 - \omega_2, \omega_2) = g(\omega_1, -\omega_1 - \omega_2)$.

c) Aufgrund von b) ist die Bispektraldichte durch ihre Werte im Dreieck $D := \{0 \leq \omega_1 \leq \pi,\ 0 \leq \omega_2 \leq \omega_1,\ 2\omega_1 + \omega_2 \leq 2\pi\}$ vollständig spezifiziert.

Beweis: Von den Aussagen in a) und b) soll lediglich die Gleichung $g(\omega_1, -\omega_1-\omega_2) = g(\omega_1, \omega_2)$ gezeigt werden. Der Rest ist leicht nachzuprüfen.

$$\begin{aligned} g(\omega_1, -\omega_1-\omega_2) &= \frac{1}{(2\pi)^2} \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} \underbrace{C(h_1,h_2)}_{=C(h_1-h_2,-h_2)} e^{-i\omega_1 h_1 + i\omega_1 h_2 + i\omega_2 h_2} \\ &= \frac{1}{(2\pi)^2} \sum_{h_1=-\infty}^{\infty} \sum_{h_2=-\infty}^{\infty} C(h_1-h_2,-h_2) e^{-i\omega_1(h_1-h_2)-i\omega_2(-h_2)} \\ &= \frac{1}{(2\pi)^2} \sum_{h_1-h_2=-\infty}^{\infty} \sum_{-h_2=-\infty}^{\infty} C(h_1-h_2,-h_2) e^{-i\omega_1(h_1-h_2)-i\omega_2(-h_2)} \\ &= g(\omega_1,\omega_2). \end{aligned}$$

c) Es seien die Werte der Bispektraldichte auf dem Dreieck D gegeben. Durch die Eigenschaft $g(\omega_1,\omega_2) = g(\omega_1, -\omega_1-\omega_2)$ ist dadurch auch die Bispektraldichte auf dem Dreieck $E := \{0 \le \omega_1 \le \pi,\ -2\omega_1 \le \omega_2 \le -\omega_1,\ \omega_2 \ge -2\pi+\omega_1\}$ spezifiziert. Durch $g(\omega_1,\omega_2) = g(-\omega_1-\omega_2,\omega_2)$ ist weiter die Bispektraldichte auf $F := \{-\pi-\omega_1 \le \omega_2 \le -\omega_1,\ 0 \le \omega_2 \le \omega_1/2,\ \omega_2 \ge -2\pi-2\omega_1\}$ bestimmt. Alle anderen Punkte des Intervalls $[-\pi,\pi]\times[-\pi,\pi]$ erhält man durch Spiegelung der Dreiecke D, E und F an der Geraden $\{\omega_1 = \omega_2\}$ (dies entspricht der Abbildung $S_1 : \mathbb{R}^2 \to \mathbb{R}^2,\ S_1(\omega_1,\omega_2) = (\omega_2,\omega_1)$), Verschiebung (entspricht der 2π-Periodizität der Bispektraldichte), sowie der Punktspiegelung der Dreiecke D, E und F am Nullpunkt (entspricht der Abbildung $S_2 : \mathbb{R}^2 \to \mathbb{R}^2,\ S_2(\omega_1,\omega_2) = (-\omega_1,-\omega_2)$). □

Ist (X_t) ein stationärer linearer Prozess

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j},\ (\epsilon_t) \sim \mathit{IID}(0,\sigma^2), \tag{8.2}$$

dann heißt

$$H(\omega) := \sum_{j=-\infty}^{\infty} \psi_j e^{-i\omega j}$$

die *Transfer-Funktion* von (X_t).

8.1.4 Lemma: *Für den stationären Prozess (8.2) gilt*

a) $\gamma(h) = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}\right)$ *(Bemerkung 3.2.8)*

b) *Für die Spektraldichte gilt*
$g(\omega) = \frac{\sigma^2}{2\pi} |H(\omega)|^2 = \frac{\sigma^2}{2\pi} H(\omega)\overline{H(\omega)} = \frac{\sigma^2}{2\pi} H(\omega)H(-\omega)$.

Ist in (8.2) (X_t) stationär bis zur Ordnung 3 und $\mu_3 := E(\epsilon^3)$, so gilt außerdem

c) $C(h_1,h_2) = \mu_3 \left(\sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h_1} \psi_{j+h_2}\right)$

d) *Für die Bispektraldichte gilt* $g(\omega_1,\omega_2) = \frac{\mu_3}{(2\pi)^2} H(\omega_1)H(\omega_2)H(-\omega_1-\omega_2)$.

Beweis: Subba Rao / Gabr (1984).

8.1.5 Definition: Es sei X eine Zufallsvariable mit existierendem 3. Moment $\mathrm{E}(X^3)$. Es bezeichne μ_X den Erwartungswert und σ_X die Standardabweichung von X. Dann heißt

$$s_3 := \frac{\mathrm{E}[(X-\mu_X)^3]}{\sigma_X^3}$$

der *Skewness-Parameter* von X.

Der Skewness-Parameter misst gewissermaßen die Nicht-Symmetrie einer Verteilung. Ist X symmetrisch verteilt, so gilt $s_3 = 0$. Die Normalverteilung z.B. besitzt als symmetrische Verteilung den Skewness-Parameter 0. Ist im Modell (8.2) der Prozess (X_t) stationär bis zur Ordnung 3 und (ϵ_t) gaußverteilt, so ist demnach $\mu_3 = 0$ und nach Lemma 8.1.4 c) und d) ist sowohl die Kumulante dritter Ordnung als auch die Bispektraldichte konstant 0.

8.1.6 Definition: Es seien $g(\cdot)$ und $g(\cdot\,,\cdot)$ die (nichtnormierte) Spektral- bzw. Bispektraldichte eines bis zur Ordnung 3 stationären Prozesses (X_t). Dann wird durch

$$f(\omega_1, \omega_2) := \frac{g(\omega_1, \omega_2)}{\sqrt{g(\omega_1)g(\omega_2)g(\omega_1 + \omega_2)}}$$

die *normierte Bispektraldichte* definiert.

Ist der lineare Prozess (8.2) stationär bis zur Ordnung 3 und $\mu_3 := \mathrm{E}(\epsilon^3)$, so gilt nach Lemma 8.1.4 b) und d)

$$|f(\omega_1, \omega_2)|^2 = \frac{|g(\omega_1, \omega_2)|^2}{g(\omega_1)g(\omega_2)g(\omega_1 + \omega_2)} = \frac{\mu_3^2}{2\pi\sigma^6},\ (\omega_1, \omega_2) \in D. \tag{8.3}$$

$|f(\omega_1, \omega_2)|^2$ ist in diesem Falle also konstant für alle $(\omega_1, \omega_2) \in D$. Insbesondere ist $|f(\omega_1, \omega_2)|^2 \equiv 0$, falls (X_t) symmetrisch verteilte Zuwächse (ϵ_t) besitzt. Somit ist es möglich, anhand eines Schätzers für die Bispektraldichte Tests zu konstruieren, die die Hypothese $H_0 : \mu_3 = 0$ testen. Muss diese Hypothese verworfen werden, so kann man davon ausgehen, dass $\epsilon_t,\ t \in T$, nicht symmetrisch verteilt ist, insbesondere, dass (ϵ_t) kein Gauß-Prozess ist.

8.1.2 Ein Schätzer für die Bispektraldichte

Soll die Bispektraldichte eines bis zur Ordnung 3 stationären Prozesses geschätzt werden, so geht man im Prinzip wie bei der Schätzung der Spektraldichte vor. Man berechnet zunächst das *Periodogramm dritter Ordnung*, ein Pendant zum Periodogramm und glättet dieses anschließend mittels eines Fensters.

Man betrachte die bis zur Ordnung 3 stationäre Zeitreihe $(X_t),\ t \in \mathbb{Z}$, zu den Zeitpunkten $t = 1, \ldots, N$. Ein natürlicher Schätzer für $C(h_1, h_2)$ ist

$$\widehat{C}(h_1, h_2) := \frac{1}{N} \sum_{t=1}^{N-h} (X_t - \overline{X})(X_{t+h_1} - \overline{X})(X_{t+h_2} - \overline{X}),\ h_1, h_2 \geq 0,\ h = \max(0, h_1, h_2).$$

Nun definiere man für $\omega_1, \omega_2 \in [-\pi, \pi]$

$$I(\omega_1, \omega_2) \quad := \quad \frac{1}{(2\pi)^2} \sum_{h_1=-(N-1)}^{N-1} \sum_{h_2=-(N-1)}^{N-1} \widehat{C}(h_1, h_2) e^{-i\omega_1 h_1 - i\omega_2 h_2}\ .$$

Man nennt $I(\cdot\,,\cdot)$ das *Periodogramm dritter Ordnung* von (X_t). Das Periodogramm dritter Ordnung ist ein asymptotisch erwartungstreuer Schätzer für die Bispektraldichte, jedoch wie das Periodogramm zweiter Ordnung i. A. nicht konsistent (s. Brillinger / Rosenblatt (1967)). Also muss auch $I(\cdot\,,\cdot)$ durch ein geeignetes Window geglättet werden. In Subba Rao / Gabr (1984) werden verschiedene Bispektral-Windows eingeführt, eines davon soll an dieser Stelle vorgestellt werden.

Es sei G die Ellipse $\{(\theta_1, \theta_2) :\ \theta_1^2 + \theta_2^2 + \theta_1\theta_2 \leq \pi^2\}$. Die Funktion

$$W(\theta_1, \theta_2) := \begin{cases} \frac{\sqrt{3}}{\pi^3}\left(1 - \frac{1}{\pi^2}(\theta_1^2 + \theta_2^2 + \theta_1\theta_2)\right), & \text{falls } (\theta_1, \theta_2) \in G, \\ 0\,, & \text{sonst,} \end{cases}$$

heißt *Optimum Bispektral Window*.

Man wähle nun eine Folge (M_N) so, dass $\frac{M_N^2}{N} \to 0$ für $M_N, N \to \infty$. (Näheres zur Wahl von M_N bei gegebenem $N \in \mathbb{N}$ siehe Subba Rao / Gabr (1984)). Zu gegebenem $N \in \mathbb{N}$ ist dann ein konsistenter Schätzer für die Bispektraldichte gegeben durch (vgl. Subba Rao / Gabr (1984))

$$\hat{g}(\omega_1, \omega_2) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} M_N^2 W(M_N(\theta_1 - \omega_1), M_N(\theta_2 - \omega_2)) I(\theta_1, \theta_2) d\theta_1 d\theta_2.$$

Somit ist

$$\hat{f}(\omega_1, \omega_2) := \frac{\hat{g}(\omega_1, \omega_2)}{\sqrt{\hat{g}(\omega_1)\hat{g}(\omega_2)\hat{g}(\omega_1 + \omega_2)}}$$

ein Schätzer für die normierte Bispektraldichte.

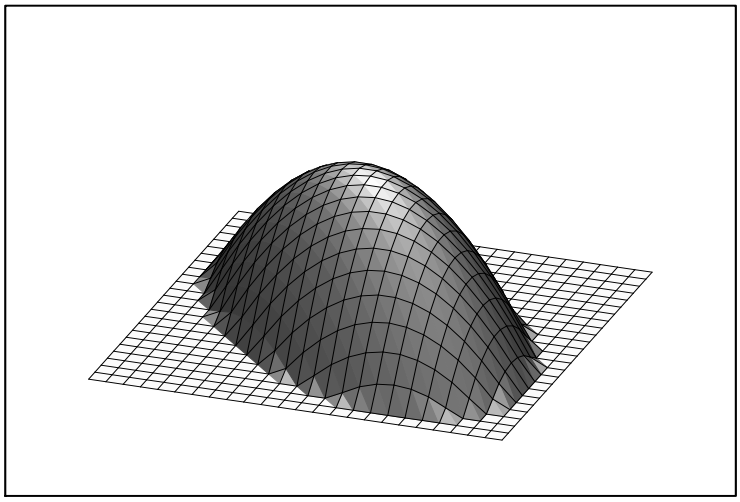

Abbildung 8.1: Das Optimum Bispektral-Window

8.1.3 Ein Test für $\mu_3 = 0$ und Linearität

Wie bereits erwähnt, gilt für einen linearen Prozess mit drittem Moment $\mu_3 = 0$, dass die normierte Bispektraldichte $f(\omega_1, \omega_2)$ konstant ist für alle $i, j \in \mathbb{N}$. Insbesondere trifft dies für einen Prozess mit gaußverteilten Zuwächsen zu. Abschnitt 8.1.2 erlaubt eine Beurteilung von $|f(\omega_1, \omega_2)|^2$ anhand graphischer Kriterien. In der Regel ist es jedoch wünschenswert, die Hypothese

$$H_0 : \ (X_t) \textit{ ist linear und stationär bis zur Ordnung } 3 \textit{ mit } \mu_3 = 0$$

auch zu testen. Subba Rao und Gabr (1984) testen H_0 in zwei Schritten. Hier soll ein einfacher Test von Hinich (1982) vorgestellt werden, der auf der χ^2-Verteilung basiert.

Zunächst wird der Bereich D aus Bemerkung 8.1.3 mit einem Punktegitter L überzogen. Hierfür sei $M_N = N^\eta$ für ein $\eta \in (\frac{1}{2}, 1)$. Hinich (1982) merkt an, dass zu Testzwecken η nur geringfügig größer als $\frac{1}{2}$ gewählt werden sollte. Man setze $\omega_m = \frac{M_N}{N}(2m-1)\pi$, $\omega_n = \frac{M_M}{N}(2n-1)\pi$ und

$$L = \left\{ (\omega_m, \omega_n) : n = 1, \ldots, m, \ 1 \leq 2m + n \leq \frac{N}{M_N} + \frac{3}{2} \right\}.$$

Anschließend wähle man ein feineres Gitter $\tilde{L}$ so, dass jeder Gitterpunkt (ω_m, ω_n) von L der Mittelpunkt eines Quadrates $Q_{m,n}$ ist, das M_N^2 Gitterpunkte des Gitters $\tilde{L}$ enthält. Diese Gitterpunkte müssen nicht notwendigerweise in D enthalten sein. Für jeden Punkt (ω_i, ω_j) des Gitters $\tilde{L}$ wird nun $I(\omega_i, \omega_j)$ berechnet. Anschließend

wird das Periodogramm dritter Ordnung auf einfache Weise geglättet, indem für jeden Gitterpunkt (ω_m, ω_n) von L

$$\hat{g}\left(\frac{M_N}{N}(2m-1)\pi, \frac{M_N}{N}(2n-1)\pi\right) := \frac{1}{M_N^2} \sum_{(\omega_i,\omega_j)\in Q_{m,n}} I(\omega_i, \omega_j)$$

als Schätzer für die Bispektraldichte verwendet wird. Besitzt ein Quadrat $Q_{m,n}$ Punkte (ω_i, ω_j) außerhalb des Bereiches D, so werden diese Punkte bei der Mittelbildung nicht berücksichtigt. In diesem Fall muss der Faktor $\frac{1}{M_N^2}$ durch einen geeigneten Vorfaktor ersetzt werden (ersetze M_N^2 durch die Anzahl der Summanden). Nach Hinich (1982) ist $\hat{g}(\cdot\,,\cdot)$ ein konsistenter Schätzer für $g(\cdot\,,\cdot)$.

Entsprechend schätzt man $g(\cdot\ )$ durch Mittelung von M_N benachbarten Periodogramm-Ordinaten (Periodogramm zweiter Ordnung, s. Abschnitt 4.2.1) und erhält für jeden Gitterpunkt (ω_m, ω_n) von L den Schätzer

$$\left|\hat{f}\left(\omega_m, \omega_n\right)\right|^2 = \frac{\left|\hat{g}\left(\omega_m, \omega_n\right)\right|^2}{\hat{g}(\omega_m)\hat{g}(\omega_n)\hat{g}\left(\omega_m + \omega_n\right)}.$$

Es bezeichne nun $q_{m,n}$ die Anzahl der Punkte in $D \cap Q_{m,n}$, wobei die Punkte auf dem Rand von $D \cap Q_{m,n}$ doppelt gezählt werden. Setze

$$\widehat{Y}_{m,n} := \frac{2\pi}{N^{1-4\eta} \cdot q_{m,n}} \cdot \left|\hat{f}\left(\frac{M_N}{N}(2m-1)\pi, \frac{M_N}{N}(2n-1)\pi\right)\right|^2.$$

8.1.7 Lemma: *Es sei (X_t) ein linearer Prozess gemäß (8.2), stationär bis zur Ordnung 3. Dann ist unter der Hypothese H_0 die Teststatistik $T := 2\sum_{(m,n)\in L} \widehat{Y}_{m,n}$ approximativ χ^2-verteilt mit $2A$ Freiheitsgraden, wobei A die Anzahl der Gitterpunkte (ω_m, ω_n) in L ist.*

Beweis: Hinich (1982).

Das Testverfahren ist nun klar. Man verwerfe H_0, falls $T > \chi^2_{2A,1-\alpha}$, wobei $\chi^2_{2A,1-\alpha}$ das $(1-\alpha)-$Quantil der χ^2-Verteilung mit $2A$ Freiheitsgraden ist.

Muss der Test verworfen werden, so kommen als Ursachen in Frage

(i) Die Zuwächse (ϵ_t) sind nicht symmetrisch verteilt,

(ii) Die Zuwächse (ϵ_t) sind nicht *iid*-verteilt,

(iii) Der Prozess (X_t) ist nicht stationär bis zur Ordnung 3,

(iv) Es liegt kein linearer Prozess vor.

Zur Klärung der Frage, ob überhaupt ein linearer Prozess vorliegt, hat Hinich (1982) einen Test entwickelt, der auf dem Quartilsabstand der nicht-zentralen χ^2-Verteilung (siehe z.B. Johnson et al. (1995)) basiert. Subba Rao und Gabr (1984) haben einen entsprechenden F-Test entwickelt.

Verwirft ein solcher Test die Hypothese eines linearen Prozesses, stellt sich die Frage, wie im Falle (iv) ein nichtlinearer Prozess modelliert werden kann. Mögliche Modellansätze, z.B. *Bilineare Prozesse*, *Threshold-Autoregressive Prozesse*, *Exponentiell-Autoregressive Prozesse*, werden z.B. in Priestley (1980) eingeführt.

Anhang A. Begriffe aus der Funktionalanalysis

A.1 Grundlegendes aus der Funktionalanalysis

A.1.1 Definition: Es sei $\mathbb{K}$ ein Körper ($\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$) und E eine nichtleere Menge. Ist für alle $x, y \in E$ und $\alpha \in \mathbb{K}$ eine Summe $x + y \in E$ und ein Produkt $\alpha x \in E$ definiert, so dass

(i) $x + (y + z) = (x + y) + z$ (Assoziativität),

(ii) $x + y = y + x$ (Kommutativität),

(iii) es existiert ein Nullelement in E,

(iv) für jedes $x \in E$ existiert ein inverses Element $-x$,

(v) $\alpha(x + y) = \alpha x + \alpha y$,

(vi) $(\alpha + \beta)x = \alpha x + \beta x$,

(vii) $(\alpha\beta)x = \alpha(\beta x)$,

($viii$) $1 \cdot x = x$,

so heißt E ein *Vektorraum* über $\mathbb{K}$.

A.1.2 Beispiel: Der $\mathbb{R}^n := \{(x_1, \ldots, x_n)^\top : x_i \in \mathbb{R}, i = 1, \ldots, n\}$ ist ein Vektorraum über $\mathbb{R}$.

A.1.3 Beispiel: Der Raum $C[0, 1]$ aller reellwertigen, auf dem Intervall $[0, 1]$ stetigen Funktionen ist ein Vektorraum über $\mathbb{R}$.

A.1.4 Definition: Eine nichtleere Teilmenge U eines Vektorraumes E heißt *(linearer) Unterraum* von E, wenn für alle $x, y \in U$ und alle $\alpha \in \mathbb{K}$ gilt:

(i) $x + y \in U$,

(ii) $\alpha x \in U$.

A.1.5 Bemerkung: Für einen Vektorraum E und einen beliebigen Unterraum U von E gilt (vgl. Heuser (1992)):

(i) U enthält das Nullelement von E und ist selbst ein Vektorraum.

(ii) Der Schnitt beliebig vieler Unterräume von E ist wieder ein Unterraum von E.

(iii) $\{0\}$ ist ein Unterraum von E.

A.1.6 Bemerkung und Definition: Es sei M eine nichtleere Teilmenge des Vektorraumes E. Nach Bemerkung A.1.5 (ii) ist der Schnitt aller Unterräume, die M enthalten, ein Unterraum von E. Man nennt diesen Unterraum den *von M erzeugten* (oder *aufgespannten*) *Unterraum* oder die *lineare Hülle von M* und schreibt $\langle M \rangle$ oder $\mathrm{sp}(M)$.

A.1.7 Bemerkung: Die lineare Hülle einer Menge M entspricht der Menge aller (endlichen) Linearkombinationen von Elementen aus M.

A.1.8 Definition: Sei E ein Vektorraum über einem Körper $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$). Ist für jedes Paar (x, y) mit $x, y \in E$ eine Zahl $\langle x, y\rangle \in \mathbb{K}$ so definiert, dass

(i) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$,

(ii) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle,\ \alpha \in \mathbb{K}$,

(iii) $\langle y, x\rangle = \overline{\langle x, y\rangle}$, ($\overline{\langle x, y\rangle}$ die zu $\langle x, y\rangle$ konjugiert komplexe Zahl),

(iv) $\langle x, x\rangle \geq 0\ \forall\ x \in E$ und $\langle x, x\rangle = 0 \Leftrightarrow x = 0$,

so heißt $\langle x, y\rangle$ *Innenprodukt.* Der Raum E, versehen mit einem Innenprodukt, heißt *Innenproduktraum.*

A.1.9 Beispiel: Auf dem Vektorraum $\mathbb{R}^n$ ist durch $\langle x, y\rangle := \sum_{i=1}^n x_i y_i = x^\top y,\ x = (x_1, \ldots, x_n)^\top, y = (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$, ein Innenprodukt definiert.

A.1.10 Definition: Es sei E ein Vektorraum über einem Körper $\mathbb{K}$. Ist jedem Element $x \in E$ eine Zahl $\|x\|$ so zugeordnet, dass gilt

(i) $\|x\| \geq 0\ \forall\ x \in E$ und $\|x\| = 0 \Leftrightarrow x = 0$,

(ii) $\|\alpha x\| = |\alpha|\|x\|,\ \alpha \in \mathbb{K}$,

(iii) $\|x + y\| \leq \|x\| + \|y\|$,

so heißt $\|x\|$ die *Norm* von x. Der Raum E, versehen mit einer Norm, heißt *normierter Raum.*

Ist nun auf einem Vektorraum E über $\mathbb{K}$ ein Innenprodukt $\langle\cdot,\cdot\rangle$ gegeben, dann wird durch $\|x\| := \sqrt{\langle x, x\rangle}$ eine Norm auf E definiert (vgl. z. B. Heuser (1992)), die *durch* $\langle\cdot,\cdot\rangle$ *induzierte Norm.*

A.1.11 Beispiel: Die euklidische Norm $\|x\| = \sqrt{\sum_{i=1}^n x_i^2},\ x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, ist die durch das Innenprodukt aus Beispiel A.1.9 induzierte Norm des $\mathbb{R}^n$.

A.1.12 Satz: (Schwarz'sche Ungleichung) *Ist E ein Innenproduktraum, so gilt für $x, y \in E$ stets*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|.$$

Das Gleichheitszeichen gilt genau dann, wenn x und y linear abhängig sind.

Beweis: s. Heuser (1992).

Ist (x_n) eine Folge in E derart, dass zu jedem $\epsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, so dass für alle $n, m \geq n_0$ gilt $\|x_n - x_m\| < \epsilon$, so spricht man von einer *Cauchyfolge.*

A.1.13 Definition: Ein normierter Raum E heißt *vollständig,* falls jede Cauchyfolge (x_n) aus E gegen ein Element aus E *konvergiert*, falls also für jede Cauchyfolge $(x_n) \subset E$ ein $x \in E$ existiert mit

$$\|x_n - x\| \to 0\ (n \to \infty). \tag{A.1}$$

Ein vollständiger normierter Raum wird auch *Banachraum* genannt.

A.1.14 Beispiel: Der $\mathbb{R}^n$, versehen mit der euklidischen Norm, ist ein Banachraum.

A.1.15 Beispiel: Der Raum $C[0, 1]$ aller reellwertigen, auf dem Intervall $[0, 1]$ stetigen Funktionen, versehen mit der Maximumsnorm

$$\|x\| = \max_{t \in [0,1]} |x(t)|\ , \tag{A.2}$$

ist ein Banachraum.

A.1.16 Definition: Eine Teilmenge M eines normierten Raumes E heißt *abgeschlossen*, wenn der Grenzwert jeder konvergenten Folge $(x_n) \subset M$ ein Element aus M ist.

A.1.17 Definition: Es sei M eine Teilmenge des normierten Raumes E. Der Schnitt aller abgeschlossenen Teilmengen, die M enthalten, heißt *Abschluss* von M. Man schreibt $\overline{M}$. Die Menge $\overline{M} \cap \overline{E \backslash M}$ heißt der *Rand* von M.

A.1.18 Bemerkung: Der Abschluss $\overline{M}$ ist die Menge alle Grenzwerte konvergenter Folgen aus M. Es gilt

$$M = \overline{M} \iff M \text{ abgeschlossen.}$$

Ist $\overline{M} = E$, so sagt man, M liege *dicht* in E.

A.1.19 Bemerkung: Es sei U ein Untervektorraum des **vollständigen** normierten Raumes E. Dann ist U selbst ein normierter Raum und als solcher genau dann vollständig, wenn er als Unterraum von E abgeschlossen ist. (Für einen **un**vollständigen normierten Raum E folgt aus der Abgeschlossenheit des Unterraumes U **nicht** die Vollständigkeit von U.)

A.1.20 Definition: Einen Innenproduktraum, der bezüglich seiner durch das Innenprodukt $\langle \cdot, \cdot \rangle$ induzierten Norm vollständig ist, nennt man *Hilbertraum.*

A.1.21 Beispiel: Der Vektorraum $\mathbb{R}^n$, versehen mit dem Innenprodukt $\langle x, y \rangle := \sum_{i=1}^n x_i y_i$ ist ein Hilbertraum.

A.1.22 Definition: Sind $x_1, x_2, \ldots$ Elemente des normierten Raumes E, so sagt man, die Reihe $\sum_{i=1}^m x_i$ *konvergiere* gegen ein Element aus E, wenn ein $x \in E$ existiert mit

$$\|x - \sum_{i=1}^{m} x_i\| \overset{m \to \infty}{\longrightarrow} 0.$$

In diesem Falle schreibt man $x = \sum_{i=1}^{\infty} x_i$.

A.1.23 Definition: In einem Innenproduktraum E heißen zwei Elemente x, y zueinander *orthogonal* (i.Z. $x \perp y$), wenn ihr Innenprodukt verschwindet, also falls $\langle x, y \rangle = 0$.

Eine Folge $\{u_1, u_2, \ldots\}$ von Elementen $u_i \in E$ heißt *Orthogonalfolge*, falls die u_j paarweise orthogonal sind, falls also gilt $\langle u_i, u_j \rangle = 0,\ i \neq j$. Eine Orthogonalfolge $\{u_1, u_2, \ldots\}$ heißt *Orthogonalbasis* für E, falls die u_i, u_j zueinander orthogonal sind $(i \neq j)$ und jedes Element $x \in E$ darstellbar ist in der Form

$$x = \sum_{i=1}^{\infty} a_i u_i,\ a_i \in \mathbb{K},\ i = 1, 2, \ldots \tag{A.3}$$

Allgemein nennt man eine Menge zueinander orthogonaler Elemente *Orthogonalsystem.*

Eine Menge $\{u_1, u_2, \ldots\}$ von Elementen eines normierten Raumes E heißt *Orthonormalfolge* , falls die u_i paarweise orthogonal sind und für jedes Element gilt $\|u_i\| = 1$. Ist $\{u_1, u_2, \ldots\}$ zusätzlich eine Basis von E, ist also jedes $x \in E$ darstellbar in der Form (A.3), so heißt $\{u_1, u_2, \ldots\}$ *Orthonormalbasis.*

A.1.24 Beispiel: Im $\mathbb{R}^n$ bilden die Einheitsvektoren $e_i,\ i = 1, \ldots n$, die an der i-ten Stelle eine 1 und sonst Nullen besitzen, eine Orthonormalbasis.

A.1.25 Satz: *Es sei $\{u_1, u_2, \ldots\}$ eine Orthonormalfolge des Hilbertraumes E und $a_i \in \mathbb{K},\ i \in \mathbb{N}$. Dann konvergiert $\sum_{i=1}^{\infty} a_i u_i$ gegen ein Element aus E genau dann, wenn $\sum_{i=1}^{\infty} |a_i|^2 < \infty$. Ist x ein Element aus dem Hilbertraum E mit*

$$x = \sum_{i=1}^{\infty} a_i u_i,$$

so sind die Koeffizienten a_i eindeutig bestimmt durch $a_i = \langle x, u_i \rangle,\ i \in \mathbb{N}$.

Den Beweis findet der Leser z.B. in Heuser (1992) oder Bachman et al. (2000).

A.2 Ergänzungen zur Fourier-Theorie

In Anhang A.2 werden einige Begriffe eingeführt, die in der Literatur zur Fourier-Theorie häufig verwendet werden.

A.2.1 Faltungen

Unter der *Faltung* (engl. *convolution*) zweier reellwertiger Funktionen φ und ψ versteht man das Integral

$$(\varphi * \psi)(t) := \int_{-\infty}^{\infty} \varphi(s)\psi(t-s)ds.$$

A.2.1 Lemma: *Bezeichnet f die Fourier-Transformierte der quadratisch integrierbaren Funktion φ und g die Fourier-Transformierte von $\psi \in L^2$, so ist die Fourier-Transformierte h der Faltung $\varphi * \psi$ gegeben durch*

$$h(\omega) = f(\omega) \cdot g(\omega) \cdot 2\pi.$$

Beweis: Es ist

$$\begin{aligned} h(\omega) &\overset{(1.18)}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\varphi(s)\psi(t-s)ds\right)e^{-i\omega t}dt \\ &\overset{v=t-s}{=} \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi(s)\psi(v)e^{-i\omega(s+v)}dsdv \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\varphi(s)e^{-i\omega s}ds\int_{-\infty}^{\infty}\psi(v)e^{-i\omega v}dv \\ &= f(\omega)\cdot g(\omega)\cdot 2\pi. \end{aligned} \tag{A.4}$$

□

A.2.2 Lemma: *Die Fourier-Transformierte k des Produktes $\varphi(t)\cdot\psi(t)$ der quadratisch integrierbaren Funktionen φ und ψ ist die Faltung der Fourier-Transformierten von φ und ψ, also*

$$k(\omega) = (f * g)(\omega).$$

Beweis:

$$\begin{aligned} \int_{-\infty}^{\infty}(f*g)(\omega)e^{i\omega t}d\omega &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}f(\lambda)\cdot g(\omega-\lambda)d\lambda\right)e^{i\omega t}d\omega \\ &\overset{\nu=\omega-\lambda}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\lambda)g(\nu)e^{i(\lambda+\nu)t}d\lambda d\nu \\ &= \int_{-\infty}^{\infty}f(\lambda)e^{i\lambda t}d\lambda\cdot\int_{-\infty}^{\infty}g(\nu)e^{i\nu t}d\nu \\ &= \varphi(t)\cdot\psi(t). \end{aligned}$$

$\varphi\cdot\psi$ ist also die inverse Fourier-Transformierte von k und somit, wegen der Eindeutigkeit in L^2, k die Fourier-Transformierte von $\varphi\cdot\psi$. □

A.2.2 Distributionen und Dirac-Impuls

Motivation: Ein physikalischer Impuls, z.B. ein Lichtblitz, benötigt in der Realität eine bestimmte Zeitdauer, innerhalb derer er sich entwickeln und auch wieder abklingen kann. Abbildung A.1 zeigt eine Funktion, die z.B. als Energie eines Lichtblitzes im Laufe der Zeit interpretiert werden kann. Im Allgemeinen ist sowohl die Dauer als auch die genaue Form des Impulses nicht bekannt und auch nicht von Interesse. Von Interesse ist vielmehr der **Effekt**, den dieser Impuls z.B. bei seiner Messung ergibt.

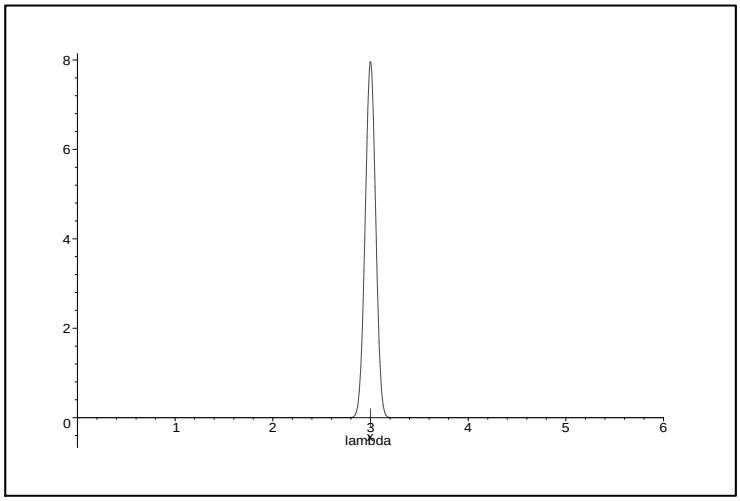

Abbildung A.1: Physikalischer Impuls

In Abbildung A.1 erreicht die Energie des Lichtblitzes zur Zeit $\lambda \in \mathbb{R}$ ihr Maximum. Es soll die gesamte Energie (also die Energie über den gesamten Zeitraum) gemessen werden und wir nehmen an, die Messung ergebe einen Wert $\phi(\lambda) \in \mathbb{R}$. Diesen Messvorgang könnte man theoretisch durch Integration der Funktion aus Abbildung A.1 beschreiben, doch die genaue Form dieser Funktion ist unbekannt. Zur Behebung dieses Problems behilft man sich mit Distributionen (verallgemeinerten Funktionen).

A.2.3 Definition: Eine auf dem $\mathbb{R}^n$ definierte Funktion ϕ heißt *finit*, wenn sie außerhalb einer beschränkten Menge $M \subset \mathbb{R}^n$ verschwindet. Die Menge aller im $\mathbb{R}^n$ finiten und beliebig oft stetig differenzierbaren Funktionen $\phi : \mathbb{R}^n \to \mathbb{C}$ bezeichnet man mit $\mathcal{D}$. Die Elemente von $\mathcal{D}$ nennt man *Grund-* oder *Testfunktionen.*

A.2.4 Definition: Man sagt, die Folge $(\phi_k) \subset \mathcal{D}$ *konvergiert in* $\mathcal{D}$ *gegen 0*, falls

(i) eine beschränkte Menge M existiert, so dass alle ϕ_k außerhalb M verschwinden

(ii) $\sup_{x\in\mathbb{R}^n} |\phi_k(x)| \overset{k\to\infty}{\longrightarrow} 0$

(iii) $\sup_{x\in\mathbb{R}^n} |(\frac{\partial}{\partial x_1})^{p_1} \ldots (\frac{\partial}{\partial x_n})^{p_n} \phi_k(x)| \overset{k\to\infty}{\longrightarrow} 0$ für alle $p_1, \ldots, p_n \in \mathbb{N}$.

Eine Folge (ϕ_k) konvergiert in $\mathcal{D}$ gegen ϕ, falls $(\phi_k - \phi) \to 0$. Die Schreibweise $\phi = \sum_{i=1}^{\infty} \phi_i$ bedeutet, dass die Folge der Teilsummen $\sum_{i=1}^{k} \phi_i$ gegen ϕ konvergiert.

Die Definition der Konvergenz in $\mathcal{D}$ fordert also neben der gleichmäßigen Konvergenz auf $\mathbb{R}^n$ (Bedingung (ii)) auch, dass die Folgen der partiellen Ableitungen (beliebiger Ordnung) gleichmäßig auf $\mathbb{R}^n$ konvergieren (iii). Ist außerdem M_k die definitionsgemäß zu ϕ_k existierende Menge, außerhalb derer ϕ_k verschwindet, so bedeutet (i), dass die Vereinigung aller Mengen M_k beschränkt sein muss.

A.2.5 Definition: Eine Abbildung $T : \mathcal{D} \to \mathbb{C}$ heißt *Funktional auf* $\mathcal{D}$. T heißt *linear*, falls gilt

$$T(\lambda\phi + \mu\psi) = \lambda T(\phi) + \mu T(\psi) \text{ für alle } \phi, \psi \in \mathcal{D} \text{ und } \lambda, \mu \in \mathbb{C}.$$

Ein Funktional T heißt *stetig*, falls aus $\phi_n \to \phi$ stets $T(\phi_n) \to T(\phi)$ folgt (für jede Folge $(\phi_n) \subset \mathcal{D}$).

Ein Funktional auf $\mathcal{D}$ ist eine komplexwertige (oder reellwertige) Abbildung, die auf einer Menge von **Funktionen** definiert ist. Linearität und Stetigkeit in $\mathcal{D}$ sind also völlig analog zu Linearität und Stetigkeit von Funktionen auf $\mathbb{R}$ definiert.

A.2.6 Definition: Eine *Distribution* oder *verallgemeinerte Funktion* ist ein stetiges lineares Funktional $T : \mathcal{D} \to \mathbb{C}$.

A.2.7 Beispiel: Der *Dirac-Impuls* δ ist jenes Funktional, welches einer Funktion $\phi : \mathbb{R} \to \mathbb{C},\ \phi \in \mathcal{D}$, ihren Wert an der Stelle 0 zuordnet. Man schreibt

$$\delta(\phi) := \int_{-\infty}^{\infty} \phi(t)\delta(t)dt := \phi(0). \tag{A.5}$$

Aufgrund der Eigenschaft (A.5) stellt der Dirac-Impuls δ eine bequeme Schreibweise dar, um einen physikalischen Impuls zu beschreiben. Man stellt sich δ daher häufig als einen unendlich kurzen, unendlich starken Impuls vor. Findet dieser Impuls nicht an der Stelle 0, sondern an einer beliebigen Stelle $\lambda \in \mathbb{R}$ statt, so schreibt man

$$\delta(t-\lambda)(\phi) := \int_{-\infty}^{\infty} \phi(t)\delta(t-\lambda)dt := \phi(\lambda). \tag{A.6}$$

Der Effekt einer Folge von Impulsen an den Stellen $\lambda_1, \lambda_2, \ldots$ lässt sich damit schreiben in der Form

$$(\sum_{i=1}^{\infty} \delta(t-\lambda_i))(\phi) := \sum_{i=1}^{\infty} (\delta(t-\lambda_i)(\phi)) = \sum_{i=1}^{\infty} \phi(\lambda_i). \tag{A.7}$$

A.2.8 Bemerkung: Es soll betont werden, dass eine Distribution **keine auf dem $\mathbb{R}^n$ definierte Funktion im herkömmlichen Sinne ist**, dass man vielmehr eine Testfunktion ϕ auf sie anwenden muss, um ihr einen Wert zuzuweisen. Häufig ist es jedoch sinnvoll, eine Distribution in Abhängigkeit von $t \in \mathbb{R}$ zu betrachten (man vergleiche die Schreibweisen (A.6) und (A.7)). Dies führt in gewissem Sinne auf eine Verallgemeinerung des Funktionen-Begriffes. Man bezeichnet Distributionen daher auch als verallgemeinerte Funktionen.

Eine kurze, empfehlenswerte Einführung in die Theorie der Distributionen bietet Göpfert / Riedrich (1994).

Anhang B. Begriffe aus der Wahrscheinlichkeitstheorie

B.1 Grundlegendes

B.1.1 Definition: Es sei X eine Zufallsvariable. Dann nennt man dasjenige Wahrscheinlichkeitsmaß P^X, das jeder geeigneten Teilmenge A des Zustandsraumes von X den Wert $P^X(A) = P(X \in A)$ zuordnet, die *Verteilung* von X.

B.1.2 Definition: Ist P ein Wahrscheinlichkeitsmaß und A ein Ereignis, so sagt man, A gelte *P-fast sicher (P-f.s.)*, falls $\mathrm{P}(A) = 1$.

Es seien $X_1, \ldots, X_n$ Zufallsvariablen mit gemeinsamer Verteilungsfunktion F und für $j = 1, \ldots, n$ bezeichne F_j die Verteilungsfunktion von X_j. Dann sind die Zufallsvariablen $X_1, \ldots, X_n$ genau dann *(stochastisch) unabhängig*, wenn für alle $t_1, \ldots, t_n \in \mathbb{R}$ gilt

$$F(t_1, \ldots, t_n) = F_1(t_1) \cdot \ \ldots \ \cdot F_n(t_n).$$

Im Falle einer diskreten Verteilung sind $X_1, \ldots, X_n$ genau dann *unabhängig*, wenn für alle $x_1, \ldots, x_n \in \mathbb{R}$:

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot \ \ldots \ \cdot P(X_n = x_n).$$

Besitzen alle $X_1, \ldots, X_n$ dieselbe Verteilung, so sagt man, $X_1, \ldots, X_n$ sind *identisch verteilt*. Sind $X_1, \ldots, X_n$ unabhängig und identisch verteilt mit Verteilungsfunktion F, so schreibt man

$$X_1, \ldots, X_n \overset{iid}{\sim} F.$$

Ist (X_t) eine Zeitreihe mit $X_1, X_2, \cdots \overset{iid}{\sim} F,\ \mathrm{E}(X_i) = \mu,\ \mathrm{Var}(X_i) = \sigma^2$, schreibt man

$$(X_t) \sim IID(\mu, \sigma^2).$$

Dabei steht "*iid*" bzw. "IID" für "independent and identically distributed".

B.1.3 Definition: Zwei Zufallsvariablen X und Y heißen *unkorreliert*, wenn

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y). \tag{B.1}$$

(B.1) ist gleichbedeutend mit $\mathrm{Cov}(X, Y) = 0$. Die Zufallsvariablen $X_1, X_2, \ldots$ heißen unkorreliert, falls $\mathrm{Cov}(X_i, X_j) = 0$ für alle $i \neq j$.

Unabhängige Zufallsvariablen sind stets unkorreliert. Aus der Unkorreliertheit folgt jedoch (außer bei normalverteilten Zufallsvariablen) **nicht** die Unabhängigkeit.

B.1.4 Definition: Es sei P_ϑ ein von einem Parameter $\vartheta \in \Theta$ abhängiges Wahrscheinlichkeitsmaß und E_ϑ bezeichne den entsprechenden Erwartungswert. Dann heißt eine Folge von Schätzern $(T_n)_{n\geq 1}$ für ϑ *asymptotisch erwartungstreu*, falls für alle $\vartheta \in \Theta$ gilt

$$\lim_{n\to\infty} \mathrm{E}_\vartheta(T_n) = \vartheta.$$

B.1.5 Definition: Es sei P_ϑ ein von einem Parameter $\vartheta \in \Theta$ abhängiges Wahrscheinlichkeitsmaß. Dann heißt eine Folge von Schätzern $(T_n)_{n\geq 1}$ für ϑ *konsistent*, falls für alle $\vartheta \in \Theta$ gilt

$$\lim_{n\to\infty} P_\vartheta(|T_n - \vartheta| \geq \epsilon) = 0 \ \forall \epsilon > 0. \tag{B.2}$$

B.1.6 Bemerkung: Der Einfachheit halber sagt man auch, der Schätzer T_n sei asymptotisch erwartungstreu bzw. konsistent für ϑ. Man beachte, dass (B.2) gleichbedeutend ist mit $T_n \xrightarrow{P_\vartheta} \vartheta$ (s. Abschnitt B.4.2).

B.1.7 Bemerkung: Ist T_n asymptotisch erwartungstreu und gilt $\mathrm{Var}(T_n) \overset{n\to\infty}{\longrightarrow} 0$, so ist T_n konsistent.

B.1.8 Definition: Es seien $X_1, \ldots, X_n$ Zufallsvariablen mit gemeinsamer Dichte f_ϑ, wobei ϑ ein unbekannter Parameter ist. (Z.B. $\vartheta = (\mu, \sigma^2)$ im Falle $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$.) Zu den Zufallsvariablen $X_1, \ldots, X_n$ seien Realisierungen $x_1, \ldots, x_n$ gegeben. Dann heißt die Funktion

$$L_x(\vartheta) = f_\vartheta(x_1, \ldots, x_n)$$

die *Likelihood-Funktion* zu $x = (x_1, \ldots, x_n)$. Die Likelihood-Funktion beschreibt die Wahrscheinlichkeit, dass unter Annahme des Parameters ϑ die Werte $x_1, \ldots, x_n$ beobachtet werden. Es ist also jener Wert $\hat{\vartheta}$ ein plausibler Schätzer für ϑ, der zu gegebenen beobachteten Werten $x_1, \ldots, x_n$ die Likelihoodfunktion bzgl. ϑ maximiert. Ein solcher Schätzer heißt *Maximum-Likelihood-Schätzer* für ϑ.

Die Maximum-Likelihood-Schätzung ist ein parametrisches Verfahren, d.h., abgesehen von dem zu schätzenden Parameter ϑ muss die Verteilung der $X_1, \ldots, X_n$ bekannt sein. Im Allgemeinen wird für Maximum-Likelihood-Verfahren eine Normalverteilung der Zufallsvariablen $X_1, \ldots, X_n$ vorausgesetzt, da sich dann $\ln(L_x(\vartheta))$ leicht berechnen und ableiten lässt.

B.2 Spezielle Verteilungen

B.2.1 Definition: Für $a \in \mathbb{R}$ sei δ_a dasjenige Maß auf $\mathbb{R}$ mit

$$\delta_a(A) = \begin{cases} 1, & a \in A, \\ 0, & a \notin A. \end{cases}$$

Man nennt δ_a das *Einpunktmaß* oder die *Einpunktverteilung in a*. Eine Zufallsvariable mit dem Einpunktmaß als Verteilung heißt *singulär verteilt*. Eine Zufallsvariable ist genau dann singulär verteilt, wenn sie f.s. (fast sicher, vgl. Definition B.1.2) konstant ist.

B.2.2 Definition: Es sei X eine Zufallsvariable mit Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}),\ \sigma^2 > 0,$$

dann heißt X *normalverteilt* mit Erwartungswert μ und Varianz σ^2, i.Z. $X \sim \mathcal{N}(\mu, \sigma^2)$.

B.2.3 Bemerkung: Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, dann besitzt die Zufallsvariable $Y := \nu \pm \tau \cdot X$ eine $\mathcal{N}(\nu \pm \tau\mu, \tau^2 \cdot \sigma^2)$-Verteilung.

B.2.4 Bemerkung: Ist $X \sim \mathcal{N}(\mu, \sigma^2)$ und $Y \sim \mathcal{N}(\nu, \tau^2)$, dann besitzt die Zufallsvariable $X \pm Y$ die Verteilung $\mathcal{N}(\mu \pm \nu, \sigma^2 + \tau^2)$.

B.2.5 Definition: Es sei $X = (X_1, \ldots, X_d)^\top$ ein d-dimensionaler Zufallsvektor und für jedes $X_j,\ j = 1, \ldots, d$, existiere der Erwartungswert $\mathrm{E}(X_j)$ von X_j. Dann heißt

$$\mathrm{E}(X) := (\mathrm{E}X_1, \ldots, \mathrm{E}X_d)^\top$$

der *Erwartungswertvektor* von X. Ist zusätzlich $\mathrm{E}X_j^2 < \infty\ \forall j$, so heißt die $d \times d$-Matrix

$$\Sigma := (\mathrm{Cov}(X_i, X_j))_{i,j=1}^d$$

die *Kovarianzmatrix* von X.

B.2.6 Definition: Der Zufallsvektor $X = (X_1, \ldots, X_d)^\top$ besitzt genau dann eine *d-dimensionale Normalverteilung*, falls für jedes $c = (c_1, \ldots, c_d)^\top \in \mathbb{R}^d$ die Zufallsvariable

$$c^\top X = \sum_{i=1}^d c_i X_i$$

eindimensional normalverteilt ist.

Die d-dimensionale Normalverteilung ist durch Erwartungswertvektor und Kovarianzmatrix eindeutig bestimmt.

B.2.7 Definition: Sind $X_1, \ldots, X_k$ stochastisch unabhängig und je $\mathcal{N}(0,1)$-verteilt, so heißt die Verteilung von

$$Y := X_1^2 + \cdots + X_k^2$$

χ^2-Verteilung mit k Freiheitsgraden (χ_k^2-Verteilung).

B.2.8 Definition: Sind $X \sim \chi_m^2$ und $Y \sim \chi_n^2$ stochastisch unabhängig, so nennt man die Verteilung der Zufallsvariablen

$$F := \frac{\frac{1}{m}X}{\frac{1}{n}Y}$$

eine F-Verteilung mit m und n Freiheitsgraden, i.Z. $F \sim F_{m,n}$.

B.2.9 Definition: Sind $X \sim \chi_m^2$ und $Y \sim \chi_n^2$ stochastisch unabhängig, dann heißt die Verteilung der Zufallsvariablen

$$B := \frac{X}{X+Y}$$

eine β-Verteilung mit Parametern $\frac{m}{2}$ und $\frac{n}{2}$, i.Z. $B \sim \beta_{\frac{m}{2},\frac{n}{2}}$.

B.2.10 Bemerkung: Die Dichte der Normalverteilung ist auch als *Gauß'sche Glockenkurve* bekannt, die Normalverteilung wird auch als *Gaußverteilung* bezeichnet. Die Normal-, Chi-Quadrat-, Beta- und F-Verteilungen sind, neben vielen anderen, ausführlich in den Büchern Johnson et al. (1994) bzw. Johnson et al. (1995) beschrieben.

B.3 Der Raum $L^2(P)$

B.3.1 Definition: Es sei P ein Wahrscheinlichkeitsmaß auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Für $k \in \mathbb{R},\ k > 0$, bezeichnet $L^k(P)$ die Menge aller reellwertigen Zufallsvariablen X auf $(\Omega, \mathcal{A}, P)$ mit

$$\mathrm{E}|X|^k < \infty.$$

Sind X und Y zwei Zufallsvariablen aus $L^k(P)$ mit $X = Y$ P-f.s., so werden X und Y in $L^k(P)$ miteinander identifiziert.

B.3.2 Bemerkung: $L^k(P)$ ist ein Vektorraum über $\mathbb{R}$, denn

$$\begin{array}{ll} (i) & X \in L^k(P) \quad \Rightarrow \alpha X \in L^k(P)\ (\alpha \in \mathbb{R}), \\ (ii) & X, Y \in L^k(P) \quad \Rightarrow X + Y \in L^k(P). \end{array}$$

Dabei folgt (ii) aus der Ungleichung

$$\begin{aligned} |X+Y|^k &\leq (|X| + |Y|)^k \\ &\leq (2\max(|X|,|Y|))^k \\ &\leq 2^k(|X|^k + |Y|^k). \end{aligned}$$

Für $k \geq 1$ wird auf $L^k(P)$ durch

$$\|X\|_k := \left(\mathrm{E}|X|^k\right)^{\frac{1}{k}}$$

eine Norm definiert, d.h. es gilt

$$\begin{array}{ll} (i) & \|X\|_k \geq 0, \text{ wobei } \|X\|_k = 0 \Leftrightarrow X = 0 \ P\text{-f.s.}, \\ (ii) & \|\alpha X\|_k = |\alpha| \|X\|_k,\ \alpha \in \mathbb{R}, \\ (iii) & \|X+Y\|_k \leq \|X\|_k + \|Y\|_k. \end{array}$$

Die Normeigenschaften (i) bis (iii) sind mit dem entsprechenden maßtheoretischen Hintergrund leicht zu überprüfen. Interessierte Leser werden in diesem Zusammenhang auf die maßtheoretische Standardliteratur (z.B. Billingsley (1995)) verwiesen.

B.3.3 Definition: Sind $X, X_1, X_2, \ldots$ Zufallsvariablen aus $L^k(P)$, $k \geq 1$, mit

$$\mathrm{E}|X_n - X|^k \overset{n\to\infty}{\longrightarrow} 0,$$

so sagt man, die Folge $(X_n)_{n\geq 1}$ *konvergiert im k-ten Mittel* gegen X, i.Z., $X_n \overset{L^k(P)}{\longrightarrow} X$.

B.3.4 Satz: *Der normierte Raum $L^k(P)$, $k \geq 1$, ist vollständig, d.h. jede Cauchyfolge $(X_n)_{n\geq 1}$ in $L^k(P)$ konvergiert im k-ten Mittel gegen ein $X \in L^k(P)$. $L^k(P)$ ist also ein Banachraum.*

Beweis: z.B. Billingsley (1995).

Von besonderer Bedeutung ist der normierte Raum $L^2(P)$ aller reellwertigen Zufallsvariablen mit existierendem zweiten Moment $\mathrm{E}(X^2)$. Auf diesem Banachraum wird durch

$$\langle X, Y\rangle := \mathrm{E}(XY) \tag{B.3}$$

ein Innenprodukt definiert, denn mit (B.3) sind die definierenden Eigenschaften eines Innenproduktes erfüllt:

$$\begin{array}{lll}
(i) & \langle X+Y, Z\rangle & = \langle X, Z\rangle + \langle Y, Z\rangle, \\
(ii) & \langle \alpha X, Y\rangle & = \alpha\langle X, Y\rangle,\ \alpha \in \mathbb{R}, \\
(iii) & \langle Y, X\rangle & = \langle X, Y\rangle, \\
(iv) & \langle X, X\rangle & \geq 0, \quad \text{wobei } \langle X, X\rangle = 0 \Leftrightarrow X = 0\ P\text{-f.s.}
\end{array}$$

Wegen $\langle X, X\rangle = \mathrm{E}(X^2)$ induziert das Innenprodukt (B.3) die Norm auf dem vollständigen normierten Raum $L^2(P)$. Der Raum $L^2(P)$ ist also ein Hilbertraum.

B.3.5 Bemerkung: Nach Definition sind zwei Zufallsvariablen X und Y aus $L^2(P)$ genau dann zueinander orthogonal, falls $\langle X, Y\rangle = 0$, d.h

$$X \perp Y \Longleftrightarrow \mathrm{E}(XY) = 0.$$

Sind $X, Y \in L^2(P)$ mit $\mathrm{E}X = \mathrm{E}Y = 0$, dann sind X und Y genau dann zueinander orthogonal, wenn sie unkorreliert sind.

Der folgende Satz ist in der Stochastik von großer Bedeutung und soll daher für einen beliebigen Hilbertraum $\mathcal{H}$, versehen mit einer Norm $\|\cdot\|$, formuliert werden. In Bezug auf den Hilbertraum $L^2(P)$ ersetze man also $\mathcal{H}$ durch $L^2(P)$, das Element $y \in \mathcal{H}$ durch eine Zufallsvariable $Y \in L^2(P)$, etc.

B.3.6 Satz: (Projektionssatz) *Es sei $\mathcal{F}$ ein abgeschlossener Unterraum des Hilbertraumes $\mathcal{H}$ und $y \in \mathcal{H}$ beliebig. Dann existiert genau ein $y^* \in \mathcal{F}$ mit*

$$\|y - y^*\| = \min_{x\in\mathcal{F}} \|y - x\| \tag{B.4}$$

und es gilt

$$y^* \in \mathcal{F} \text{ und } \|y - y^*\| = \min_{x\in\mathcal{F}} \|y - x\| \iff y^* \in \mathcal{F} \text{ und } y - y^* \perp \mathcal{F}.$$

Beweis: z.B. Brockwell / Davis (1991).

Der Projektionssatz besagt also, dass ein Element y^* aus $\mathcal{F}$ genau dann den Abstand zu y minimiert (bezgl. der Norm in $\mathcal{H}$ und über alle Elemente aus $\mathcal{F}$), wenn $y - y^*$ orthogonal zu $\mathcal{F}$ ist. Der Projektionssatz gewährleistet auch die Eindeutigkeit dieses Elementes y^*.

B.3.7 Definition: Das eindeutig bestimmte Element $y^* \in \mathcal{F}$ aus Satz B.3.6 heißt *Orthogonalprojektion (OGP) von y auf $\mathcal{F}$.* Der Operator $\mathrm{pr}_{\mathcal{F}}$, der jedem $y \in \mathcal{H}$ seine Orthogonalprojektion $\mathrm{pr}_{\mathcal{F}} y = y^*$ zuordnet, heißt OGP-Operator auf $\mathcal{F}$. Ein Operator M ist genau dann OGP-Operator auf $\mathcal{F}$, wenn gilt:

$$\begin{array}{ll}
\text{(i)} & y \in \mathcal{F} \implies My = y, \\
\text{(ii)} & y \perp \mathcal{F} \implies My = 0.
\end{array}$$

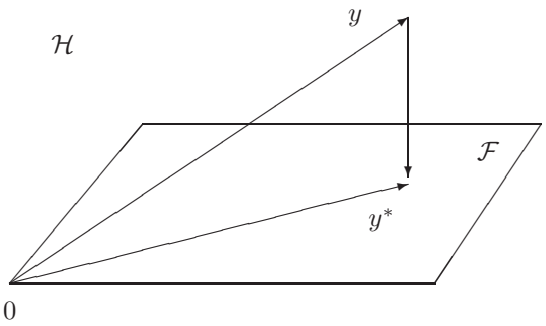

Abbildung B.1: Orthogonalprojektion

Abbildung B.1 veranschaulicht die Orthogonalprojektion graphisch.

B.3.8 Bemerkung: Für eine Projektionsabbildung $\mathrm{pr}_{\mathcal{F}}$ gilt stets

$$\mathrm{pr}_{\mathcal{F}} = \mathrm{pr}_{\mathcal{F}}^2. \tag{B.5}$$

Ist $\mathrm{pr}_{\mathcal{F}}$ eine Orthogonalprojektion, so ist $\mathrm{pr}_{\mathcal{F}}$ zusätzlich *selbstadjungiert*, d.h. es gilt

$$\langle \mathrm{pr}_{\mathcal{F}} x, y\rangle = \langle x, \mathrm{pr}_{\mathcal{F}} y\rangle \text{ für alle } x, y \in \mathcal{H}. \tag{B.6}$$

Ist $\mathcal{H} = \mathbb{R}^d$ und M die Abbildungsmatrix eines OGP-Operators, so entsprechen (B.5) und (B.6) den Eigenschaften

$$\begin{array}{lll} & \text{(i)} & M^2 = M, \\ \text{und} & \text{(ii)} & M^\top = M, \end{array}$$

d.h., M ist symmetrisch (ii) und *idempotent* (i).

Beweis: Christensen (1987).

B.4 Konvergenzarten

B.4.1 Konvergenz in $L^2(P)$

B.4.1 Definition: Sind $X, X_1, X_2, \ldots$ Zufallsvariablen aus $L^2(P)$ und gilt

$$\mathrm{E}|X_n - X|^2 \overset{n\to\infty}{\longrightarrow} 0,$$

so heißt $(X_n)_{n\geq 1}$ *im quadratischen Mittel gegen X konvergent*, i.Z.

$$X_n \overset{L^2(P)}{\longrightarrow} X \text{ oder l.i.m. } X_n = X \text{ in } L^2(P).$$

B.4.2 Lemma: *Sind $X, X_1, X_2, \ldots, Y, Y_1, Y_2, \ldots$ Zufallsvariablen aus $L^2(P)$ und gilt $X_n \overset{L^2(P)}{\longrightarrow} X$, $Y_n \overset{L^2(P)}{\longrightarrow} Y$, so gilt auch*

$$X_n + Y_n \overset{L^2(P)}{\longrightarrow} X + Y.$$

Beweis:

$$\mathrm{E}[(X_n + Y_n - X - Y)^2] = \mathrm{E}[(X_n - X)^2] + \mathrm{E}[(Y_n - Y)^2] + 2 \underbrace{\mathrm{E}[(X_n - X)(Y_n - Y)]}_{\leq E[(X_n - X)^2]^{1/2} E[(Y_n - Y)^2]^{1/2}} \to 0.$$

□

B.4.3 Lemma: *Es seien X_k, $k \in \mathbb{N}$, normalverteilte Zufallsvektoren und X_k konvergiere in $L^2(P)$ gegen X, d.h. $E(\|X_k - X\|_2^2) \to 0$. Dann ist auch X normalverteilt und für den Erwartungswertvektor bzw. die Kovarianzmatrix gilt*

$$E(X) = \lim_{k\to\infty} E(X_k) \text{ und } Cov(X) = \lim_{k\to\infty} Cov(X_k).$$

Beweis: s. Øksendal (1995).

B.4.2 Stochastische Konvergenz

Es seien $X, X_1, X_2, \ldots$ $\mathbb{R}^d$-wertige Zufallsvariablen.

B.4.4 Definition: Man sagt, die Folge $(X_n)_{n\geq 1}$ *konvergiert stochastisch gegen X (i.Z. $X_n \xrightarrow{P} X$)*, falls gilt

$$\mathrm{P}(\|X_n - X\| > \epsilon) \overset{n\to\infty}{\longrightarrow} 0 \ \forall \epsilon > 0.$$

Dabei ist $\|\cdot\|$ eine beliebige Norm des $\mathbb{R}^d$. Besitzt X eine Einpunktverteilung δ_a im Punkt $a \in \mathbb{R}^d$, so schreibt man auch $X_n \xrightarrow{P} a$.

B.4.5 Lemma: *Es seien $X, X_1, X_2, \ldots$ Zufallsvektoren mit Werten in $\mathbb{R}^d$. Dann gilt für jede stetige Funktion $h : \mathbb{R}^d \to \mathbb{R}^s$*

$$X_n \xrightarrow{P} X \Longrightarrow h(X_n) \xrightarrow{P} h(X).$$

Beweis: z.B. Gänssler / Stute (1977).

B.4.3 Verteilungskonvergenz

Es seien $X, X_1, X_2, \ldots$ $\mathbb{R}^d$-wertige Zufallsvariablen mit zugehörigen Verteilungsfunktionen $F, F_1, F_2, \ldots$

B.4.6 Definition: Die Folge $(X_n)_{n\geq 1}$ *konvergiert nach Verteilung (*oder *konvergiert schwach) gegen X*, falls gilt

$$\lim_{n\to\infty} F_n(x) = F(x)$$

für alle Stetigkeitsstellen x von F. Die Verteilung von X heißt *Grenzverteilung von* (X_n). Man schreibt

$$X_n \xrightarrow{\mathcal{D}} X.$$

B.4.7 Bemerkung: (Für $d = 1$:) Die Grenzverteilung einer schwach konvergenten Folge ist eindeutig bestimmt, d.h. aus $X_n \xrightarrow{\mathcal{D}} X$ und $X_n \xrightarrow{\mathcal{D}} Y$ folgt, dass X und Y dieselbe Verteilung besitzen.

Beweis: z.B. Gänssler / Stute (1977).

B.4.8 Bemerkung: (Für $d = 1$:) Aus $X_n \xrightarrow{P} X$ folgt $X_n \xrightarrow{\mathcal{D}} X$.

Beweis: z.B. Billingsley (1995).

B.4.9 Bemerkung: (Für $d = 1$:) Besitzt X eine Einpunktverteilung δ_a für eine reelle Zahl $a \in \mathbb{R}$, so folgt aus der Verteilungskonvergenz $X_n \xrightarrow{\mathcal{D}} X$ die stochastische Konvergenz $X_n \xrightarrow{P} a$.

Beweis: z.B. Billingsley (1995).

B.4.10 Satz: (Cramér-Wold) *Es seien $X, X_1, X_2, \ldots$ d-dimensionale Zufallsvektoren. Dann gilt*

$$X_n \xrightarrow{\mathcal{D}} X \Longleftrightarrow a^\top X_n \xrightarrow{\mathcal{D}} a^\top X \ \forall a = (a_1, \ldots, a_d) \in \mathbb{R}^d.$$

Beweis: z.B. Billingsley (1995).

Für beliebige normierte Räume S definieren wir die Verteilungskonvergenz wie folgt:

B.4.11 Definition: Es sei S ein normierter Raum und $X, X_n\ ,\ n \in \mathbb{N}$, Zufallselemente aus S mit zugehörigen Verteilungen $P, P_n\ ,\ n \in \mathbb{N}$. Gilt für jedes Ereignis A mit $P(\partial A) = 0$ (∂A bezeichnet den Rand der Menge A, siehe Definition A.1.17)

$$\lim_{n\to\infty} P_n(A) = P(A),$$

so sagt man, P_n *konvergiert schwach* gegen P (i.Z. $P_n \Rightarrow P$) oder X_n *konvergiert nach Verteilung* gegen X (i.Z. $X_n \xrightarrow{\mathcal{D}} X$).

B.4.12 Satz: (Abbildungssatz) *Es seien S und S' normierte Räume und $X, X_1, X_2, \ldots$ Zufallselemente mit Werten in S sowie $X_n \xrightarrow{\mathcal{D}} X$ in S. Weiter sei $h : S \to S'$ eine stetige Abbildung. Dann gilt*

$$h(X_n) \xrightarrow{\mathcal{D}} h(X) \text{ in } S'.$$

Beweis: z.B. Billingsley (1968).

B.4.13 Lemma: *Es sei S ein normierter Raum, $X, X_1, X_2, \ldots, Y_1, Y_2, \ldots$ Zufallselemente mit Werten in S sowie $a \in S$. Dann gilt*

$$\left.\begin{array}{l} X_n \xrightarrow{\mathcal{D}} X \\ Y_n \xrightarrow{P} a \end{array}\right\} \Longrightarrow (X_n, Y_n) \xrightarrow{\mathcal{D}} (X, a).$$

Beweis: z.B. Billingsley (1968).

B.4.14 Lemma: *Gilt $X_n \xrightarrow{\mathcal{D}} X$ sowie $Y_n \xrightarrow{P} a$, so folgt*

$$\begin{array}{rcl} X_n + Y_n & \xrightarrow{\mathcal{D}} & X + a \\ X_n \cdot Y_n & \xrightarrow{\mathcal{D}} & X \cdot a. \end{array}$$

Beweis: folgt sofort aus Lemma B.4.13 und dem Abbildungssatz.

B.4.15 Satz: *Es sei S ein normierter Raum und $X, X_n, Y_n,\ n \in \mathbb{N}$, Zufallselemente mit Werten in S. Dann gilt*

$$\left.\begin{array}{l} X_n \xrightarrow{\mathcal{D}} X \\ \|X_n - Y_n\| \xrightarrow{P} 0 \end{array}\right\} \Longrightarrow Y_n \xrightarrow{\mathcal{D}} X.$$

Beweis: z.B. Billingsley (1968).

B.4.16 Definition: Es sei g von beschränkter Variation auf $[0,1]$ und B eine reelle, normale Brown'sche Bewegung. Weiter sei $0 = z_{n0} < \cdots < z_{nn} = 1$ mit $\max_{j=1}^{n} |z_{nj} - z_{n,j-1}| \overset{n\to\infty}{\longrightarrow} 0$. Dann definiert man

$$\int_0^1 g(z)dB(z) := \text{l.i.m.}_{n\to\infty} \sum_{j=1}^{n} g(z_{nj})[B(z_{nj}) - B(z_{n,j-1})] \text{ (Grenzwert in } L^2(P)).$$

B.4.17 Lemma: *Ist B eine reelle, normale Brown'sche Bewegung und g von beschränkter Variation auf $[0,1]$, so gilt*

$$\int_0^1 g(z)dB(z) \sim \mathcal{N}\left(0, \int_0^1 g(z)^2 dz\right).$$

Beweis: Es seien $z_{n0} < \cdots < z_{nn}$ wie in Definition B.4.16. Aus $B(z_{nj}) - B(z_{n,j-1}) \sim \mathcal{N}(0, z_{nj} - z_{n,j-1})$ folgt (da die Zuwächse der Brown'schen Bewegung unabhängig sind), dass

$$\sum_{j=1}^{n} g(z_{nj})[B(z_{nj}) - B(z_{n,j-1})] \tag{B.7}$$

ebenfalls normalverteilt ist mit Erwartungswert 0 und Varianz $\sum_{j=1}^{n} g(z_{nj})^2[z_{nj} - z_{n,j-1}]$. Da (B.7) in $L^2(P)$ gegen $\int_0^1 g(z)dB(z)$ konvergiert, folgt mit Lemma B.4.3 die Behauptung. □

B.4.18 Definition: Ist B eine reelle, normale Brown'sche Bewegung und g von beschränkter Variation auf $[0,1]$, so definiert man

$$\int_0^1 B(z)dg(z) := \text{l.i.m.}_{n\to\infty} \sum_{j=1}^{n} B(z_{n,j-1})[g(z_{nj}) - g(z_{n,j-1})] \text{ (Grenzwert in } L^2(P)).$$

mit $z_{n0} < \cdots < z_{nn}$ wie in Definition B.4.16.

B.4.19 Lemma: *Es gilt*

$$\int_0^1 g(z)dB(z) = g(1)B(1) - \int_0^1 B(z)dg(z).$$

Beweis: Es seien $z_{n0} < \cdots < z_{nn}$ wie in Definition B.4.16. Dann ist

$$\begin{aligned}
\sum_{j=1}^{n} g(z_{nj})[B(z_{nj}) - B(z_{n,j-1})] &= g(z_{n1})[B(z_{n1}) - B(z_{n0})] + g(z_{n2})[B(z_{n2}) - B(z_{n1})] \\
&\quad + \cdots + g(z_{nn})[B(z_{nn}) - B(z_{n,n-1})] \\
&= \underbrace{-g(z_{n1})B(z_{n0})}_{=0 \ P-f.s.} - \underbrace{\sum_{j=2}^{n} B(z_{n,j-1})[g(z_{nj}) - g(z_{n,j-1})]}_{\overset{L^2(P)}{\longrightarrow} \int_0^1 B(z)dg(z)} \\
&\quad + \underbrace{g(z_{nn})B(z_{nn})}_{=g(1)B(1)}.
\end{aligned}$$

□

B.4.20 Lemma: *Ist B eine reelle, normale Brown'sche Bewegung, so gilt*

$$\int_0^1 B(z)dz \sim \mathcal{N}(0, \frac{1}{3}).$$

Beweis: Da B ein Gauß-Prozess ist, besitzt jede Linearkombination von Komponenten eines Vektors $(B(z_{n0}), \ldots, B(z_{nn}))^\top$ eine Normalverteilung, insbesondere auch die Summe $\sum_{i=1}^{n} B(z_{ni})(z_{ni} - z_{n,i-1})$. Somit ist $\int_0^1 B(z)dz$ als $L^2(P)$-Grenzwert dieser Summe nach Lemma B.4.3 ebenfalls normalverteilt. Nach Lemma B.4.19 und Lemma B.4.17 ist

$$\mathrm{E}\left(\int_0^1 B(z)dz\right) = \mathrm{E}(B(1)) - \mathrm{E}\left(\int_0^1 zdB(z)\right) = 0 \ P - \text{f.s.}$$

Außerdem gilt

$$\begin{aligned}
\mathrm{Var}\left(\int_0^1 B(z)dz\right) &= \mathrm{Var}(B(1)) + \mathrm{Var}\left(\int_0^1 zdB(z)\right) - 2\mathrm{Cov}\left(B(1), \int_0^1 zdB(z)\right) \\
&= 1 + \frac{1}{3} - 2\lim_{n\to\infty} \sum_{j=1}^{n} z_{nj} \underbrace{\mathrm{Cov}(B(1), B(z_{nj}) - B(z_{n,j-1}))}_{\mathrm{Cov}(B(1),B(z_{nj})) - \mathrm{Cov}(B(1),B(z_{n,j-1}))} \\
&= 1 + \frac{1}{3} - 2\lim_{n\to\infty} \sum_{j=1}^{n} z_{nj}(z_{nj} - z_{n,j-1}) \\
&= 1 + \frac{1}{3} - 2\int_0^1 zdz = \frac{1}{3}.
\end{aligned}$$

□

Anhang C. Buchstaben und Symbole

Tabelle C.1: Die gängigsten griechischen Buchstaben

α		alpha	μ		mü
β		beta	ν		nü
γ	Γ	gamma	ξ	Ξ	xi
δ	Δ	delta	π	Π	pi
ϵ		epsilon	ρ		rho
ζ		zeta	σ	Σ	sigma
η		eta	τ		tau
θ, ϑ	Θ	theta	ϕ, φ	Φ	phi
ι		jota	χ		chi
κ		kappa	ψ	Ψ	psi
λ	Λ	lambda	ω	Ω	omega

Tabelle C.2: Abkürzungen und Symbole

$\exists$	es existiert	∇	Nabla (Differenzenoperator)
$\forall$	für alle	∂	partielle Ableitung
$\setminus$	Mengendifferenz	∂M	Rand der Menge M
$\oplus$	direkte Summe	$M^\perp$	orthogon. Komplement von M
$\perp$	orthogonal zu	$\overline{M}$	Abschluss von M
pr_U	Projektionsabb. auf U	$\langle M \rangle$	lineare Hülle von M
$A^\top$	Transponierte der Matrix A	$\mathrm{sp}(M)$	lineare Hülle von M
A^-	verallgemeinerte Inverse v. A	$\overline{\langle M \rangle}$	abgeschl. lineare Hülle von M
$C(A)$	Bildraum von A	$\overline{\mathrm{sp}}(M)$	abgeschl. lineare Hülle von M
$*$	Faltung	$\overline{X}$	arithm. Mittel des Vektors X
$\approx$	approximativ	$\overline{X_{j\cdot}}$	arithm. Mittel des Vektors X_j
$\sim$	besitzt die Verteilung	$\overline{z}$	zu z konjugiert komplexe Zahl
$\overset{iid}{\sim}$	unabh. u. identisch verteilt	$[x]$	größte ganze Zahl $\leq x$
$\xrightarrow{\mathcal{D}}$	Verteilungskonvergenz	$\lvert x \rvert$	Betrag von x
$\xrightarrow{P}$	stochastische Konvergenz	$\lVert x \rVert$	Norm von x
$\xrightarrow{L_2(P)}$	Konvergenz in $L_2(P)$	$\langle x, y \rangle$	Innenprodukt von x und y
$\hat{\theta}$	Schätzer für θ	$\mathrm{sgn}(x)$	Vorzeichen von x

Danksagung

Mein besonderer Dank gilt Herrn Prof. Dr.-Ing. Bernhard Heck und Herrn Prof. Dr. Wolfgang Bischoff für ihre Förderung und Unterstützung, für unermüdliches Korrekturlesen und für viele fruchtbare Anregungen. Dank gebührt auch Herrn Dr.-Ing. Jochen Howind sowie Herrn Prof. Dr.-Ing. Hansjörg Kutterer für wertvolle Hinweise und eine gute Zusammenarbeit. Der Deutschen Forschungsgemeinschaft (DFG) danke ich für die finanzielle Unterstützung des Projektes. Der DGK sei recht herzlich für die Veröffentlichung des Berichts gedankt.

Karlsruhe, im Juli 2006 Annette Teusch

Literatur

[1] Anderson T. W., Darling D. A. (1952): *Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes*; Annals of Mathematical Statistics, **23**, 193-212.

[2] Anderson T. W. (1993): *Goodness of Fit Tests for Spectral Distributions*; The Annals of Statistics, Vol. 21, No. 2, 830-847.

[3] Ash R. B., Gardner M. F. (1975): *Topics in Stochastic Processes*; Academic Press.

[4] Bachman G., Narici L., Beckenstein E. (2000): *Fourier and Wavelet Analysis*; Springer.

[5] Bartlett M. S. (1937): *Properties of Sufficiency and Statistical Tests*; Proceedings of the Royal Society, Series A, Vol. 160, 268-282.

[6] Bartlett M. S. (1966): *An Introduction to Stochastic Processes with Special Reference to Methods and Applications*; Cambridge University Press.

[7] Beran J. (1994): *Statistics for Long-Memory Processes*; Chapman & Hall.

[8] Billingsley P. (1968): *Convergence of Probability Measures*; Wiley.

[9] Billingsley P. (1995): *Probability and Measure*; 3rd Ed., Wiley.

[10] Bischoff W. (1998): *A functional central limit theorem for regression models*; The Annals of Statistics, Vol. 26, No. 4, 1398-1410.

[11] Bischoff W., Heck B., Howind J., Teusch A. (2005): *A procedure for testing the assumption of homoscedasticity in least squares residuals. A case study of GPS carrier-phase observations*; Journal of Geodesy, Vol. 78, 397-404.

[12] Bischoff W., Heck B., Howind J., Teusch A. (2006): *A procedure for estimating the variance function of linear models and for checking the appropriateness of estimated variances: a case study of GPS carrier-phase observations*; Journal of Geodesy, Vol. 79, 694-704.

[13] Box G. E. P., Cox D. R. (1964): *An Analysis of Transformations*; Journal of the Royal Statistical Society, Series B, Vol. 26, 211-246.

[14] Box G. E. P., Pierce D. A. (1970): *Distribution of Residual Autocorrelations in Autoregressiv-Integrated Moving Avarage Time Series Models*; Journal of the American Statistical Association, **65**, 1509-1526.

[15] Brillinger D. R., Rosenblatt M. (1967): *Asymptotic Theory of Estimates of k-th Order Spectra* in *SPECTRAL ANALYSIS of TIME SERIES* (Editor B. Harris); Wiley.

[16] Brockwell P. J., Davis R. A. (1988): *Simple consistent estimation of the coefficients of a linear filter*; Stochastic Processes and Their Applications, **22**, 47-59.

[17] Brockwell P. J., Davis R. A. (1991): *Time Series: Theory and Methods*; Springer.

[18] Brockwell P. J., Davis R. A. (1996): *Introduction to Time Series and Forecasting*; Springer.

[19] Brown R. L., Durbin J., Evans J. M. (1975): *Techniques for Testing the Constancy of Regression Relationship over Time*; Journal of the Royal Statistical Society, Series B, Vol. 37,149-192.

[20] Carroll R. J., Ruppert D. (1988): *Transformation and Weighting in Regression*; Chapman & Hall.

[21] Christensen R. (1987): *Plane Answers to Complex Questions*; Springer.

[22] Conover W. J. (1971): *Practical Nonparametric Statistics*; Wiley.

[23] Cook R. D., Weisberg S. (1982): *Residuals and Influence in Regression*; Chapman & Hall.

[24] Dahlhaus R., Giraitis L. (1998): *On the Optimal Segment Length for Parameter Estimates for Locally Stationary Time Series*; Journal of Time Series Analysis, Vol. 19, No. 6, 629-655.

[25] Daubechies I. (1988): *Orthonormal Bases of Compactly Supported Wavelets*; Communications on Pure and Applied Mathematics, **41**, 909-996.

[26] David H. A.(1970): *Order Statistics*; Wiley.

[27] Dette H., Munk A. (1998): *Testing heteroscedasticity in nonparametric regression*; Journal of the Royal Statistical Society, Series B, Vol. 60, No. 4, 693-708.

[28] Durbin J. (1969): *Tests for serial correlation in regression analysis based on the periodogram of least squares residuals*; Biometrika, **56**, 1-15.

[29] Feller W. (1948): *On the Kolmogorov-Smirnov theorems for empirical distributions*; Annals of Mathematical Statistics, **19**, 177-189.

[30] Ferguson Th. S. (1973): *A Bayesian Analysis of some Nonparametric Problems*; The Annals of Statistics, Vol. 1, No. 2, 209-230.

[31] Fisk P. R. (1975): *Discussion of the Paper by Brown, Durbin and Evans*; Journal of the Royal Statistical Society, Series B, Vol. 37, No. 2, 164-166.

[32] Gänssler P., Stute W. (1977): *Wahrscheinlichkeitstheorie*; Springer.

[33] Godolphin E. J., de Tullio M. (1978): *Invariance Properties of Uncorrelated Residual Transformations*; Journal of the Royal Statistical Society, Series B, Vol. 40, No. 3, 313-321.

[34] Göpfert A., Riedrich Th. (1994): *Funktionalanalysis*; Teubner.

[35] Goldfeld S. M., Quandt R. E. (1965): *Some Tests for Heteroscedasticity*; Journal of the American Statistical Association, **60**, 539-547.

[36] Grenander U., Rosenblatt M. (1957): *Statistical Analysis of Stationary Time Series*; Wiley.

[37] Graybill F. A. (1976): *Theory and Application to the Linear Model*; Wadsworth.

[38] Hart B. I. (1942): *Significance Levels for the Ratio of the Mean Square Successive Difference to the Variance*; Annals of Mathematical Statistics, **19**, 445-447.

[39] Hartley H. O. (1950): *Maximum F-ratio as a shortcut test for heterogeneity of variance*; Biometrika, **37**, 308-312.

[40] Hartung J. (1987): *Statistik*; Oldenburg Verlag.

[41] Hesse Ch. (2003): *Angewandte Wahrscheinlichkeitstheorie*; Vieweg.

[42] Heuser H. (1992): *Funktionalanalysis*; Teubner.

[43] Heuser H. (1986[1]): *Lehrbuch der Analysis, Teil 1*; Teubner.

[44] Heuser H. (1986[2]): *Lehrbuch der Analysis, Teil 2*; Teubner.

[45] Higgins J. R. (1996): *Sampling Theory in Fourier and Signal Analysis*; Oxford University Press.

[46] Hinich M. J. (1982): *Testing for Gaussianity and Linearity of a Stationary Time Series*; Journal of Time Series Analysis, Vol. 3, No. 3, 169-176.

[47] Horn R. A., Johnson Ch. R. (1988): *Matrix Analysis*; Cambridge University Press.

[48] Hsu D. A. (1977): *Tests for Variance Shift at an Unknown Time Point*; Applied Statistics, Vol. 26, No. 3, 279-284.

[49] Johnson N. L., Kotz S., Balakrishnan N. (1994): *Continous Univariate Distributions, Vol. 1*; Wiley.

[50] Johnson N. L., Kotz S., Balakrishnan N. (1995): *Continous Univariate Distributions, Vol. 2*; Wiley.

[51] Johnson N. L., Kotz S., Balakrishnan N. (2000): *Continous Multivariate Distributions, Vol. 1*; Wiley.

[52] Kalman R. E. (1960): *A new approach to linear filtering and prediction problems*; Transactions of the ASME, Journal of Basic Engineering, 82D, 35-45.

[53] Katznelson Y. (1968): *An introduction to HARMONIC ANALYSIS*; Wiley.

[54] Kendall M., A. Stuart, J. K. Ord (1983): *The Advanced Theory of Statistics, Vol. 3*; Griffin & Co.

[55] Klees R., Haagmans R. (Eds.) (2000): *Wavelets in the Geosciences*; Springer.

[56] Kulperger R. J., Lockhart R. A. (1998): *Tests of Independence in Time Series*; Journal of Time Series Analysis, Vol. 19, No. 2, 165-186.

[57] Lehmann E. L. (1986): *Testing Statistical Hypotheses*; Wiley.

[58] Ljung G. M., Box G. E. P. (1978): *On a measure of lack of fit in time series models*; Biometrika, **65**, 2, 297-303.

[59] Mallat S. (1999): *A Wavelet Tour of Signal Processing*; Academic Press.

[60] McGilchrist C. A., Sandland R. L. (1979): *Recursive Estimation of the General Linear Model with Dependent Errors*; Journal of the Royal Satistical Society, Series B, Vol. 41, No. 1, 65-68.

[61] Neuburger E. (1972): *Einführung in die Theorie des linearen Optimalfilters*; Oldenburg Verlag.

[62] Nievergelt Y. (1999): *Wavelets Made Easy*; Birkhäuser.

[63] Øksendal B. (1995): *Stochastic Differential Equations*; Springer.

[64] Pearson E. S., Hartley H.O. (1970): *Biometrika Tables for Statisticians, Vol. 1*; Cambridge Univeersity Press.

[65] Percival D. B., Walden A. T. (2000): *Wavelet Methods for Time Series Analysis*; Cambridge Series in Statistical and Probabilistic Mathematics.

[66] Priestley M. B. (1980): *State-Dependent Models: A General Approach to Non-Linear Time Series Analysis*; Journal of Time Series Analysis, Vol. 1, No. 1, 47-71.

[67] Priestley M. B. (1981[1]): *Spectral Analysis and Time Series, Vol. 1*; Academic Press.

[68] Priestley M. B. (1981[2]): *Spectral Analysis and Time Series, Vol. 2*; Academic Press.

[69] Schoknecht A. (2001): *Analyse auf Homoskedastizität in Linearen Regressionsmodellen und Anwendung der Resultate zur Prognose von branchenspezifischen Insolvenzquoten*; Diplomarbeit am Institut für Mathematische Stochastik der Universität Karlsruhe (unveröffentlicht).

[70] Seber G. A. F. (1977): *Linear Regression Analysis*; Wiley.

[71] Seber G. A. F., Wild C. J. (1989): *Nonlinear Regression*; Wiley.

[72] Strichartz R. S. (1993): *How to make Wavelets*; The American Mathematical Monthly, Vol. 100, No. 6, 539-556.

[73] Subba Rao T., Gabr M. M. (1984): *An Introduction to Bispectralanalysis and Bilinear Time Series Models*; Springer.

[74] Tanaka K. (1996): *Time Series Analysis*; Wiley.

[75] Tanizaki H. (1995): *Asymptotically Exact Confidence Intervals of CUSUM and CUSUMSQ Tests*; Communications in Statistics - Simulations, Vol. 24, No. 4, 1019-1036.

[76] Theil H. (1965): *The Analysis of Disturbances in Regression Analysis*; Journal of the American Statistical Association, **60**, 1067-1079.

[77] Theil H. (1968): *A Simplification of the BLUS Procedure for Analyzing Regression Disturbances*; Journal of the American Statistical Association, **63**, 242-251.

[78] Vidakovic B. (1999): *Statistical Modelling by Wavelets*; Wiley Series in Probability and Statistics.

[79] Walter W. (1974): *Einführung in die Theorie der Distributionen*; B.I.-Wissenschaftsverlag.

[80] Wludyka P. S., Nelson P. R. (1997): *An Analysis of Means-Type Test for Variances From Normal Populations*; Technometrics, Vol. 39, No. 3, 274-285.

Index