

Statistik und ihre Anwendungen

Jens-Peter Kreiß
Georg Neuhaus

Einführung in die Zeitreihenanalyse



Springer

Wissenschaftliche Berater:

Prof. Dr. Holger Dette · Prof. Dr. Wolfgang Härdle

Jens-Peter Kreiß · Georg Neuhaus

Einführung in die Zeitreihenanalyse

Mit 86 Abbildungen und 8 Tabellen

 Springer

Professor Dr. Jens-Peter Kreiß
Technische Universität Carolo-Wilhelma zu Braunschweig
Institut für Mathematische Stochastik
Carl-Friedrich-Gauß-Fakultät für Mathematik und Informatik
Pockelstraße 14
38106 Braunschweig
j.kreiss@tu-bs.de

Professor Dr. Georg Neuhaus
Universität Hamburg
Schwerpunkt Mathematische Statistik
und Stochastische Prozesse
Department Mathematik
Bundesstraße 55
20146 Hamburg
neuhaus@math.uni-hamburg.de

ISBN-10 3-540-25628-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-25628-1 Springer Berlin Heidelberg New York

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <<http://dnb.ddb.de>> abrufbar.

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media
springer.de

© Springer-Verlag Berlin Heidelberg 2006
Printed in Germany

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: Design & Production, Heidelberg

SPIN 11381266 154/3153-5 4 3 2 1 0 – Gedruckt auf säurefreiem Papier

Für unsere Kinder
Alexander, Franziska und Johanna
Sabine und Susanne

Vorwort

Die vorliegende Einführung in die Zeitreihenanalyse hat ihre Wurzeln in einer Vorlesung zu diesem Thema, die der zweite Autor vor mehr als 20 Jahren an der Universität Hamburg gehalten hat und in der der erste Autor zum ersten Mal in Kontakt mit diesem Gebiet gekommen ist. Das Ziel war damals und ist heute, wichtige Bereiche der Zeitreihenanalyse einem mathematisch orientierten Leserkreis auf einem angemessenen Niveau zu vermitteln.

Unsere Darstellung beginnt mit Beispielen von Zeitreihen, wie sie uns in unserem täglichen Leben begegnen und versucht erste interessierende Fragestellungen aufzuwerfen. Daran schließt sich eine Diskussion grundlegender Begriffe wie Stationarität, Autokovarianz und Autokorrelation von Zeitreihen, Spektralmaß und Spektraldichte sowie Vorhersage von Zeitreihen an. Hierbei begegnen uns auch Modelle für Zeitreihen, wie ein weißes Rauschen oder autoregressive Zeitreihen, zum ersten Mal. Bei der Vorstellung der Sonnenfleckenzahlen (sunspot numbers) gehen wir auch auf die Historie dieser klassischen Zeitreihe ein.

Ausgerüstet mit den Grundwerkzeugen der Zeitreihenanalyse behandeln wir anschließend in Kapitel 5 ein zentrales theoretisches Resultat, nämlich den Spektralsatz der Zeitreihenanalyse. Ausgehend von relativ leicht zu interpretierenden zyklischen Zeitreihen definieren und motivieren wir ein stochastisches Integral als Skalarprodukt erhaltende lineare Abbildung (Isometrie), mit dessen Hilfe wir jede stationäre Zeitreihe darstellen können. Durch das Zurückgreifen auf Isometrien haben wir für das stochastische Integral eine Reihe von Rechenregeln zur Hand, die uns den Umgang mit diesem und auch den Beweis des Spektralsatzes sehr erleichtern. Bei der Betrachtung von Filtern profitieren wir ebenfalls von diesem Vorgehen, sodass sich der abstrakte Zugang zum Spektralsatz lohnt.

Einfache autoregressive Zeitreihen konnten wir relativ früh einführen und direkt auf ihre Eigenschaften hin untersuchen. Für die Behandlung der grundlegenden *autoregressive moving average* (kurz: *ARMA*)-Zeitreihen können wir in Kapitel 7 auf den Spektralsatz und vor allem auf die Filtertechnik

zurückgreifen. So kann diese zentrale Klasse linearer Zeitreihenmodelle effizient und vollständig behandelt werden. Insbesondere die Frage nach der Existenz (bzw. der eindeutigen Existenz) von Lösungen von $ARMA$ -Gleichungen und die Struktur der Autokovarianzfunktion wird erschöpfend behandelt.

Mit Kapitel 10 beginnen wir den statistischen Teil unserer Einführung in die Zeitreihenanalyse. Hierin stellen wir zunächst das aktuelle Konzept der schwachen Abhängigkeit vor, welches auf Paul Doukhan zurück geht und über Forderungen an die Kovarianzstruktur von Zufallsvariablen definiert wird. Durch den Einsatz dieses Konzeptes vermeiden wir einerseits die Behandlung von Mischungseigenschaften, um die man bei der asymptotischen Verteilungstheorie für nichtlineare, abhängige Zufallsvariable kaum herum kommt. Andererseits stellen wir dem Leser ein modernes Konzept und ein zentrales Grenzwertresultat aus der Literatur zur Verfügung, das über die Behandlung von speziellen Zeitreihenmodellen hinaus zur Anwendung kommen kann. Natürlich können wir das Konzept der schwachen Abhängigkeit nicht in seiner gesamten Breite behandeln, aber wir formulieren und beweisen alle in dieser Einführung verwendeten Resultate vollständig. Mit dem erwähnten zentralen Grenzwertsatz für schwach abhängige Zufallsvariable können wir die asymptotische Normalität der empirischen Autokovarianzfunktion für lineare Zeitreihen zeigen. Hieraus lassen sich in Kapitel 11 asymptotische Verteilungen, insbesondere für die Schätzer der Parameter einer autoregressiven Zeitreihe, herleiten. Prominentestes Beispiel für derartige Schätzer sind die Yule-Walker-Parameterschätzer, denen wir uns ausführlich zuwenden. Aber auch auf der Methode der kleinsten Quadrate oder dem Maximum-Likelihood-Ansatz basierende Parameterschätzer werden für AR - und $ARMA$ -Zeitreihen vorgestellt. Die theoretischen Ergebnisse und insbesondere die asymptotischen Normalverteilungsaussagen ergänzen wir durch kleine Simulationsstudien.

Wir verlassen im Anschluss die statistischen Fragen im Zeitbereich der Zeitreihenanalyse und wenden uns in Kapitel 12 dem Schätzen im sogenannten Spektralbereich, d.h. im Wesentlichen dem Schätzen der Spektraldichte einer stationären Zeitreihe, zu. Dort steht zunächst das Periodogramm mit seinen Eigenschaften im Vordergrund. Da dieses aber keinen konsistenten Schätzer der Spektraldichte abgibt, wenden wir uns der Suche nach derartigen Schätzern zu. Dies führt uns zu Glättungsverfahren für das Periodogramm und sogenannten Lag-Window-Schätzern. Auch hier haben wir uns um eine in sich abgeschlossene Darstellung – einschließlich der asymptotischen Verteilungsaussagen – bemüht. Es versteht sich dabei von selbst, dass im Rahmen einer Einführung in die Zeitreihenanalyse nicht alle Aspekte und Feinheiten der Spektraldichteschätzung behandelt werden können. Ein Einblick in diesen wesentlichen Bereich der Zeitreihenanalyse ist aber unabdingbar. Zur Veranschaulichung der Glättungstechniken werden zusätzlich die Ergebnisse einiger Simulationen dargelegt.

Die Kapitel 10 bis 12 enthalten mehrere Resultate, die im Beweis einer längeren Argumentation bedürfen. Wir haben uns bemüht, die wesentlichen

Beweisideen zu vermitteln, aber an der einen oder anderen Stelle längere technische Rechnungen unterdrückt.

Kapitel 13 gibt einen Einblick in den Bereich der Ordnungswahl von *ARMA*-Zeitreihenmodellen. Im Gegensatz zu den bisherigen Kapiteln greifen wir an dieser Stelle mitunter auf heuristische Argumentationsweisen zurück. Eine mathematisch vollständige Behandlung erfordert einen recht hohen technischen Aufwand, der uns für eine Einführung in das Gebiet der Zeitreihenanalyse nicht angebracht erscheint. Auch hier findet sich das ein oder andere Simulationsbeispiel.

Eine moderne Einführung in die Zeitreihenanalyse kann nicht ohne eine Einführung in die Grundlagen finanzieller Zeitreihen auskommen. Bedingt durch den enormen Bedarf an adäquater Modellierung von umfangreichen finanziellen Datenreihen, die fast immer in Form von Zeitreihen vorliegen, und verstärkt durch die Würdigung der bedingt heteroskedastischen Modellierung von Zeitreihen (vgl. Engle (1982)) durch die Verleihung des Preises der Schwedischen Nationalbank zu Ehren Alfred Nobels (auch Nobelpreis genannt) an Robert F. Engle im Jahre 2003, wurden in diesem Umfeld in den vergangenen Jahren neuartige Modellklassen entwickelt, die einen eigenständigen Bereich innerhalb der Zeitreihenanalyse begründet haben. Wir werden die Grundlagen der von Engle eingeführten *ARCH*-Zeitreihen und deren Erweiterung zu *GARCH*-Modellen in Kapitel 14 ausführlich und mathematisch rigoros behandeln. Gleichzeitig wollen wir am Beispiel des Deutschen Aktienindex (DAX) das Potential dieser Modelle für die Beschreibung derartiger finanzieller Zeitreihen untersuchen.

Zum Abschluss stellen wir in Kapitel 15 grundlegende Resultate über multivariate stationäre Zeitreihen zusammen. Dieses Kapitel ist lediglich als eine Art Übersicht zu verstehen, ohne dass wir hier in die Tiefe gehen.

Eine Reihe von Anhängen mit häufig benötigten Formeln, grundlegenden mathematischen Resultaten und Konvergenzbegriffen der Stochastik sollen den vorliegenden Text ergänzen und in sich abschließen.

Die vorliegenden Inhalte sind von beiden Autoren wiederholt in Lehrveranstaltungen an der Universität Hamburg und der Technischen Universität Carolo-Wilhelmina zu Braunschweig vermittelt worden. Unsere Vorstellung ist, dass diese Einführung in die Zeitreihenanalyse als Grundlage für eine Vorlesung in den Bereichen Mathematik, Finanz- und Wirtschaftsmathematik und Ökonometrie auch von Dozentinnen und Dozenten verwendet werden kann, deren eigenes Arbeitsgebiet nicht die Zeitreihenanalyse ist. Für eine Vorlesung im Umfang von vier Semesterwochenstunden ist allerdings eine Auswahl zu treffen. Insbesondere wird man aus den Kapiteln 9 bis 15 nur Teile behandeln können. Die im Anschluss an die jeweiligen Kapitel gestellten Aufgaben reichen sicher hin, um Übungen zur einer solchen Veranstaltung zu konzipieren. Ebenso kann der Text auch zur eigenständigen Einarbeitung in das Gebiet oder im Rahmen eines Seminars verwendet werden. Hierzu sind

unseres Erachtens Grundvorlesungen über Analysis und Lineare Algebra sowie Grundkenntnisse der Stochastik und Statistik notwendig.

Unser Dank geht an die vielen Personen, die uns bei der Erstellung dieses Textes unterstützt haben. Herr Jörg Ohle und insbesondere Frau Simone Kohlmann haben die handschriftlichen Manuskripte unermüdlich nach L^AT_EX übertragen. Unsere Mitarbeiter Andreas Dürkes, Christian Hagel, Frank Palkowski und Volker Rehbock haben die geschriebenen Kapitel mit großer Sorgfalt gelesen und zahlreiche, über viele Schreibfehler hinausgehende, Verbesserungsvorschläge gemacht, die wir gerne aufgegriffen haben. Nicht zuletzt haben Studierende uns in unseren Lehrveranstaltungen auf eine Reihe von Punkten hingewiesen, die verbesserungsbedürftig waren. Schließlich gilt unser Dank dem Springer-Verlag und hier insbesondere Herrn Clemens Heine und Frau Lilith Braun, die das Projekt initiiert und zuvorkommend begleitet haben.

Wir hoffen, mit der vorliegenden Einführung dazu beizutragen, dass zukünftige Studierende einen ansprechenden Einstieg in das interessante Gebiet der Zeitreihenanalyse finden.

Braunschweig und Hamburg,
März 2006

Jens-Peter Kreiß
Georg Neuhaus

Inhaltsverzeichnis

1	Einführung	1
1.1	Beispiele für Zeitreihen	4
1.2	Trendschätzung	9
1.3	Schätzung saisonaler Anteile in Zeitreihen	12
	Aufgaben	14
2	Stationarität und grundlegende Modelle der Zeitreihenanalyse	17
2.1	Stationarität von Zeitreihen	17
2.2	Grundlegende stationäre Zeitreihenmodelle	22
2.3	Empirische Autokovarianzen und Autokorrelationen	31
2.4	Gaußsche Zeitreihen	38
2.5	Die partielle Autokorrelation	39
	Aufgaben	43
3	Die Autokovarianz und die Autokorrelation	47
3.1	Grundlegende Eigenschaften	47
3.2	Spektralmaß und Spektraldichte	49
	Aufgaben	63
4	Lineare Vorhersage bei endlicher Vergangenheit	65
4.1	Die rekursive Gram–Schmidt–Orthogonalisierung	66
4.2	Die Levinson–Rekursion	69
	Aufgaben	72
5	Der Spektralsatz für stationäre Zeitreihen	75
5.1	Die Spektraldarstellung zyklischer Zeitreihen	75
5.2	Maße mit orthogonalen Werten und ein stochastisches Integral	78
5.3	Der Spektralsatz	83
5.4	Eine Substitutionsregel für stochastische Integrale	87
	Aufgaben	88

6	Filterung stationärer Zeitreihen	91
6.1	Grundbegriffe und einfache Eigenschaften von Filtern	91
6.2	Spezielle Filter	95
6.3	Zweiseitige MA-Reihen	106
	Aufgaben	106
7	ARMA-Modelle	109
7.1	Definition und Existenz von ARMA-Reihen	109
7.2	Kausalität und Invertibilität von ARMA-Reihen	116
7.3	Lineare L_1 -Filter	119
	Aufgaben	120
8	Die Autokovarianz und Autokorrelation von ARMA-Reihen im reellen Fall	123
8.1	Die Berechnung der Autokovarianzen von ARMA-Reihen aus der MA-Darstellung	123
8.2	Die Differenzengleichung für die Koeffizienten der MA-Darstellung	126
8.3	Die Differenzengleichung für die Autokovarianzen und die Yule-Walker-Gleichungen bei AR-Reihen	128
8.4	Identifizierbarkeit der Parameter von ARMA-Zeitreihen	133
	Aufgaben	136
9	Deterministische und rein nicht-deterministische Zeitreihen	139
9.1	Die Wold-Zerlegung	140
9.2	Approximation durch AR- und MA-Reihen	146
	Aufgaben	149
10	Asymptotische Eigenschaften von Schätzverfahren in linearen Zeitreihenmodellen	153
10.1	Einfache asymptotische Eigenschaften des Stichprobenmittels und der Stichprobenautokovarianz	154
10.2	Schwache Abhängigkeit	159
10.3	Ein zentraler Grenzwertsatz für schwach abhängige Zufallsvariable	161
10.4	Asymptotische Normalität des Stichprobenmittels und der Stichprobenautokovarianz	169
	Aufgaben	180
11	Parameterschätzung in ARMA-Modellen	183
11.1	Parameterschätzung für autoregressive Zeitreihen	184
11.2	Maximum-Likelihood Schätzer im autoregressiven Modell	193
11.3	Parameterschätzung in autoregressiven Modellen mit wachsender Ordnung	201

11.4 Parameterschätzung für ARMA–Zeitreihen	212
Aufgaben	224
12 Schätzen im Spektralbereich	227
12.1 Parametrische Spektraldichteschätzung	230
12.2 Das Periodogramm	232
12.3 Eigenschaften des Periodogramms	235
12.4 Lag–Window–Schätzer der Spektraldichte	243
12.5 Das geglättete Periodogramm	255
12.6 Konfidenzintervalle für die Spektraldichte	259
12.7 Das integrierte Periodogramm	262
Aufgaben	270
13 Modellierung mit ARMA–Zeitreihen	277
13.1 ARIMA–Zeitreihen	278
13.2 Ordnungswahl in ARMA–Zeitreihen	279
13.3 Threshold Zeitreihenmodelle	292
Aufgaben	292
14 Grundlagen finanzieller Zeitreihen	295
14.1 GARCH–Modelle	298
14.2 Parameterschätzung in GARCH–Modellen	309
14.3 Anwendung der GARCH–Methodik	318
Aufgaben	322
15 Grundlagen multivariater Zeitreihen	325
15.1 Multivariate Spektraltheorie	328
15.2 Multivariate Filter	335
15.3 Der quadratische Kohärenzkoeffizient und verwandte Größen ..	337
15.4 Schätzer der Spektraldichtematrix	343
15.5 Multivariate ARMA–Reihen	344
15.6 Schätzung des Mittelwertvektors und der Autokovarianzmatrix einer multivariaten Zeitreihe	350
15.7 Lineare Vorhersage bei multivariaten Zeitreihen	354
15.8 Zustandsraummodelle	355
15.9 Der Kalman–Filter zur linearen Vorhersage	358
Aufgaben	361
A Anhang	363
A.1 Einige nützliche Formeln	363
A.2 Integration komplexer Funktionen	364
A.3 Elementare Hilbertraum Theorie	366
A.4 Lösungen einer homogenen Differenzengleichung	372
A.5 Konvergenzbegriffe in der Stochastik	374
A.6 Die Moore–Penrose–Inverse	380

Literaturverzeichnis	381
Index	385

Einführung

Zeitreihen begegnen uns in den verschiedensten Bereichen des täglichen Lebens. Überall werden uns Grafiken mit zeitlichen Verläufen von Anzahlen präsentiert. Prominente Beispiele sind sicherlich die monatlichen Arbeitslosenzahlen in Deutschland oder der fast kontinuierlich dokumentierte Stand des Deutschen Aktienindex DAX. Aber auch Verkaufszahlen, Bevölkerungs- oder Einkommensentwicklungen sowie Unfallzahlen im Straßenverkehr werden in sogenannten Zeitreihen übersichtlich und informativ dargestellt. So können sich viele Menschen in unserer Gesellschaft unter Zeitreihen und deren Analyse etwas vorstellen. Dies ist bei weitem nicht für jedes Wissenschaftsgebiet der Fall. Gleichzeitig reicht die Zeitreihenanalyse in ganz verschiedene Gebiete der Wissenschaft hinein, etwa in die Wirtschaftswissenschaften, die Natur- und Ingenieurwissenschaften und nicht zuletzt in die Mathematik. Wir nähern uns den Zeitreihen von einer mathematisch stochastischen Seite, werden aber die vielfältigen anwendungsbezogenen Fragestellungen im Auge behalten und versuchen, auch dem nicht originär aus dem mathematischen Umfeld stammenden Leser grundlegende konzeptionelle Ansätze nahe zu bringen.

Zunächst ist die Frage zu beantworten, was eine Zeitreihe eigentlich ist. Grundsätzlich bezeichnen wir mit einer Zeitreihe eine Folge von oftmals reellen Größen, die in einer diskreten (und eben nicht kontinuierlichen) Zeit anfallen, aufgezeichnet oder verarbeitet werden sollen. Die monatlichen Arbeitslosenzahlen in Deutschland (vgl. Abbildung 1.1), die von der Bundesagentur für Arbeit in Nürnberg bereitgestellt werden, sind ein allgegenwärtiges Beispiel. Wir gehen davon aus, dass die zu den verschiedenen Zeitpunkten anstehenden Größen zufallsabhängig sind und deshalb sinnvoll als Zufallsvariable aufgefasst werden können. Falls die Zeitabstände zwischen den einzelnen Zufallsvariablen zusätzlich äquidistant sind, also ohne Einschränkung (in einem geeignetem Messsystem) den Abstand eins besitzen, gelangen wir zu dem grundlegenden Modell

$$X = (X_t : t \in T), \quad T = \mathbb{N} \text{ oder } \mathbb{Z}, \quad (1.1)$$

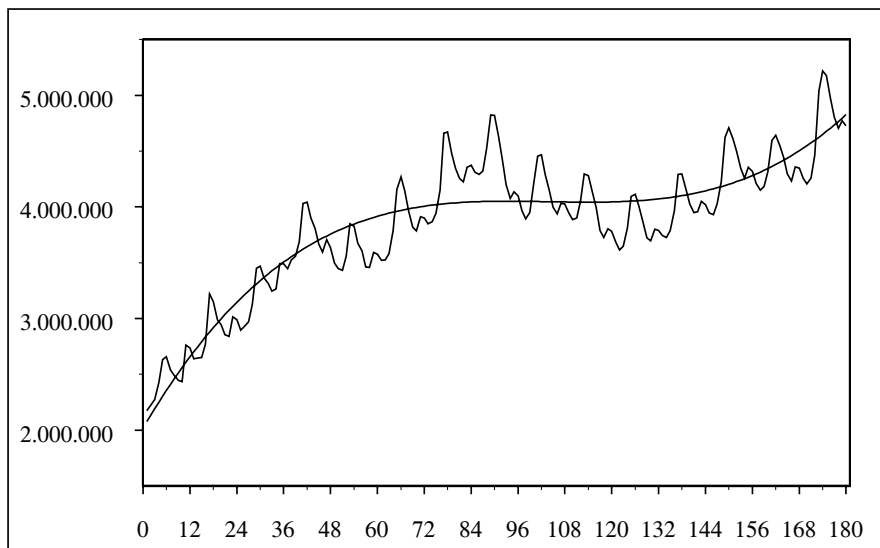


Abb. 1.1 Monatliche Arbeitslosenzahlen in der Bundesrepublik Deutschland von September 1990 bis August 2005 mit einem geschätzten kubischen Trend

mit (reellwertigen) Zufallsvariablen $X_t, t \in T$, auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Obwohl die tatsächlichen Realisierungen einer Zeitreihe fast immer zu einem gewissen Zeitpunkt (sagen wir $t = 1$) beginnen, benutzt man in der Zeitreihenanalyse trotzdem sehr häufig die Modellierung über die Zeitmenge $T = \mathbb{Z}$, also mit unendlicher Vergangenheit, um Resultate in technisch einfacherer Weise formulieren und beweisen zu können. Die Zeitmenge $T = \mathbb{N}$ würde immer wieder eine gesonderte Behandlung der Startwerte erfordern, der wir aus dem Wege gehen wollen. Natürlich ist gerade die Abhängigkeitsstruktur der Zufallsvariablen X_t zu verschiedenen Zeitpunkten für die Zeitreihenanalyse besonders interessant. Deshalb spielt der für die Mathematische Statistik sehr wichtige Fall der stochastisch unabhängigen und identisch verteilten (independent and identically distributed – i.i.d. –) Zufallsvariablen im Bereich der Zeitreihenanalyse nur die Rolle eines – allerdings grundlegenden – Eingangsmodells im Rahmen des sogenannten *weißen Rauschens* (vgl. Beispiel 2.5).

Im folgenden Abschnitt betrachten wir tatsächliche Zeitreihen aus unserem gesellschaftlichen Umfeld, aber auch einige klassische Datensätze der Zeitreihenanalyse, wie die historischen Wasserstände des Nils oder die Canadian Lynx Data. Mit der im Rahmen der historischen Entwicklung der Zeitreihenanalyse bedeutenden Zeitreihe der Sonnenfleckenzahlen (sunspot numbers) befassen wir uns in Kapitel 2 gesondert. Weiter behandeln wir in zwei Abschnitten die Schätzung und Bereinigung von sogenannten Trend- und Saisonkomponenten

bei Zeitreihen. Die stärker mathematisch geprägte Analyse von Zeitreihen fußt nämlich zum überwiegenden Teil auf Zeitreihen, denen ein über die Zeit stabiler Generierungsmechanismus zugrunde liegt und die dadurch ein sogenanntes stationäres Verhalten zeigen. Hierauf werden wir dann in Kapitel 2 vertieft eingehen. Die dort vorgestellten Modelle und entwickelten Methoden befassen sich also mit der Abhängigkeitsstruktur der Zufallsgrößen nach vollständiger Eliminierung von etwaigen saisonalen oder Trendeinflüssen.

Obwohl in der praktischen Anwendung derartige Effekte immer nur unvoll-

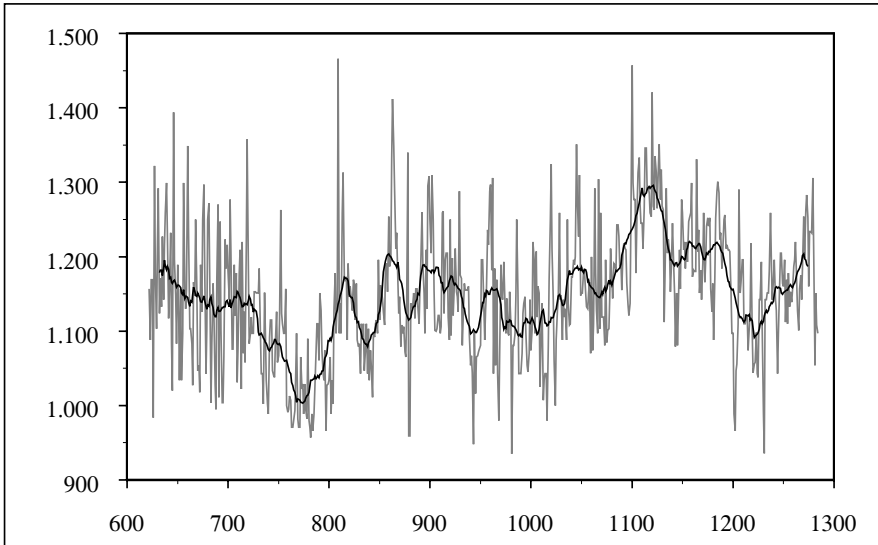


Abb. 1.2 Jährliche minimale Wasserstände des Nils von 622 bis 1284 gemessen am Nilometer auf der Nilinsel ar-Randa bei Kairo mit einem über ein gleitendes Mittel der Länge 21 geschätzten Trend

ständig und höchstens approximativ entfernt werden können, finden die im Verlauf dieses Buches behandelten Methoden trotzdem vielfache Anwendung auf lediglich saison- und trendbereinigte Daten. Dies mag auf den ersten Blick verwundern, aber die vollständige Einbeziehung aller in diesem mehrstufigen Verfahren eingesetzten Methoden würde zu einer äußerst unübersichtlichen und mit technischen Details überladenen Darstellung des Gebietes der Zeitreihenanalyse führen. Zusätzlich ist die stochastische Modellierung grundsätzlich dem mathematisch kaum zu fassenden Risiko eines nicht adäquaten Modells ausgesetzt. Wir folgen in dieser Einführung in die Zeitreihenanalyse dem Vorbild vieler Lehrbücher und basieren die wesentlichen Teile der Analyse auf stationären Zeitreihen, die wir in Definition 2.1 formal einführen werden. Trotzdem behandeln wir an dieser Stelle grundlegende Techniken zur Bereinigung

von Trend- oder Saisoneinflüssen, mit denen Zeitreihenanalytiker vertraut sein sollten.

Ein Ziel der Zeitreihenanalyse ist eine adäquate Beschreibung von in zeitlicher Abfolge auftretenden und voneinander abhängigen Daten. Insbesondere sucht man nach einer angemessenen Modellierung der Abhängigkeitsstruktur, um vertrauenswürdige Prognosen für zukünftige Werte abgeben zu können. Ein weiteres wichtiges Ziel einer mathematisch geprägten Zeitreihenanalyse ist es, ein tieferes Verständnis der verwendeten Modelle und Methodiken in diesem Bereich zu gewinnen.

1.1 Beispiele für Zeitreihen

Im Rahmen der historischen Entwicklung der Zeitreihenanalyse haben sich einige Beispiele für Zeitreihen einen festen Platz erobert. Dazu gehören die sogenannten Sonnenfleckenzahlen (sunspot numbers), denen wir uns in Kapitel 2 ausführlich und mit einigen historisch interessanten Details ergänzt zuwenden werden. Die Aufzeichnung der Sonnenfleckenzahlen geht bis ins frühe 17. Jahrhundert zurück. Wesentlich älter sind die Aufzeichnungen der Ägypter über

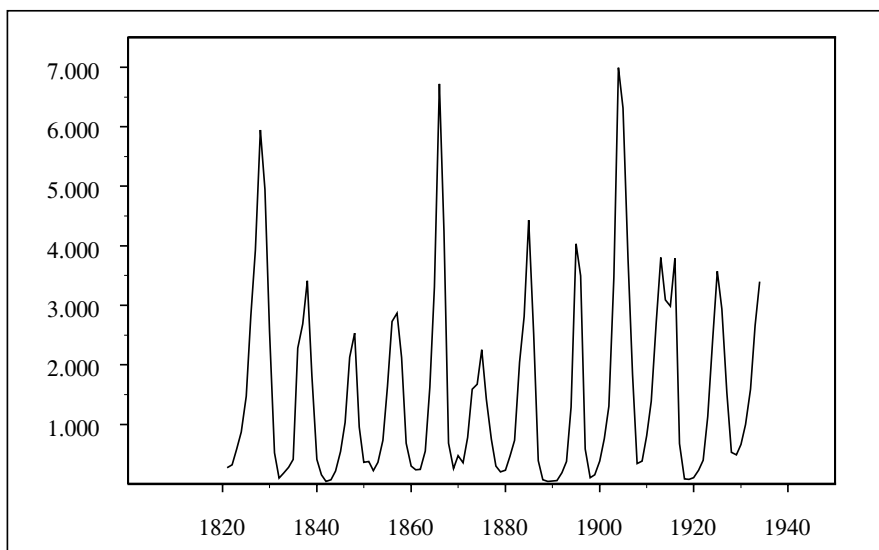


Abb. 1.3 Anzahl jährlich gefangener Luchse in Kanada (MacKenzie River) für den Zeitraum 1821 bis 1934

die Wasserstände des für die ägyptische Landwirtschaft so wichtigen Flusses Nil. Die auf der Nilinsel ar-Randa in der Nähe von Kairo errichteten Nilometer zeichneten jeweils am 3. Juli eines Jahres Wasserstände (sogenannte

jährliche Minima) auf. Diese liegen für den Zeitraum 622–1284 (vgl. Abbildung 1.2) lückenlos vor (siehe Toussoun (1925) oder Beran (1994)). Nicht nur die Erträge sondern auch die Besteuerung der Felder hing vom Nilstand ab, der bis zum Bau des Staudammes bei Assuan (1892–1902) alljährlich gemessen wurde. Dieser Datensatz ist auf der Website <http://lib.stat.cmu.ed> von StatLib erhältlich. Diese Website wird vom Department of Statistics at Carnegie Mellon University gepflegt und ist eine wahre Fundgrube für statistisch interessante Datensätze.

Eine weitere in der Literatur vielfach untersuchte Zeitreihe ist die Anzahl der

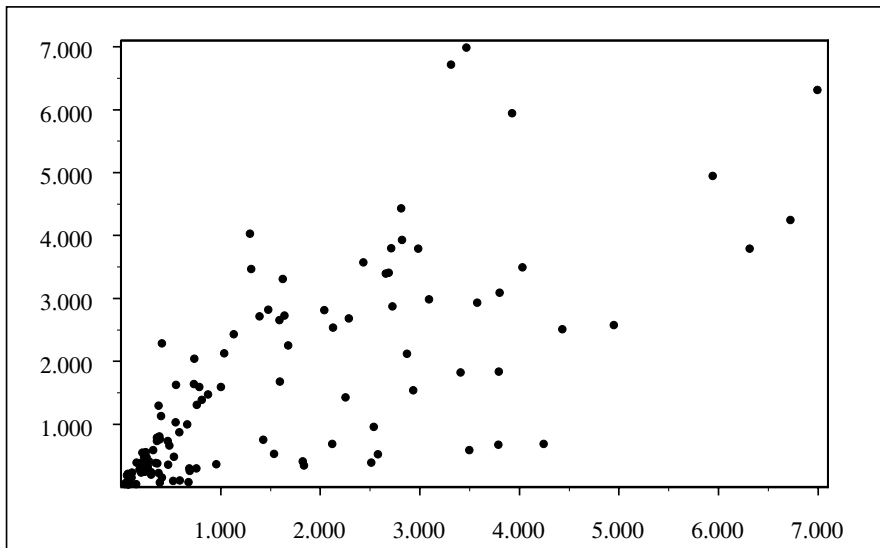


Abb. 1.4 $X_{t-1} - X_t$ -Plot der Luchsdaten aus Abbildung 1.3

in einem bestimmten Gebiet Kanadas (MacKenzie River) gefangenen Luchse (vgl. Campbell/Walker (1977)). Diese Canadian Lynx Data sind in Abbildung 1.3 dargestellt. Für diese Daten gibt Abbildung 1.4 einen sogenannten $X_{t-1} - X_t$ -Plot, in dem die Punkte $(X_{t-1}, X_t), t = 2, \dots, n$, eingetragen sind, wieder. Ein solcher Plot gibt einen ersten Eindruck über die mögliche Abhängigkeitsstruktur zweier benachbarter Beobachtungen. Für die Canadian Lynx Data erkennt man in Abbildung 1.4 einen recht deutlichen positiven Zusammenhang der Beobachtungen aus zwei aufeinander folgenden Jahren.

Tatsächlich stellt die grafische Darstellung und Aufbereitung der Daten in der Regel den ersten Schritt bei der Analyse von Zeitreihen dar.

Als weitere Beispiele für ökonomische Zeitreihen seien neben den Arbeitslosenzahlen (Abbildung 1.1), die Reihe der täglichen Schlusskurse des DAX (Abbildung 1.5) für den Zeitraum 1. Januar 1990 bis 30. September 2005 sowie

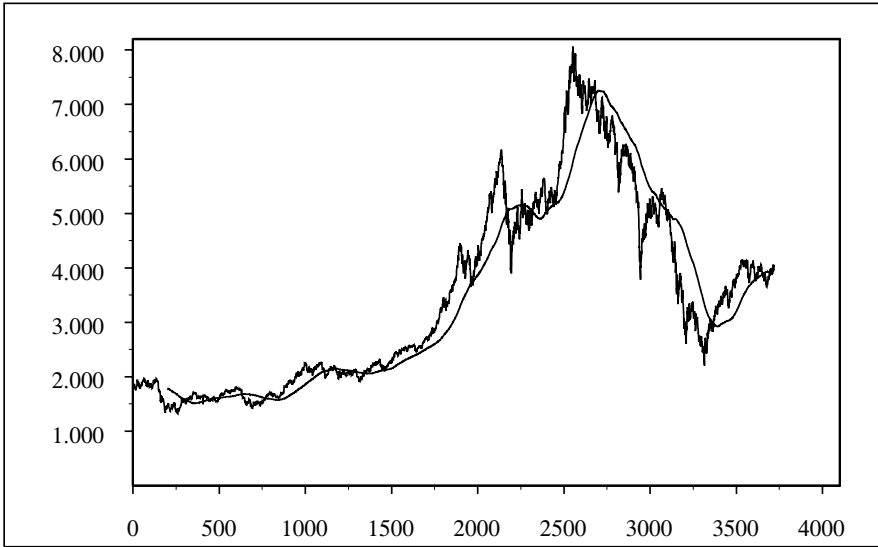


Abb. 1.5 Tägliche Schlusskurse des Deutschen Aktienindex DAX für den Zeitraum Januar 1990 bis September 2005 zusammen mit einem gleitenden Mittel über die jeweils 200 vorangegangenen Werte (sogenannter 200 Tage Durchschnitt)

der tägliche US-Dollar/DM –Wechselkurs für den Zeitraum vom 1. Juli 1990 (Beginn der Währungsunion mit der ehemaligen DDR) bis zur Einführung des Euro bargeldes am 1. Januar 2002 (Abbildung 1.6) erwähnt. Diese Beispiele für sogenannte Finanzzeitreihen werden wir im Kapitel 14 erneut aufgreifen. Wesentlich für die dortige Untersuchung sind die sogenannten Renditen, d.h. die prozentualen Veränderungen des Kurses von heute gegenüber dem von gestern. Für die Zeitreihe des DAX sind derartige Renditen in Abbildung 1.7 zu sehen. Bei genauem Hinsehen erkennt man in der Zeitreihe des DAX (vgl. Abbildung 1.5) gleich zu Beginn der Datenreihe ein abruptes Absinken des Kurses. Dieses Ereignis koinzidiert deutlich mit dem Beginn des Golfkrieges zu Beginn des Jahres 1991. Kurz danach finden wir einen singulären fast zehnprozentigen Verlust des DAX, der auf den Tag des Bekanntwerden des seinerzeitigen Putschversuches in der ehemaligen Sowjetunion gegen Michail Gorbatschow fällt und der mit bloßem Auge in dieser Darstellung nur sehr schwer zu erkennen ist. Beide Ereignisse sind sehr viel deutlicher in den Renditewerten des DAX (vgl. Abbildung 1.7) zu erkennen. Zu Beginn dieser Datenreihe beobachten wir eine erhöhte Schwankungsbreite (sogenannte Volatilität) des Kurses, die einem nervösen Markt zu eigen ist, wie er in der Zeit des Golfkrieges vorgelegen hat. Im weiteren Verlauf der Zeitreihe der DAX-Renditen finden wir immer wieder ruhigere und auch volatilere Marktphasen (im Wesentlichen charakterisiert durch die beobachtete Schwankungsbreite), die für

finanzielle Zeitreihen sehr typisch ist. Insbesondere ist der singuläre, fast zehnprozentige Verlust an nur einem einzigen Tag zu Beginn in der Zeitreihe der Renditen sehr deutlich zu erkennen.

Diesen kleinen Reigen an Beispielen wollen wir mit der monatlichen Zeitreihe der Anzahl der durch Verkehrsunfälle getöteten Personen in Deutschland für

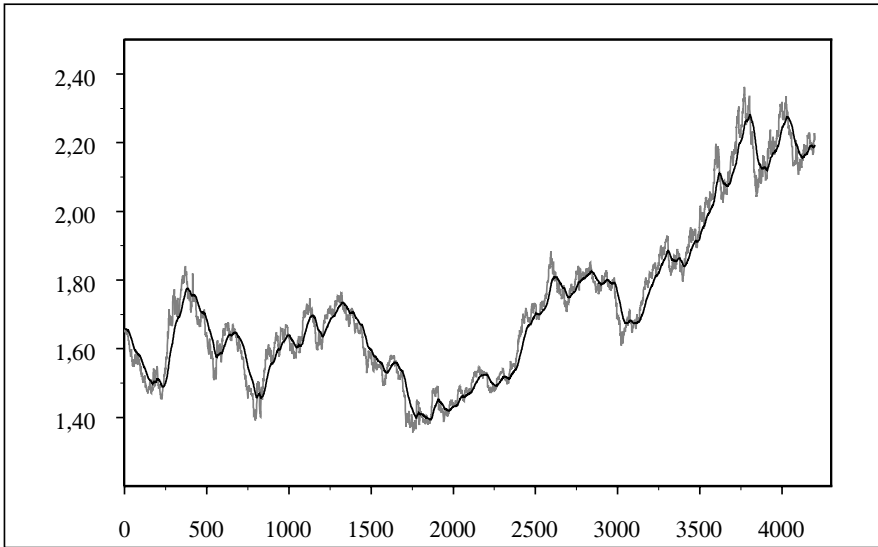


Abb. 1.6 Täglicher US-Dollar/DM-Wechselkurs für den Zeitraum 1. Juli 1990 bis 31. 12. 2001 mit einer Glättung gemäß der Methode des exponentiellen Glättens mit Parameter 0.97

den Zeitraum seit Anfang 1991 abschließen (vgl. Abbildung 1.8). Die Daten stammen vom Statistischen Bundesamt, Wiesbaden (Genesis-Online, Code 46241). Diese große Datenbank (<http://www.statistik-portal.de>) enthält eine Vielzahl von interessanten Zeitreihen aus den verschiedensten Bereichen in Deutschland und ist eine lohnende Quelle bei der Suche nach relevanten Datensätzen. Man erkennt in der Zeitreihe der Verkehrsunfallzahlen mit tödlichem Ausgang den deutlichen, auffallend linearen und fallenden Trend sowie eine klare saisonale Komponente, auf deren Schätzung wir noch zu sprechen kommen werden.

In vielen weiteren realen Zeitreihen, wie etwa den Arbeitslosenzahlen, entdeckt man ebenfalls einen zeitlichen Trend und/oder saisonale Einflüsse. In den folgenden beiden Abschnitten wollen wir uns um diese Fragestellung kümmern.

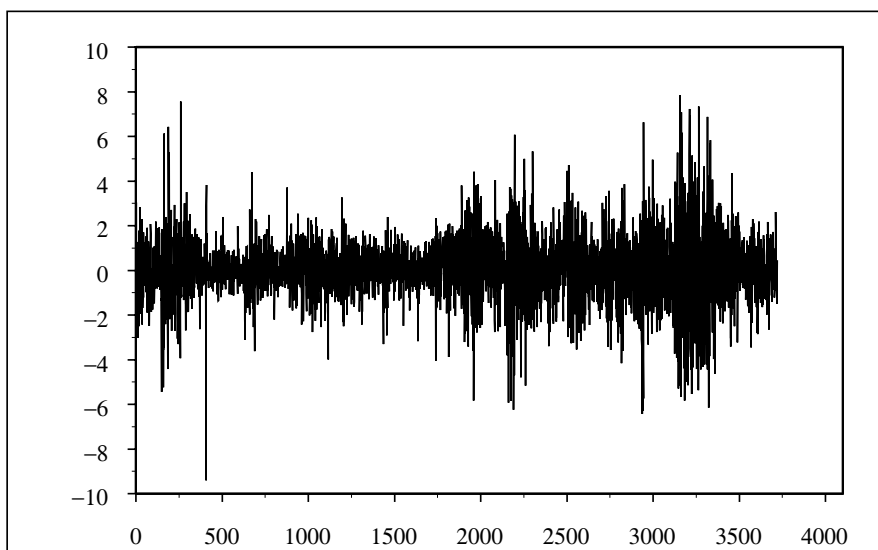


Abb. 1.7 Tägliche Renditen des DAX aus Abbildung 1.5

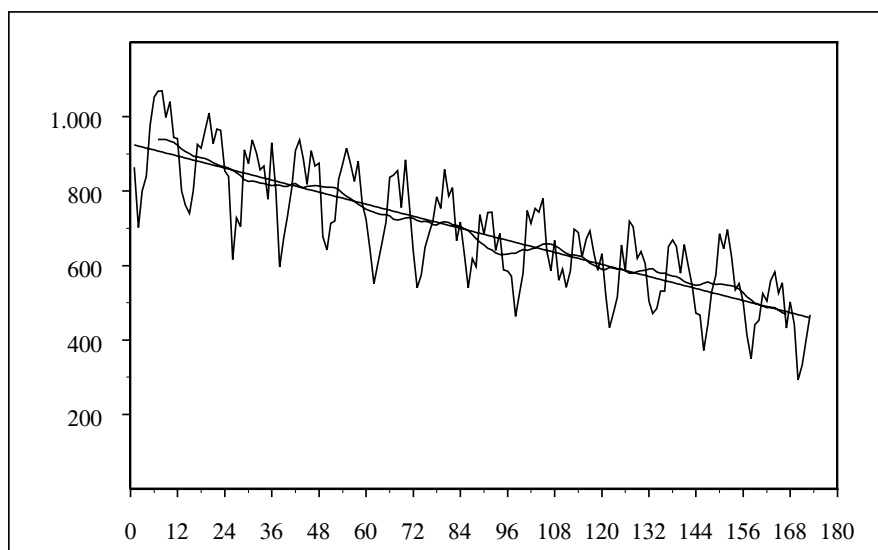


Abb. 1.8 Monatliche Anzahl der durch Verkehrsunfälle in Deutschland getöteten Personen von Januar 1991 bis Mai 2005 zusammen mit einem geschätzten linearen Trend und einem gemäß des gleitenden Mittels über 12 Monate geschätzten Trend

1.2 Trendschätzung

Um Trendkomponenten in Zeitreihendaten zu schätzen und später dann auch eliminieren zu können, bieten sich zunächst Methoden an, die aus dem Bereich der linearen Regression bekannt sind.

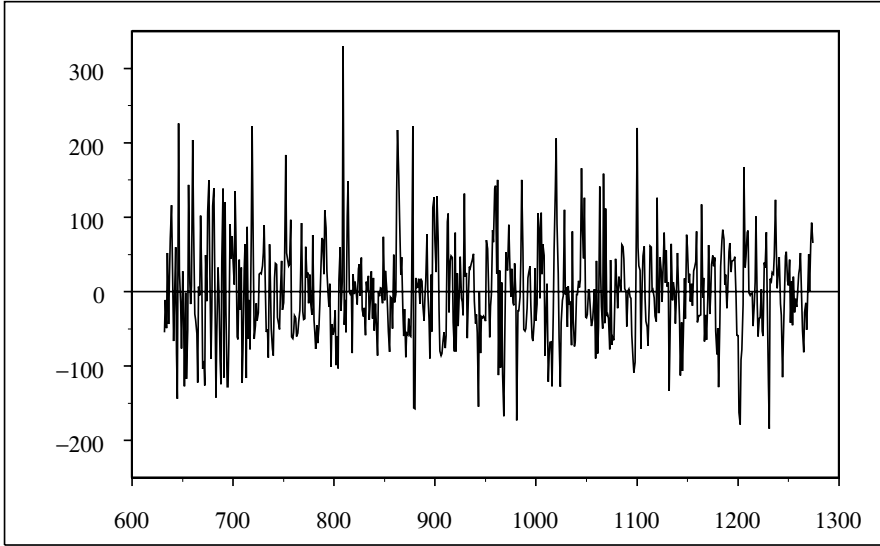


Abb. 1.9 Minimale Nilwasserstände (vgl. Abbildung 1.2) bereinigt um den dort angegebenen Trend

Beispiel 1.1 (Lineare Trendmodelle)

Für gegebene reelle Funktionen m_0, m_1, \dots, m_k auf \mathbb{Z} bezeichnen wir

$$m(t) = \sum_{i=0}^k a_i m_i(t), \quad \forall t \in \mathbb{Z}, \quad (1.2)$$

für $a_0, \dots, a_k \in \mathbb{R}$, als lineares Trendmodell. Für $k = 1$, $m_0(t) = 1$ und $m_1(t) = t$ erhalten wir den linearen Trend $a_0 + a_1 \cdot t$, sowie für beliebiges k mit $m_i(t) = t^i$, $i = 0, \dots, k$, einen polynomialen Trend.

Die Schätzung in einem linearen Trendmodell (1.2) führt man in Analogie zur Regression etwa mit der Methode der kleinsten Quadrate durch, indem man a_0, \dots, a_k so wählt, dass der Ausdruck

$$\sum_{t=1}^n \left(X_t - \sum_{i=0}^k a_i m_i(t) \right)^2 \quad (1.3)$$

minimal wird. Dabei bezeichnen X_1, \dots, X_n die gegebenen Beobachtungen der zugrunde liegenden Zeitreihe. Unterstellen wir für die Arbeitslosenzahlen in Abbildung 1.1 und für den dargestellten Zeitraum einen polynomialen Trend der Ordnung $k = 3$, so erhalten wir den geschätzten Trend

$$\hat{m}(t) = 2.018.898 + 59.815,60 \cdot t - 582,12 \cdot t^2 + 1,87 \cdot t^3, \quad (1.4)$$

der in Abbildung 1.1 mit dargestellt ist. \square

Aus der Abbildung 1.1 erkennt man, dass ein Trend in gewisser Weise eine Glättung der beobachteten Zeitreihe darstellt. Deshalb kommen auch aus der Statistik bekannte Glättungsmethoden zur Trendschätzung in Frage.

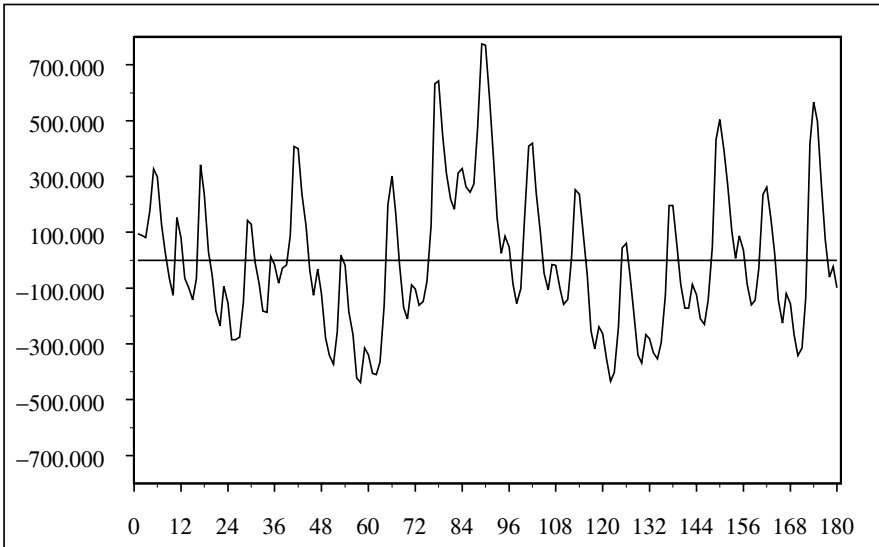


Abb. 1.10 Trendbereinigte Arbeitslosenzahlen (vgl. mit Abbildung 1.1)

Beispiel 1.2 (Gleitendes Mittel)

Für ein gegebenes $q \in \mathbb{N}$ und reelle Gewichte w_j , $j = -q, \dots, q$, heißt

$$\hat{m}(t) = \sum_{j=-q}^q w_j X_{t+j}, \quad t = q+1, \dots, n-q, \quad (1.5)$$

ein Trendschätzer auf Basis eines gleitendes Mittels der Länge $2q+1$ für die Beobachtungen X_1, \dots, X_n , wobei n genügend groß sein muss, damit (1.5) Sinn macht. Wichtig ist die Einhaltung der Normierungsbedingung $\sum_{j=-q}^q w_j = 1$,

die z.B. erforderlich ist, damit das gleitende Mittel eine konstante Folge $X_t = c \forall t$ unverändert lässt.

Im Falle $w_j = w_{-j}$, $j = 1, \dots, q$, heißt das gleitende Mittel *symmetrisch* und im Falle $w_j = 0$ für $j \geq 1$ heißt es *einseitig*. Das einseitig gleitende Mittel erlaubt eine Bestimmung der Trendschätzung bis hin zum Zeitpunkt der aktuellen Beobachtung und findet insbesondere bei Börsenkursen Anwendung (vgl. Abbildung 1.5 für den DAX mit $q = 200$).

Typische Wahlen für die Gewichte w_j sind $w_j = 1/(2q + 1)$, $j = -q, \dots, q$. Mit $q_{\text{Nil}} = 10$ haben wir exakt diese Trendschätzung eingesetzt, um die in Abbildung 1.2 wiedergegebene Situationen für die Nildaten zu erhalten.

Im Falle des einseitigen gleitenden Mittels ist insbesondere die Situation $q = +\infty$ und $w_{-j} = \alpha^j \cdot (1 - \alpha)$, $j = 0, 1, 2, \dots$, ($0 < \alpha < 1$) unter dem Namen *exponentielles Glätten* bekannt, siehe auch Aufgabe 4.1 und Beispiel 6.9. Es gilt

$$\hat{m}(t) = (1 - \alpha) \sum_{j=0}^{\infty} \alpha^j X_{t-j} = (1 - \alpha) \cdot X_t + \alpha \cdot \hat{m}(t - 1). \quad (1.6)$$

Um die Konvergenz der Reihe wollen wir uns erst später kümmern. Sie gilt tatsächlich sowohl fast sicher als auch im quadratischen Mittel, vgl. Beispiel 2.13. Formal müssen für die Berechnung von $\hat{m}(t)$ unendlich viele Beobachtungen der Vergangenheit vorhanden sein. Dies umgeht man, indem man den Startwert $\hat{m}(1) = X_1$ festsetzt und die weiteren Werte rekursiv auf Basis der Daten X_1, \dots, X_n gemäß der zweiten Gleichung in (1.6) berechnet. Mit $\alpha = 0,97$ haben wir diesen Trendschätzer mittels exponentiellem Glätten für die Wechselkursdaten in Abbildung 1.6 bestimmt. \square

Selbstverständlich können weitere aus dem Bereich des Glättens bekannte Methodiken, wie etwa Splines-Techniken, zum Trendschätzen herangezogen werden. Die in Beispiel 1.2 vorgestellte Methode des gleitenden Mittels korrespondiert zu sogenannten Kernglättungsmethoden der Statistik, wobei aufgrund der diskreten Zeit die Kernfunktion durch die Gewichte w_j festgelegt wird und die Länge q des gleitenden Mittels der Bandweite eines Kernschätzers entspricht. Je größer q gewählt wird, desto stärker werden die Ausgangsdaten geglättet.

Betrachten wir nun etwa die Nildaten nach Trendbereinigung (vgl. Abbildung 1.9), so erkennen wir eine gewisse Stabilität der Datenreihe über die Zeit. Hingegen zeigen sowohl die Arbeitslosenzahlen wie die Verkehrsunfallzahlen nach Subtraktion des geschätzten Trends (vgl. Abbildungen 1.10 und 1.11) noch deutliche saisonale Schwankungen, die es zu bereinigen gilt, bevor diese Datensätze den in den folgenden Kapiteln behandelten Methoden für stationäre Zeitreihen zugänglich werden.

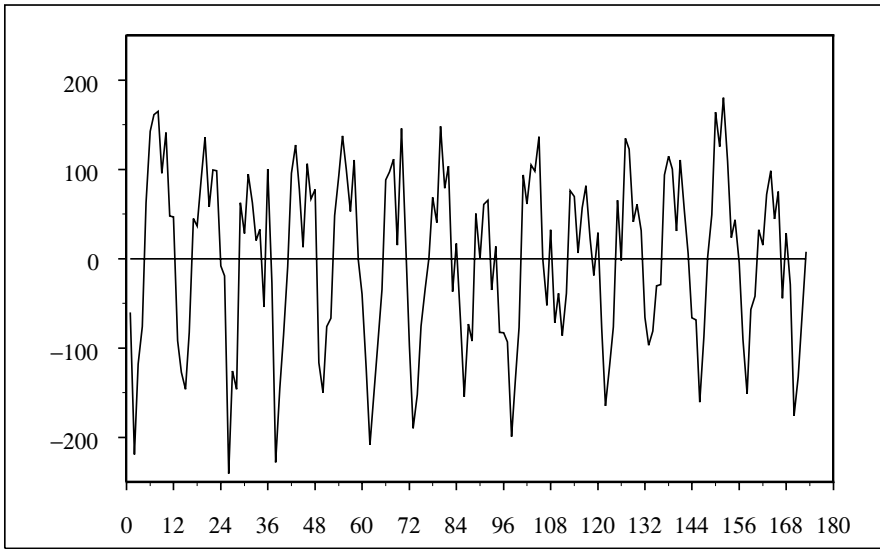


Abb. 1.11 Die trendbereinigte Anzahl von Todesopfern in Verkehrsunfällen in Deutschland (vgl. mit Abbildung 1.8)

1.3 Schätzung saisonaler Anteile in Zeitreihen

Viele Zeitreihendaten unterliegen saisonal bedingten Schwankungen. So hören wir oft, dass ein Rückgang oder auch eine Zunahme von Arbeitslosenzahlen in gewissen Monaten lediglich auf saisonale Effekte zurückzuführen ist. Das Verkehrsunfallrisiko ist verständlicherweise verknüpft mit Straßenzuständen (Eis, Schnee, Regen usw.), dem generellen Fahrverhalten oder auch mit den Lichtverhältnissen, so dass auch bei diesen Daten nach Trendbereinigung verbleibende saisonale Anteile nicht verwundern (vgl. Abbildung 1.11). Typischerweise ist die Periodenlänge d eines saisonalen Anteils bekannt. Bei monatlichen ökonomischen Daten wird man die Periodenlänge $d = 12$ unterstellen, bei Quartalsdaten hingegen die Periodenlänge $d = 4$. Gehen wir also für das folgende Beispiel von einer festen und bekannten Periodenlänge $d \in \mathbb{N}$ aus.

Beispiel 1.3 (Saisonbereinigung)

Entweder sei die zugrunde liegende Zeitreihe bereits trendbereinigt oder sie enthalte überhaupt keinen Trend. Auf Basis von trendbereinigten oder trendfreien Daten X_1, \dots, X_n gehen wir dann wie folgt vor, um eine saisonale Komponente mit der Periodenlänge d zu schätzen. Der Einfachheit halber nehmen wir an, dass m komplette Perioden der Länge d beobachtet wurden, also dass $n = m \cdot d$ gilt. Für $r = 1, \dots, d$ definieren wir

$$\tilde{s}_r = \frac{1}{m} \sum_{j=1}^m X_{r+(j-1) \cdot d}. \quad (1.7)$$

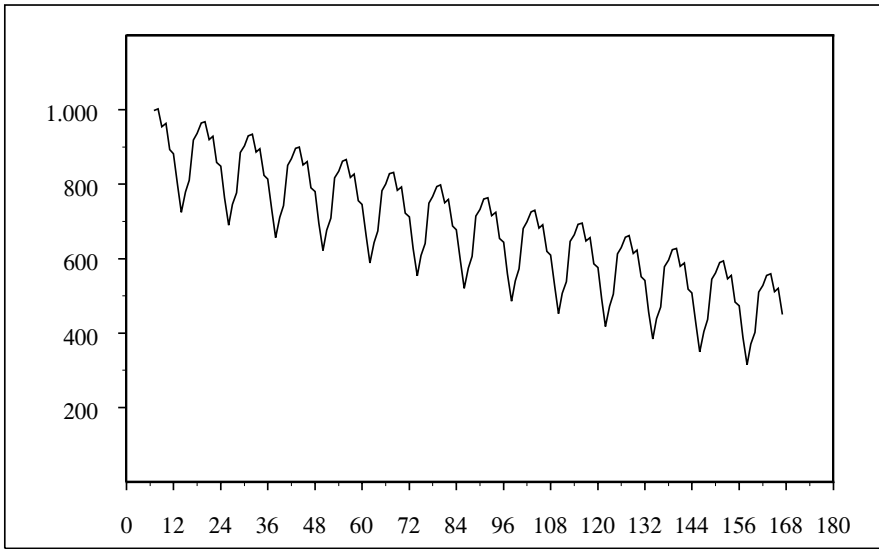


Abb. 1.12 Geschätzter saisonaler Anteil der Unfallzahlen zzgl. eines nach Saisonbereinigung erneut geschätzten linearen Trends (vgl. mit Abbildung 1.8)

Entweder bezeichnet man den saisonalen Anteil direkt mit $\tilde{s}_1, \dots, \tilde{s}_d$ und setzt gemäß $\tilde{s}_t = \tilde{s}_{t-d}$ für $t > d$ periodisch fort, oder man erreicht durch

$$\hat{s}_r = \tilde{s}_r - \frac{1}{d} \sum_{j=1}^d \tilde{s}_j, \quad r = 1, \dots, d, \quad (1.8)$$

dass die saisonalen Anteile $\hat{s}_1, \dots, \hat{s}_d$ stets $\sum_{r=1}^d \hat{s}_r = 0$ erfüllen, und setzt diese dann ebenfalls periodisch fort.

Konkret erhalten wir für die ursprünglichen Unfallzahlen U_t gemäß Abbildung 1.8 nach Bereinigung des durch das folgende gleitende Mittel

$$\hat{m}_t = \frac{1}{12} (0.5 U_{t-6} + U_{t-5} + \dots + U_{t+5} + 0.5 U_{t+6}), \quad t = 7, \dots, n-6, \quad (1.9)$$

geschätzten Trends, die in Abbildung 1.11 angegebenen trendbereinigten Unfallzahlen. Dieses gleitende Mittel ist auf die Trendschätzung bei Perioden der Länge 12 zugeschnitten, da es von diesen unbeeinflusst bleibt. Für die so trendbereinigten Unfallzahlen erhalten wir gemäß (1.8) die saisonalen Anteile für die Monate Januar bis Dezember wie sie in Tabelle 1.1 aufgeführt sind. Man erkennt, dass insbesondere in den Monaten Mai bis Oktober überdurchschnittlich viele tödlich verletzte Personen in Verkehrsunfällen zu beklagen sind. Die Wintermonate weisen dagegen eher unterdurchschnittliche Werte auf. Nun kann man entweder von den trendbereinigten Unfallzahlen

Tabelle 1.1 Geschätzte saisonale Komponente der Verkehrsunfallzahlen

Monat	1	2	3	4	5	6	7	8	9	10	11	12
\hat{s}_k	-90	-160	-102	-67	44	64	94	101	56	68	0	-8

noch jeweils die zugehörigen saisonalen Anteile subtrahieren und erhält

$$U_t - \hat{m}_t - \hat{s}_t, \quad t = 7, \dots, n - 6, \quad (\hat{s}_t = \hat{s}_{t-12} \text{ für } t > 12), \quad (1.10)$$

oder man kann alternativ auf Basis der lediglich saisonbereinigten Zeitreihe $U_t - \hat{s}_t$, $t = 1, \dots, n$, erneut eine Schätzung der Trendkomponente durchführen und diesen neu geschätzten Trend zur Bereinigung verwenden. Im Grundsatz ändert sich an der Darstellung (1.10) dadurch nichts. Wir sind diesen zweiten Weg gegangen und stellen in Abbildung 1.12 die geschätzte Saisonkomponente für die Verkehrsunfallzahlen zuzüglich eines nach Saisonbereinigung erneut geschätzten linearen Trends dar. Abbildung 1.13 gibt schließlich die auf diese Weise trend- und saisonbereinigten Unfallzahlen wieder. \square

Die Formel (1.10) legt nahe, für die Modellierung von Zeitreihen mit Saison- und Trendkomponente den folgenden Ansatz

$$X_t = Z_t + m_t + s_t, \quad t = 1, \dots, n, \quad (1.11)$$

zu wählen. Dabei stellen (m_t) bzw. (s_t) die deterministischen – also nicht vom Zufall beeinflussten – Trend- bzw. saisonalen Anteile dar. (Z_t) ist eine Zeitreihe mit stationärem Verhalten, wie wir sie im weiteren Verlauf dieses Buches betrachten wollen. Die eigentlich zugrunde liegende Zeitreihe (X_t) entsteht durch additive Überlagerung dieser drei Anteile.

Wir wollen die Überlegungen zu Trend- und Saisonanteilen hiermit abschließen und uns ab dem nächsten Kapitel ausschließlich mit sogenannten stationären Zeitreihen – dann allerdings in mathematisch weit rigorosere Form – auseinandersetzen.

Aufgaben

Aufgabe 1.1 Gegeben sei eine Zeitreihe $X = (X_t : t \in \mathbb{N})$, die geglättet werden soll. Zur Berechnung der geglätteten Werte \hat{X}_t , $t \geq m + 1$, für gegebenes $m \in \mathbb{N}$ wird in den Punkten (r, X_r) , $r = t - m, \dots, t + m$, das Polynom p -ten Grades ($2m + 1 > p$)

$$P(r) = \sum_{i=0}^p a_i (r - t)^i$$

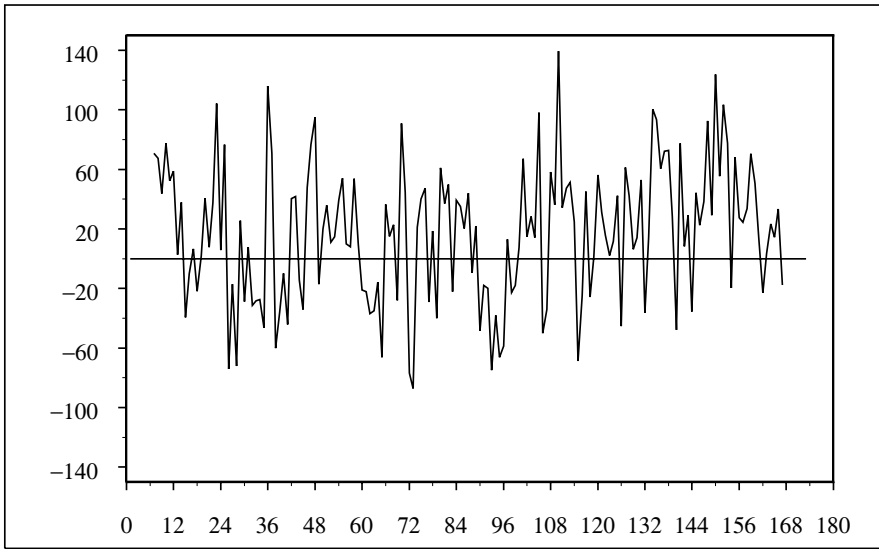


Abb. 1.13 Trend- und saisonbereinigte Anzahl der tödlich verletzten Unfallopfer in Deutschland von 1991 bis 2005 (vgl. mit Abbildung 1.8)

nach der Methode der kleinsten Quadrate angepasst, d.h. die Koeffizienten $a_i, i = 0, \dots, p$, werden bestimmt durch Minimierung der Zielfunktion

$$Z(a_0, \dots, a_p) := \sum_{s=-m}^{+m} (X_{t+s} - P(t+s))^2.$$

Man setzt dann $\hat{X}_t = P(t) = a_0$.

a) Durch Differentiation leite man ein möglichst einfaches Gleichungssystem zur Bestimmung des interessierenden Wertes a_0 her. (In diesem System treten nur die Koeffizienten mit geradem Index i auf.)

b) Aus dem System in a) kann man herleiten, dass a_0 eine Linearkombination der Terme $\sum_{s=-m}^{+m} X_{t+s} s^{2j}$ ist, wobei die Koeffizienten c_j dieser Terme nicht von den X_{t+s} abhängen. Hieraus folgere man, dass \hat{X}_t darstellbar ist als gleitender Durchschnitt

$$\hat{X}_t = \sum_{s=-m}^{+m} X_{t+s} \alpha_s \quad (1.12)$$

mit Gewichten α_s , die nicht von X_{t+s} abhängen und

$$\alpha_s = \alpha_{-s}, \quad \sum_{s=-m}^{+m} \alpha_s = 1 \quad (1.13)$$

erfüllen.

c) Speziell für $m = 3, p = 1$ und $p = 3$ bestimme man die Koeffizienten α_s in

b) und glätte für $4 \leq t \leq 10$ die Zeitreihe

$$(X_1, \dots, X_{13}) = (0.7, 2.4, 2.0, 1.8, 2.0, 2.2, 3.2, 5.4, 5.2, 4.8, 4.6, 5.2, 6.4)$$

Hinweis: Die zweite Aussage in (1.13) kann durch Wahl der speziellen Zeitreihe $X_t = c > 0 \forall t$ bewiesen werden.

Aufgabe 1.2 Man betrachte den Trendschätzer

$$\hat{m}(t) = \frac{1}{9} (-X_{t-2} + 4X_{t-1} + 3X_t + 4X_{t+1} - X_{t+2})$$

und zeige, dass für $X_t = a_0 + a_1t + a_2t^2 + a_3t^3$ die Gleichung $\hat{m}(t) = X_t$ gilt und dass durch Bildung von $\hat{m}(t)$ eine saisonale Komponente mit der Periode 3 eliminiert wird.

Aufgabe 1.3 Man zeige, dass ein gleitendes Mittel jedes Polynom vom Grad k genau dann unverändert lässt, wenn die Gewichte w_j , $j = -q, \dots, q$, des gleitenden Mittels die Beziehungen

$$\sum_{j=-q}^q w_j = 1 \quad \text{und} \quad \sum_{j=-q}^q j^r w_j = 0 \quad \forall r = 1, \dots, k,$$

erfüllen.

Speziell für $w_j = 1/(2q+1)$ gilt also, dass das zugehörige gleitende Mittel lineare Ausdrücke $U + Vt$ für beliebige Zufallsvariable U, V unverändert lässt.

Aufgabe 1.4 Seien e_t , $t \in \mathbb{Z}$, reelle, paarweise unkorrelierte Zufallsvariable mit $Ee_t = 0$ und $Var e_t = \sigma_e^2$. Man zeige, dass für das gleitende Mittel

$$X_t = \sum_{j=-q}^q \frac{1}{2q+1} e_{t-j}, \quad \forall t \in \mathbb{Z},$$

eine zentrierte Zeitreihe (X_t) mit $Var X_t = \sigma_e/(2q+1)$ entsteht.

Stationarität und grundlegende Modelle der Zeitreihenanalyse

Mathematisch verstehen wir unter einer Zeitreihe, oft auch nur kurz *Reihe* genannt, einen reell- oder auch komplexwertigen stochastischen Prozess in diskreter Zeit, d.h. eine Familie $X = (X_t : t \in T)$ von reell- bzw. komplexwertigen Zufallsvariablen X_t (auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P)) indiziert mit einer Zeit $t \in T$. $T \subset \mathbb{R}$ stellt dabei eine diskrete Menge dar. Da wir uns nur mit einer äquidistant eingeteilten Zeitachse befassen werden, betrachten wir der Einfachheit halber stets $t \in \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ bzw. $t \in \mathbb{N} = \{1, 2, \dots\}$. Sind die Zufallsvariablen X_t \mathbb{R}^d - bzw. \mathbb{C}^d -wertig, so sprechen wir von einer multivariaten (reellen oder komplexen) Zeitreihe. Zunächst wollen wir uns aber ausschließlich dem Fall $d = 1$ zuwenden.

Im ersten Abschnitt werden die Begriffe der Stationarität und der Autokovarianz bzw. der Autokorrelation vorgestellt. Hieran anschließend beinhaltet Abschnitt 2.2 mit dem weißen Rauschen, den ARMA-Modellen und zyklischen Zeitreihen grundlegende Zeitreihenmodelle und gibt erste wichtige Resultate für diese wichtigen Prozesse an. Erste statistische Aspekte begegnen uns dann im Abschnitt 2.3 bei der Untersuchung der empirischen Autokovarianz und der sog. Yule-Walker-Parameterschätzer. Schließlich ist der Abschnitt 2.4 Gaußschen Zeitreihen gewidmet, bevor wir uns noch kurz der partiellen Autokorrelation in Abschnitt 2.5 zuwenden.

2.1 Stationarität von Zeitreihen

Da wir in den allermeisten Anwendungen nur eine einzige Realisation eines Ausschnitts X_1, X_2, \dots, X_n aus der zugrunde liegenden Zeitreihe $X = (X_t : t \in \mathbb{Z})$ beobachten können, sind statistische Rückschlüsse auf das die Zeitreihe generierende wahrscheinlichkeitstheoretische Modell bzw. Aussagen über das zukünftige Verhalten der beobachteten Zeitreihe nur möglich, wenn die wesentlichen Strukturen des beobachteten Ausschnittes X_1, \dots, X_n charakteristisch für die gesamte Zeitreihe, also auch für zukünftige Werte sind.

Eine sehr gängige und häufig unterstellte Annahme ist die der Stationarität. Die in der folgenden Definition auftretenden Erwartungswerte für komplexwertige Zufallsvariable $Y = U + iV$ mit Realteil U und Imaginärteil V sind gemäß $EY = EU + iEV$ definiert, siehe Anhang A.2. Obere Querstriche bedeuten in diesem Zusammenhang die konjugiert komplexen Größen.

Definition 2.1 (*Stationarität, Autokovarianz und Autokorrelation*)

$X = (X_t : t \in \mathbb{Z})$ sei eine reell- oder komplexwertige Zeitreihe.

i) Falls für alle $n \in \mathbb{N}$ und alle $t_1, t_2, \dots, t_n, h \in \mathbb{Z}$ gilt

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \mathcal{L}(X_{t_1+h}, \dots, X_{t_n+h}), \quad (2.1)$$

dann heißt die Zeitreihe $X = (X_t : t \in \mathbb{Z})$ streng stationär.

ii) Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit $E|X_t|^2 < \infty$ für alle $t \in \mathbb{Z}$ heißt L_2 -Zeitreihe. Eine L_2 -Zeitreihe heißt schwach stationär oder auch nur stationär, falls (2.2) und (2.3) gelten.

$$EX_t = \mu \quad \forall t \in \mathbb{Z}, \quad (\text{also unabhängig von } t) \quad (2.2)$$

$$\begin{aligned} \text{Cov}(X_{t+h}, X_t) &= E(X_{t+h} - \mu) \overline{(X_t - \mu)} \\ &\text{ist für jedes } h \in \mathbb{Z} \text{ unabhängig von } t. \end{aligned} \quad (2.3)$$

Die Funktion $\gamma = \gamma_X : \mathbb{Z} \rightarrow \mathbb{C}$ mit

$$\gamma(h) = E(X_{t+h} - \mu) \overline{(X_t - \mu)}, \quad \forall h \in \mathbb{Z}, \quad (2.4)$$

heißt dann die Autokovarianzfunktion der Zeitreihe $X = (X_t : t \in \mathbb{Z})$. Dabei ist $\gamma(0)$ stets reell und nicht negativ. Falls $\gamma(0) > 0$ ist, so heißt die Funktion $\varrho = \varrho_X : \mathbb{Z} \rightarrow \mathbb{C}$ mit

$$\varrho(h) = \frac{\gamma(h)}{\gamma(0)} \quad (2.5)$$

Autokorrelationsfunktion von $X = (X_t : t \in \mathbb{Z})$.

Zeitverschiebungen h , wie sie in obiger Definition auftreten, nennen wir wie im Englischen *Lag*. Wir werden natürlich in den allermeisten Anwendungen reellwertigen Zeitreihen oder multivariaten Zeitreihen mit reellwertigen Komponenten begegnen. Trotzdem macht es Sinn, bereits an dieser Stelle die Definition einer Zeitreihe auf den komplexen Fall auszudehnen. Wir werden später immer wieder auch auf komplexwertige Zeitreihen zurückgreifen müssen, um gewisse Argumentationen effizient darstellen zu können. Auch bei der Bezeichnung der Räume der quadratintegrierbaren Zufallsvariablen, der sog. L_2 -Räume, wollen wir unterscheiden:

- Den Hilbertraum der reellwertigen und quadratintegrierbaren Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) bezeichnen wir mit $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ oder kurz mit $L_2^{\mathbb{R}}$.

- Entsprechend bezeichnet $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ (kurz $L_2^{\mathbb{C}}$) den Hilbertraum der komplexwertigen und quadratintegrierbaren Zufallsvariablen auf (Ω, \mathcal{A}, P) .

Bemerkung 2.2

An dieser Stelle sei erwähnt, dass die so genannten L_2 -Räume natürlich nur über Äquivalenzklassen von Zufallsvariablen korrekt erklärt sind. Zwei Zufallsvariable heißen dabei äquivalent, wenn sie fast sicher, d.h. mit Wahrscheinlichkeit 1, übereinstimmen. Wir werden allerdings im Rahmen dieser Einführung nicht zwischen einer Äquivalenzklasse von Zufallsvariablen und einem Repräsentanten dieser Klasse unterscheiden.

Das Skalarprodukt $\langle \cdot, \cdot \rangle$ in diesen L_2 -Räumen ist bekanntlich wie folgt definiert

$$\langle X, Y \rangle = EX\bar{Y}. \quad (2.6)$$

Hiermit werden sowohl $L_2^{\mathbb{R}}$ wie $L_2^{\mathbb{C}}$ zu vollständigen normierten Räumen (sogar zu Hilberträumen), wobei die Norm auf den entsprechenden L_2 -Räumen über das Skalarprodukt induziert wird. Es gilt

$$\|X\| = \sqrt{\langle X, X \rangle} = \sqrt{E|X|^2}. \quad (2.7)$$

Für eine L_2 -Zeitreihe $X = (X_t : t \in \mathbb{Z})$ können wir die Erwartungswerte EX_t und die Autokovarianzen $Cov(X_{t+h}, X_t)$ mit Hilfe des Skalarproduktes wie folgt darstellen

$$EX_t = \langle X_t, 1 \rangle \quad (2.8)$$

und

$$Cov(X_{t+h}, X_t) = \langle X_{t+h} - EX_{t+h}, X_t - EX_t \rangle. \quad (2.9)$$

Außerdem gilt

$$E|X_t|^2 = \|X_t\|^2 \quad \text{und} \quad Var X_t = \|X_t - EX_t\|^2. \quad (2.10)$$

Es sei noch darauf hingewiesen, dass $\langle \cdot, \cdot \rangle$ in beiden Komponenten linear ist, wobei im $L_2^{\mathbb{C}}$ darauf zu achten ist, dass Faktoren aus der zweiten Komponente konjugiert werden müssen, wenn sie vor das Skalarprodukt gezogen werden.

Im Abschnitt 2.4 werden wir diese Skalarproduktschreibweise bereits sehr vorteilhaft einsetzen können. Natürlich stehen uns alle Rechenregeln für Skalarprodukte und Normen auf den Hilberträumen $L_2^{\mathbb{R}}$ und $L_2^{\mathbb{C}}$ zur Verfügung. Eine Zusammenstellung der für uns wesentlichen Regeln ist im Anhang A.3 zu finden. \square

Die folgende Aussage sieht man direkt ein.

Anmerkung 2.3

Für eine (reelle oder komplexe) L_2 -Zeitreihe folgt die Stationarität stets aus der strengen Stationarität. \square

Im Weiteren soll die grundlegende Eigenschaft der Stationarität und ihre Bedeutung ein wenig erläutert werden.

Die Annahme der strengen Stationarität bedeutet, dass die multivariate Verteilung eines beliebigen Abschnittes (nicht notwendig zu aufeinander folgenden Zeitpunkten) der Zeitreihe nur von der relativen Lage der entsprechenden Zeitpunkte untereinander abhängt, jedoch nicht von deren absoluter Lage auf der Zeitachse. Unter dieser Annahme können wir aus einer typischen Realisierung X_1, \dots, X_n einer reellen Zeitreihe $X = (X_t : t \in \mathbb{Z})$ wiederholt Beobachtungen über die Abhängigkeitsstruktur machen. Stellen wir uns z.B. vor, dass wir uns für die Abhängigkeitsstruktur zwischen X_t und X_{t+h} (für ein $h > 0$) interessieren. Unter der Annahme der strengen Stationarität ist diese Abhängigkeitsstruktur invariant in t und unsere Stichprobe enthält mit $(X_1, X_{1+h}), \dots, (X_{n-h}, X_n)$ $n - h$ wiederholte, aber allerdings im Allgemeinen abhängige, Beobachtungen aus der zweidimensionalen Verteilung von (X_t, X_{t+h}) .

Interessieren wir uns nur für den *linearen* Zusammenhang zwischen X_t und X_{t+h} , beschrieben durch die Kovarianz $\text{Cov}(X_{t+h}, X_t)$, und setzen wir (schwache) Stationarität voraus, so enthält unsere Stichprobe X_1, \dots, X_n ebenso wiederholte Realisierungen dieser Kovarianz, auch wenn jetzt diese Realisierungen $(X_1, X_{1+h}), \dots, (X_{n-h}, X_n)$ erschwerend zur Abhängigkeit möglicherweise auch noch jeweils verschiedenen Verteilungen entspringen können. Für stationäre (und zwar schwach stationäre) und zentrierte Zeitreihen folgt aber auf jeden Fall, dass der folgende Schätzer

$$\frac{1}{n-h} \sum_{t=1}^{n-h} X_{t+h} X_t$$

für jedes $h < n$ erwartungstreu für $EX_{t+h}X_t = \gamma(h)$ (man beachte die Zentriertheit!) ist. Die Konsistenz dieses Schätzers ist aber keineswegs klar und wird Inhalt eines noch zu beweisenden Gesetzes der großen Zahlen für Zeitreihen sein. Schon jetzt sei aber bemerkt, dass ohne weitere, über die Stationarität hinausgehende, Voraussetzungen der Nachweis der gewünschten Konsistenz unmöglich ist!

Wenn es uns nun gelingt, mittels statistischer Verfahren wie auch immer gartete Informationen über die Abhängigkeitsstruktur von X_t und X_{t+h} (für festes $h > 0$) zu gewinnen, dann können wir aus dem letzten Wert X_n der vorliegenden Stichprobe gewisse Vorhersagen über den zukünftigen Wert X_{n+h} treffen, da uns die Stationarität garantiert, dass der in der Vergangenheit beobachtete h -Schritt Zusammenhang auch für die Zukunft Bestand hat. Ohne eine zur Stationarität vergleichbare Voraussetzung erscheint eine Prognose auf zukünftige Werte der Zeitreihe dagegen gänzlich unmöglich.

Abschließend sei erwähnt, dass sich die Zeitreihenanalyse in weiten Bereichen im Rahmen einer Gewinnung von Informationen allein aus der Autokovarianz- bzw. der Autokorrelationsfunktion bewegt. Insofern werden gerade die statis-

tischen Teile dieser Einführung besonderen Wert auf die Schätzung dieser Funktionen und hieraus abgeleiteter Kenngrößen legen.

Bemerkung 2.4 (Hermite–Symmetrie der Autokovarianzfunktion)

Für eine komplexe stationäre Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit Autokovarianzfunktion γ gilt unter Verwendung der Integraldefinition komplexwertiger Funktionen aus Anhang A.2

$$\begin{aligned}\gamma(-h) &= E(X_{t-h} - \mu) \overline{(X_t - \mu)} \\ &= E(X_t - \mu) \overline{(X_{t+h} - \mu)} \\ &= \overline{E(\overline{X_t - \mu} (X_{t+h} - \mu))} \\ &= \overline{\gamma(h)}.\end{aligned}\tag{2.11}$$

Entsprechendes gilt für die Autokorrelationsfunktion ρ . Beim Nachweis beachte man, dass $\gamma(0)$ stets reell ist.

Für reelle stationäre Zeitreihen bedeutet dies, dass sowohl die Autokovarianz- als auch die Autokorrelationsfunktion eine symmetrische (und reelle) Funktion darstellt. Im reellen und im komplexen Fall reicht es also aus, beide Funktionen nur für $h = 0, 1, 2, \dots$ zu betrachten. \square

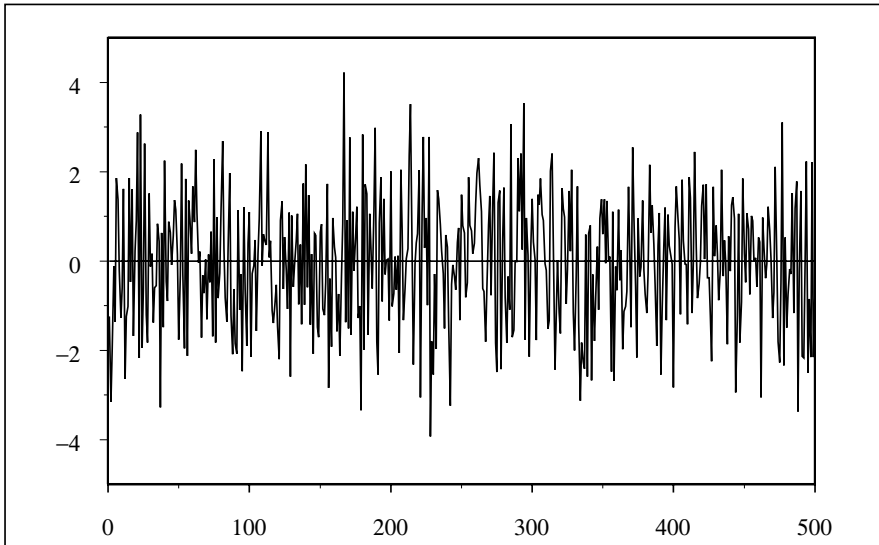


Abb. 2.1 Simulation eines normalverteilten weißen Rauschens mit Varianz 2

2.2 Grundlegende stationäre Zeitreihenmodelle

Ein einfaches aber grundlegendes Zeitreihenmodell ist das weiße Rauschen.

Beispiel 2.5 (Weißes Rauschen)

Für unkorrelierte reelle oder komplexe Zufallsvariable e_t , $t \in \mathbb{Z}$, mit $Ee_t = 0$ und $\text{Var } e_t = E|e_t|^2 = \sigma_e^2 \in (0, \infty)$ für alle $t \in \mathbb{Z}$ heißt die Zeitreihe $e = (e_t : t \in \mathbb{Z})$ ein weißes Rauschen. Die Bezeichnung *weißes* Rauschen wird erst in den Kapiteln 4 und 5 deutlich werden. Wir schreiben zukünftig gelegentlich kurz $e_t \sim (0, \sigma_e^2)$, um auszudrücken, dass Erwartungswert und Varianz von e_t die Werte 0 und σ_e^2 besitzen und entsprechend in anderen Situationen.

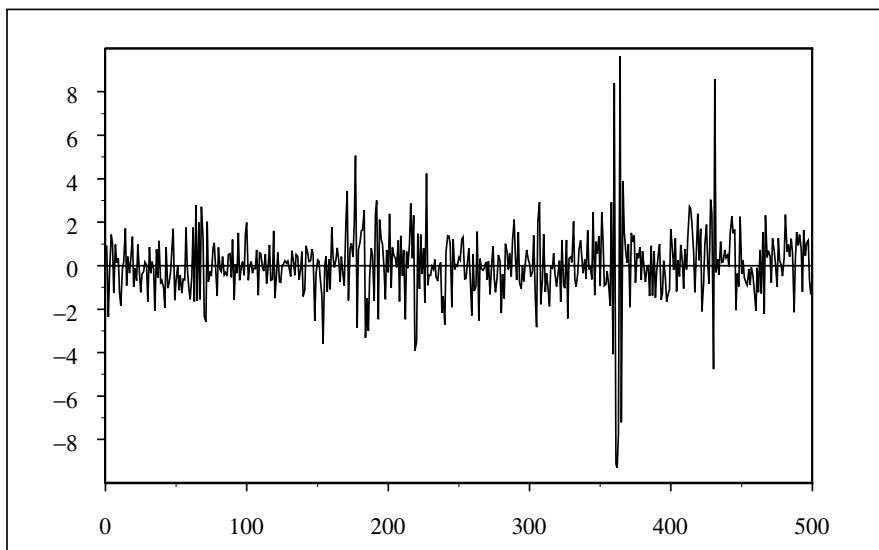


Abb. 2.2 Simulation eines nur unkorrelierten weißen Rauschens mit Varianz 2

Die Autokovarianzfunktion eines weißen Rauschens erfüllt $\gamma(0) = \sigma_e^2$ sowie $\gamma(h) = 0$ für $h \neq 0$. Entsprechend verschwindet die Autokorrelation $\varrho(h)$ für alle $h \neq 0$ und es gilt $\varrho(0) = 1$ wegen $\sigma_e^2 > 0$. Die Abbildung 2.1 zeigt eine Realisierung eines normalverteilten weißen Rauschens mit $\sigma_e^2 = 2$. Bekanntlich sind die einzelnen Zufallsvariablen e_t in diesem Fall dann sogar stochastisch unabhängig. Eine Realisierung eines nur unkorrelierten – jedoch nicht stochastisch unabhängigen und damit auch nicht normalverteilten – weißen Rauschens, ebenfalls mit Varianz $\sigma_e^2 = 2$, ist in Abbildung 2.2 zu sehen. Bei dem weißen Rauschen in Abbildung 2.2 bestehen zwar keine linearen Zusam-

menhänge zwischen den einzelnen Beobachtungen zu verschiedenen Zeitpunkten, aber sehr wohl (nicht lineare) Abhängigkeitsstrukturen. Bereits mit dem bloßen Auge sind deutliche Unterschiede in den beiden Pfaden zu entdecken. \square

Aus einem weißen Rauschen lassen sich drei wichtige Klassen linearer Zeitreihenmodelle ableiten, nämlich die sogenannten Autoregressiven (kurz: *AR*) und Moving Average (kurz: *MA*) Zeitreihen, sowie eine Zusammensetzung beider Klassen zu sog. *ARMA*-Modellen. Obwohl die übergreifenden *ARMA*-Modelle sowohl Autoregressive wie Moving Average Zeitreihen vollständig als Spezialfälle enthalten, sollen diese wegen ihrer Bedeutung jeweils in eigenen Beispielen vorgestellt werden. Beginnen wir mit den einfacher zu untersuchenden Moving Average-Zeitreihen.

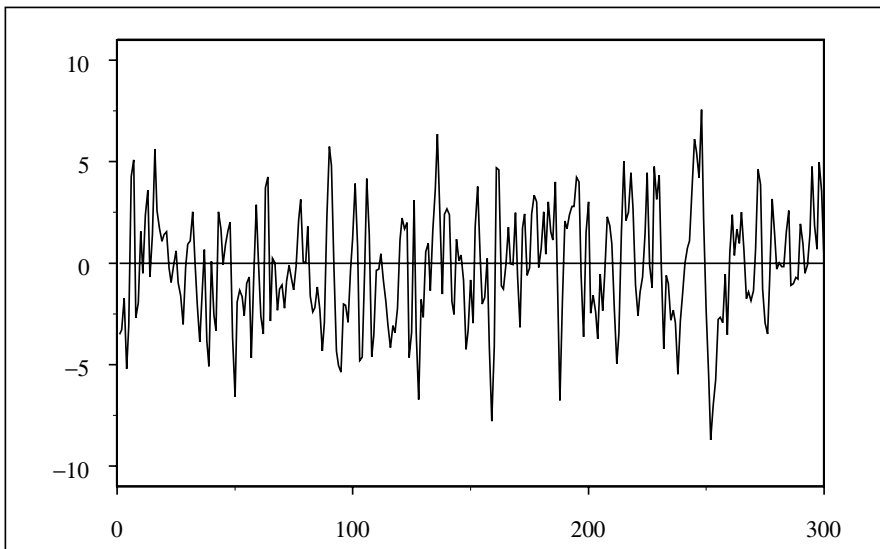


Abb. 2.3 Simulation einer $MA(1)$ -Zeitreihe mit $b_1 = 1$ und $\sigma_e^2 = 4$

Beispiel 2.6 (Moving Average (*MA*)-Zeitreihen)

$e = (e_t : t \in \mathbb{Z})$ bezeichne ein weißes Rauschen. Für reelle oder komplexe Koeffizienten b_1, \dots, b_q mit $b_q \neq 0$ und $q \in \mathbb{N}$ definiert

$$X_t = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}, \quad \forall t \in \mathbb{Z}, \quad (2.12)$$

eine stationäre Zeitreihe. $X = (X_t : t \in \mathbb{Z})$ heißt *Moving Average Zeitreihe der Ordnung q* oder kurz *MA(q)-Zeitreihe*. Im Fall eines reellen weißen Rauschens und reeller Koeffizienten ist die *MA(q)-Zeitreihe* selbstverständlich auch reell. Die Autokovarianzfunktion γ ergibt sich mit $b_0 := 1$ zu (beachte $EX_t = 0 \forall t$)

$$\begin{aligned}
\gamma(h) &= EX_{t+h}\bar{X}_t \\
&= \sum_{j,k=0}^q b_k \bar{b}_j Ee_{t+h-k}\bar{e}_{t-j} \\
&= \begin{cases} \sigma_e^2 \cdot \sum_{j=0}^{q-h} b_{j+h} \bar{b}_j & , h = 0, 1, 2, \dots, q, \\ 0 & , h > q. \end{cases} \quad (2.13)
\end{aligned}$$

Zunächst beweist die Darstellung (2.13) die oben behauptete Stationarität von $MA(q)$ -Zeitreihen. Außerdem zeigt die explizite Darstellung von γ , dass bei $MA(q)$ -Zeitreihen die Autokovarianzfunktion für alle Lags $h \geq q + 1$ verschwindet. Zeitreihenwerte, deren Zeitindizes weiter als q Einheiten auseinander liegen, sind somit stets unkorreliert. Falls das weiße Rauschen aus stochastisch unabhängigen Zufallsvariablen besteht, folgt sogar, dass X_t und X_s mit $|s - t| > q$ stochastisch unabhängig sind.

Die Abbildungen 2.3 und 2.4 zeigen Realisierungen von MA -Zeitreihen der Länge $n = 500$. Es fällt deutlich auf, dass der Pfad der $MA(7)$ -Zeitreihe in Abbildung 2.4 einen viel ruhigeren Verlauf als der Pfad der $MA(1)$ -Zeitreihe in Abbildung 2.3 aufweist. \square

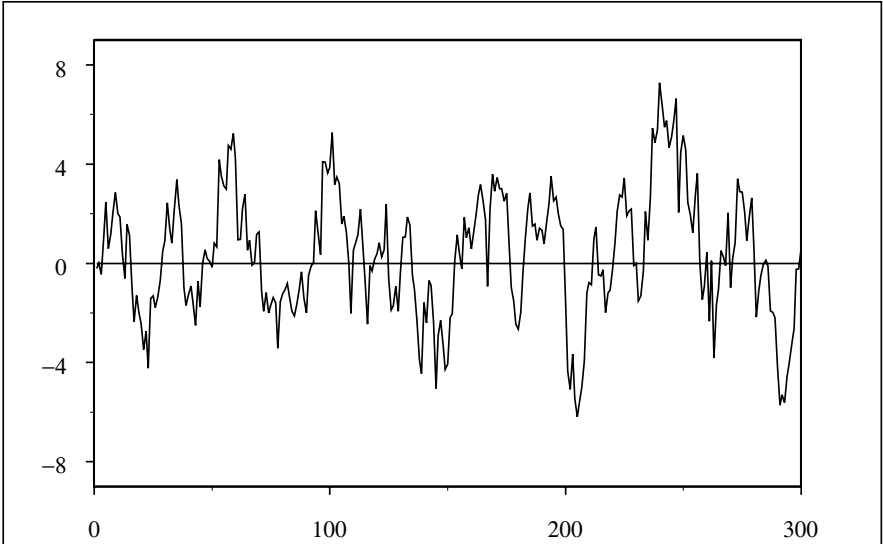


Abb. 2.4 Simulation einer $MA(7)$ -Zeitreihe mit $b_1 = \dots = b_7 = 1$ und $\sigma_e^2 = 1$

Beispiel 2.7 (AR(1)–Zeitreihe)

Wir nennen eine stationäre Zeitreihe $X = (X_t : t \in \mathbb{Z})$ *autoregressiv von der Ordnung 1*, kurz AR(1)–Zeitreihe, falls es einen Parameter $a \in \mathbb{C}$ und ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ gibt, so dass gilt

$$X_t + aX_{t-1} = e_t, \quad \forall t \in \mathbb{Z}. \quad (2.14)$$

Man nennt (2.14) eine *lineare Differenzengleichung erster Ordnung*, da abgesehen von den *Störungen* e_t die Beobachtung X_t vom vorausgehenden Wert X_{t-1} abhängt (was *erste Ordnung* bedeutet) und zwar in linearer Weise.

Im Gegensatz zu Moving Average Zeitreihen ist in diesem Fall nicht klar, ob es für ein gegebenes weißes Rauschen und einen gegebenen Parameter überhaupt derartige Zeitreihen $X = (X_t : t \in \mathbb{Z})$ gibt. Für Parameter a mit $|a| < 1$ können wir aber in der Tat stets stationäre AR(1)–Zeitreihen konstruieren, wie die folgende Überlegung zeigt. Durch wiederholte Anwendung von (2.14) erhalten wir für jedes $q \in \mathbb{N}$

$$X_t = \sum_{\nu=0}^q (-a)^\nu e_{t-\nu} + (-a)^{q+1} X_{t-(q+1)}. \quad (2.15)$$

Wir können eine stationäre Lösung des AR(1)–Schemas (unter der Voraussetzung, dass eine stationäre Lösung überhaupt existiert) also bis auf den Restterm $(-a)^{q+1} X_{t-(q+1)}$ als Moving Average Zeitreihe der Ordnung q schreiben. Wegen $|a| < 1$ gilt aufgrund der Stationarität für diesen Restterm

$$E \left| (-a)^{q+1} X_{t-(q+1)} \right|^2 = |a|^{2(q+1)} E |X_0|^2 \longrightarrow_{q \rightarrow \infty} 0.$$

Mit anderen Worten konvergiert die MA(q)–Zeitreihe auf der rechten Seite in Formel (2.15) gegen X_t für $q \rightarrow \infty$ im quadratischen Mittel (also im L_2). Falls es für (2.14) mit $|a| < 1$ und gegebenem weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ also überhaupt eine stationäre Lösung gibt, so ist diese notwendigerweise die Zeitreihe

$$X_t = \sum_{\nu=0}^{\infty} (-a)^\nu e_{t-\nu}, \quad \forall t \in \mathbb{Z}. \quad (2.16)$$

Dabei ist die Konvergenz der Reihe auf der rechten Seite (wie oben erläutert) im L_2 –Sinne zu verstehen. Man nennt eine Zeitreihe mit der Darstellung (2.16), also mit einer Darstellung als gewichtete unendliche Reihe über ein weißes Rauschen, auch Moving Average Zeitreihe mit unendlicher Ordnung (kurz MA(∞)–Zeitreihe).

Andererseits definiert die rechte Seite von (2.16) für $|a| < 1$ stets ein Element aus $L_2^{\mathbb{C}}$, da $(\sum_{\nu=0}^q (-a)^\nu e_{t-\nu} : q \in \mathbb{N})$ für jedes $t \in \mathbb{Z}$ eine Cauchy–Folge

im $L_2^{\mathbb{C}}$ und damit konvergent ist (vgl. Aufgabe 2.1). Wir können also für ein gegebenes weißes Rauschen $(e_t : t \in \mathbb{Z})$ und $a \in \mathbb{C}$ mit $|a| < 1$ eine Zeitreihe $(X_t : t \in \mathbb{Z})$ gemäß (2.16) definieren. Wegen

$$\begin{aligned} X_t + aX_{t-1} &= L_2 - \lim_{q \rightarrow \infty} \underbrace{\left(\sum_{\nu=0}^q (-a)^{\nu} e_{t-\nu} + a \sum_{\nu=0}^{q-1} (-a)^{\nu} e_{t-1-\nu} \right)}_{=e_t} \\ &= e_t, \quad \forall t \in \mathbb{Z}, \end{aligned}$$

erfüllt die Zeitreihe (2.16) das autoregressive Schema (2.14) und ist somit also eine $AR(1)$ -Zeitreihe.

Die Stationarität der Zeitreihe (2.16) wird im Folgenden nachgewiesen. Bei den hierfür notwendigen Berechnungen des Erwartungswertes und der Autokovarianz machen wir Gebrauch von den Darstellungen (2.8) und (2.9) aus Bemerkung 2.2 und der Stetigkeit des Skalarproduktes in beiden Komponenten (vgl. Anhang A.3, Satz A.9). Zunächst erhalten wir die Zentriertheit

$$\begin{aligned} EX_t &= \langle X_t, 1 \rangle = \lim_{q \rightarrow \infty} \left\langle \sum_{\nu=0}^q (-a)^{\nu} e_{t-\nu}, 1 \right\rangle \\ &= \lim_{q \rightarrow \infty} \sum_{\nu=0}^q (-a)^{\nu} \langle e_{t-\nu}, 1 \rangle = 0. \end{aligned}$$

Die Autokovarianz ergibt sich zu

$$\begin{aligned} Cov(X_{t+h}, X_t) &= EX_{t+h} \overline{X_t} = \langle X_{t+h}, X_t \rangle \\ &= \lim_{q \rightarrow \infty} \left\langle \sum_{\nu=0}^q (-a)^{\nu} e_{t+h-\nu}, \sum_{\nu=0}^q (-a)^{\nu} e_{t-\nu} \right\rangle \\ &= \lim_{q \rightarrow \infty} Cov \left(\sum_{\nu=0}^q (-a)^{\nu} e_{t+h-\nu}, \sum_{\nu=0}^q (-a)^{\nu} e_{t-\nu} \right) \\ &= \sigma_e^2 \cdot \sum_{\nu=0}^{\infty} (-a)^{\nu+h} (-\bar{a})^{\nu}, \quad \text{vgl. (2.13)} \\ &= \sigma_e^2 \cdot \frac{(-a)^h}{1 - |a|^2}, \quad h = 0, 1, 2, \dots \end{aligned} \tag{2.17}$$

Zusammen mit der obigen Eindeutigkeit haben wir also für den Fall $|a| < 1$ gezeigt, dass (2.16) eine stationäre $AR(1)$ -Zeitreihe ist und dass jede stationäre $AR(1)$ -Zeitreihe von dieser Bauart ist. (2.17) gibt damit die Autokovarianzfunktion für $AR(1)$ -Zeitreihen mit Parameter $|a| < 1$ an. Entsprechend lautet die Autokorrelationsfunktion einer $AR(1)$ -Zeitreihe

$$\varrho(h) = (-a)^h, \quad h = 0, 1, 2, \dots \tag{2.18}$$

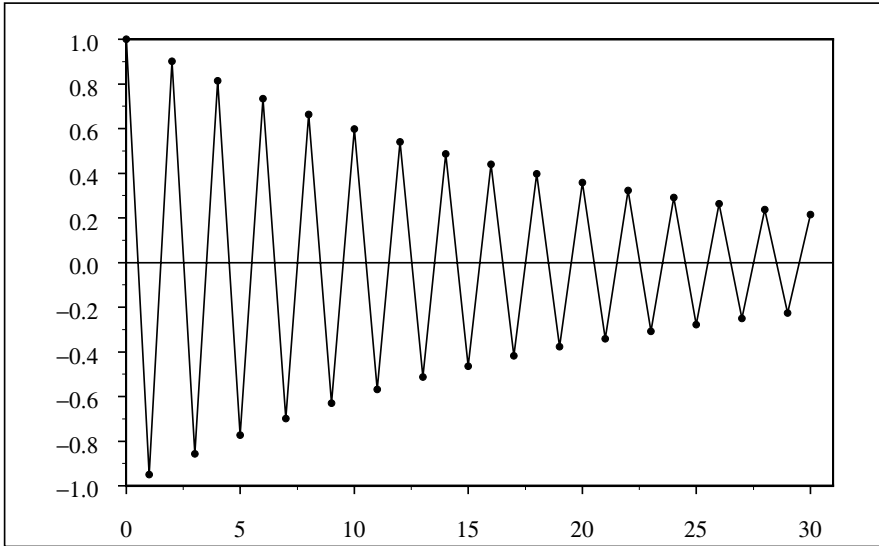


Abb. 2.5 Korrelogramm einer $AR(1)$ -Zeitreihe mit Parameter $a \in (0, 1)$

Es drängt sich die Frage auf, ob die Reihe in (2.16) noch in einem anderen Sinne konvergiert. In der Tat ist diese Reihe sogar fast sicher absolut (d.h. mit Wahrscheinlichkeit 1) konvergent. Während wir für den Nachweis der L_2 -Konvergenz auf Aufgabe 2.1 verwiesen haben, soll die fast sichere absolute Konvergenz hier gezeigt werden. Zunächst ergibt der Satz von der monotonen Konvergenz

$$E \sum_{\nu=0}^{\infty} |(-a)^{\nu} e_{t-\nu}| = \sum_{\nu=0}^{\infty} |a|^{\nu} E |e_{t-\nu}| \leq \sqrt{E |e_1|^2} \sum_{\nu=0}^{\infty} |a|^{\nu} .$$

Wegen $|a| < 1$ ist dieser Ausdruck endlich und deshalb muss auch die Reihe $\sum_{\nu=0}^{\infty} |(-a)^{\nu} e_{t-\nu}|$ fast sicher endlich sein (ansonsten wäre der Erwartungswert unendlich). Die Reihe $\sum_{\nu=0}^{\infty} (-a)^{\nu} e_{t-\nu}$ ist demnach für alle $t \in \mathbb{Z}$ fast sicher absolut konvergent. Dieser fast sichere Limes und der obige L_2 -Limes stimmen bekanntlich fast sicher überein.

Für stationäre $AR(1)$ -Prozesse klingt sowohl die Autokovarianz- wie die Autokorrelationsfunktion mit wachsendem h exponentiell ab und konvergiert für $h \rightarrow \infty$ gegen Null (vgl. (2.17) und (2.18)).

Beispiele für den Graphen von ϱ , dies sind sog. *Korrelogramme*, sind in den Abbildungen 2.5 und 2.6 für die grundsätzlich ganz unterschiedlichen Fälle $a \in (0, 1)$ und $a \in (-1, 0)$ gegeben.

Im Spezialfall $a = 0$ ist der $AR(1)$ -Prozess natürlich nichts anderes als ein weißes Rauschen.

Abschließend sei noch erwähnt, dass wegen (2.16) die $AR(1)$ -Zeitreihe bei reellem weißen Rauschen und $a \in \mathbb{R}$, $|a| < 1$, selbst auch stets reell ist. \square

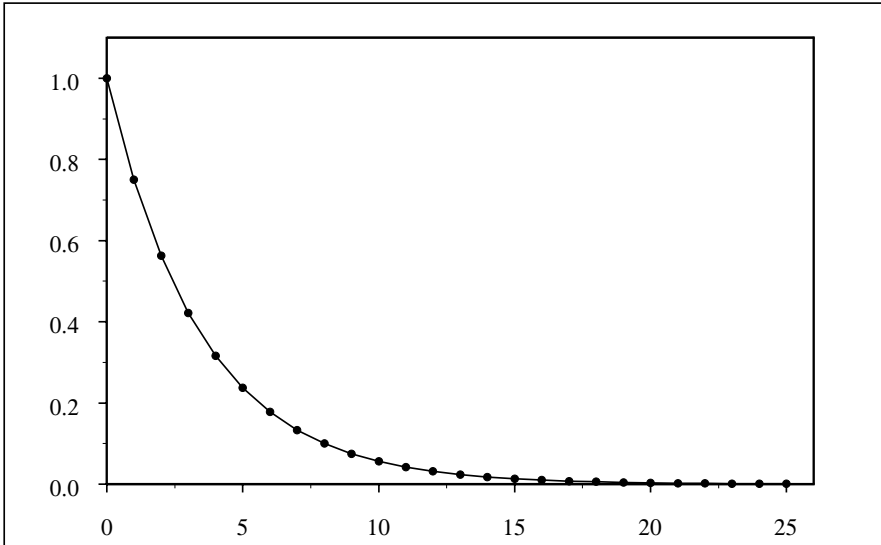


Abb. 2.6 Korrelogramm einer $AR(1)$ -Zeitreihe mit Parameter $a \in (-1, 0)$

Man kann das Vorgehen aus Beispiel 2.7 auf autoregressive Zeitreihenmodelle höherer Ordnung verallgemeinern. Dies wird auch später im Kapitel 8 geschehen. Hier soll die allgemeine Situation aber bereits kurz erläutert werden.

Wir nennen eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ *autoregressiv von der Ordnung* $p \in \mathbb{N}$, kurz $AR(p)$ -Zeitreihe mit Koeffizienten $a_1, \dots, a_p \in \mathbb{C}$, falls $a_p \neq 0$ und falls für ein geeignetes weißes Rauschen $e = (e_t : t \in \mathbb{Z})$

$$X_t + \sum_{\nu=1}^p a_\nu X_{t-\nu} = e_t, \quad \forall t \in \mathbb{Z}, \quad (2.19)$$

gilt.

Gleichung (2.19) ist eine Regressionsgleichung, wie sie aus der Statistik bekannt ist. Die Regressoren sind hier X_{t-1}, \dots, X_{t-p} , also Beobachtungen aus der Historie der interessierenden Zeitreihe selbst. X_t greift also auf die eigene Vergangenheit zurück, was zum Namen *auto*-regressiv geführt hat. Die Modellvorstellung hinter (2.19) ist, dass sich der Wert der Zeitreihe zum Zeitpunkt t als (mit $-a_1, \dots, -a_p$) gewichtete Summe der p vergangenen Werte zuzüglich einer additiven Störung durch ein weißes Rauschen ergibt.

Im Gegensatz dazu sind MA -Zeitreihen aus der Modellvorstellung entstanden, dass der Wert der Zeitreihe zum Zeitpunkt t durch zufällige „Schocks“

$e_s, s < t$, der Vergangenheit beeinflusst wird. Bei einer $MA(q)$ -Zeitreihe hat ein solcher Zufallsschock zum Zeitpunkt s Einfluss auf die Zeitreihenwerte $X_s, X_{s+1}, \dots, X_{s+q}$, d.h. er wirkt nicht nur momentan, sondern hat auch einen begrenzten Einfluss auf die Zukunft. Dies ist für viele wirtschaftswissenschaftliche Anwendungen eine durchaus realistische Annahme.

In Beispiel 2.7 haben wir die Autokovarianzfunktion einer stationären $AR(1)$ -Zeitreihe mit Parameter $|a| < 1$ explizit berechnet. Direkt aus der Modellgleichung (2.14) und der Tatsache, dass X_t nur von $e_t, e_{t-1}, e_{t-2}, \dots$ abhängt (vgl. (2.16)), sowie wegen $EX_t = 0$ erhalten wir die folgende Rekursion für $\gamma(h)$

$$\begin{aligned}\gamma(h) + a\gamma(h-1) &= EX_{t+h}\bar{X}_t + aEX_{t+h-1}\bar{X}_t \\ &= Ee_{t+h}\bar{X}_t = \begin{cases} \sigma_e^2, & h = 0 \\ 0, & h = 1, 2, \dots \end{cases} \end{aligned} \quad (2.20)$$

Dies führt auf die beiden folgenden Gleichungen

$$\begin{aligned}\gamma(1) + a\gamma(0) &= 0 \\ \gamma(0) + a\gamma(-1) &= \sigma_e^2,\end{aligned}$$

bzw. wegen $\gamma(-1) = \overline{\gamma(1)} = -\bar{a}\gamma(0)$ auf

$$\begin{aligned}\gamma(1) + a\gamma(0) &= 0 \\ (1 - |a|^2)\gamma(0) &= \sigma_e^2.\end{aligned} \quad (2.21)$$

Man nennt (2.21) die *Yule-Walker-Gleichungen* des $AR(1)$ -Modells. Bei gegebenem $\sigma_e^2 > 0$ und $|a| < 1$ erhalten wir hieraus $\gamma(0)$ und $\gamma(1)$ sowie mit (2.20) alle weiteren Autokovarianzen $\gamma(h), h \geq 1$. Umgekehrt legen die Autokovarianzen $\gamma(0)$ und $\gamma(1)$ gemäß (2.21) a und σ_e^2 fest.

Wir werden in Kapitel 8 Yule-Walker-Gleichungen für stationäre autoregressive Prozesse höherer Ordnung kennen lernen und natürlich untersuchen. Auch bei der Parameterschätzung in $AR(p)$ -Zeitreihenmodellen werden die Yule-Walker-Gleichungen eine wesentliche Rolle spielen. In diesem Zusammenhang sei für $p = 1$ auf Abschnitt 2.3 und für die allgemeine Situation auf Kapitel 8 und Kapitel 11 verwiesen.

Moving Average und Autoregressive Zeitreihen können zu einer neuen Modellklasse der sog. *ARMA*-Zeitreihen zusammengeführt werden. Die Hoffnung ist, dass man in einer zusammengesetzten Klasse mit weniger Parametern zur Beschreibung von realen Daten auskommt (vgl. Box/Jenkins (1970)).

Definition 2.8 (*ARMA(p, q)-Prozesse*)

Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt *ARMA(p, q)-Zeitreihe* mit den Ordnungen $p, q \in \mathbb{N}$ und Koeffizienten a_1, \dots, a_p und b_1, \dots, b_q mit $a_p \neq 0$ und $b_q \neq 0$, falls für ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ gilt

$$X_t + \sum_{\nu=1}^p a_\nu X_{t-\nu} = \sum_{\mu=1}^q b_\mu e_{t-\mu} + e_t, \quad \forall t \in \mathbb{Z}. \quad (2.22)$$

Die Existenz stationärer ARMA-Zeitreihen und deren Eigenschaften werden wir in Kapitel 7 detailliert untersuchen. An dieser Stelle wollen wir uns auf die Definition beschränken und noch eine weitere grundlegende Klasse von Zeitreihen, die sogenannten *zyklischen Zeitreihen*, vorstellen.

Definition 2.9 (*Reelle zyklische Zeitreihen*)

Für $M \in \mathbb{N}$ bezeichnen $e_{1j}, e_{2j}, j = 1, \dots, M$, reelle, zentrierte, quadratintegrierbare und unkorrelierte Zufallsvariable mit $\text{Var } e_{1j} = \text{Var } e_{2j} = \sigma_j^2$. $\omega_1, \dots, \omega_M \in (-\pi, \pi]$ seien fest gewählte Frequenzen. Die Zeitreihe

$$X_t = \sum_{j=1}^M (e_{1j} \cos(\omega_j t) + e_{2j} \sin(\omega_j t)), \quad \forall t \in \mathbb{Z}, \quad (2.23)$$

heißt (reelle) *zyklische Zeitreihe*.

Es handelt sich bei zyklischen Zeitreihen um eine Überlagerung von Schwingungen mit festen Frequenzen und stochastischen Amplituden.

Wie die folgende Rechnung zeigt sind zyklische Zeitreihen trotz ihrer deterministischen Frequenzen stationär. Es gilt nämlich zum einen $EX_t = 0$ sowie zum anderen

$$\begin{aligned} \gamma(h) &= EX_{t+h}X_t \\ &= \sum_{j=1}^M \{Ee_{1j}^2 \cos(\omega_j t) \cos(\omega_j(t+h)) + Ee_{2j}^2 \sin(\omega_j t) \sin(\omega_j(t+h))\} \\ &= \sum_{j=1}^M \sigma_j^2 \cos(\omega_j h), \quad \forall t, h \in \mathbb{Z}, \end{aligned} \quad (2.24)$$

wobei für die letzte Gleichung die Formel (A.5) aus Anhang A.1 verwendet wurde.

Interessanterweise treten die Frequenzen ω_j der Zeitreihe X_t in den Autokovarianzen wieder auf und zwar versehen mit einem Gewicht, welches gleich der Varianz der zur Frequenz ω_j gehörenden stochastischen Amplitude ist. Dieses Phänomen, dass die Autokovarianz die in der Zeitreihe auftretenden Frequenzen widerspiegelt, werden wir in Kapitel 5 erneut aufgreifen und erschöpfend behandeln.

Beispiel 2.10 (*ARCH(1)-Zeitreihe*)

X_0 bezeichne eine reelle Zufallsvariable mit $EX_0^2 < \infty$ und $(e_t : t \in \mathbb{N})$ sei ein von X_0 unabhängiges weißes Rauschen, bestehend aus stochastisch unabhängigen und identisch verteilten (i.i.d.) Zufallsvariablen mit Varianz 1. Man definiere dann für $a_0 > 0$ und $0 < a_1 < 1$

$$X_t = \sqrt{a_0 + a_1 X_{t-1}^2} \cdot e_t, \quad \forall t = 1, 2, \dots \quad (2.25)$$

Man erkennt, dass X_{t-1} lediglich von X_0 und e_s mit $s \leq t-1$ abhängt und deswegen unabhängig von e_t ist. (2.25) ergibt also für $t = 1, 2, \dots$

$$E X_t = E \sqrt{a_0 + a_1 X_{t-1}^2} \cdot E e_t = 0, \quad (2.26)$$

$$E X_t^2 = E (a_0 + a_1 X_{t-1}^2) \cdot E e_t^2 = a_0 + a_1 E X_{t-1}^2, \quad (2.27)$$

sowie für $h \geq 1$ und $t = 1, 2, \dots$

$$E X_t X_{t+h} = E \left(\sqrt{a_0 + a_1 X_{t+h-1}^2} X_t \right) \cdot E e_{t+h} = 0. \quad (2.28)$$

Setzt man $E X_0^2 = a_0/(1 - a_1)$, so folgt durch rekursives Einsetzen in (2.27) schließlich

$$E X_t^2 = a_0 + a_1 \frac{a_0}{1 - a_1} = \frac{a_0}{1 - a_1},$$

so dass die Zeitreihe $(X_t : t \in \mathbb{N})$ ein weißes Rauschen darstellt. Da aber X_t über (2.25) direkt von X_{t-1} abhängt, liegt lediglich Unkorreliertheit aber keine Unabhängigkeit vor. In Abbildung 2.2 ist für den Fall $a_0 = 1$ und $a_1 = 0.5$ eine Simulation einer ARCH(1)-Zeitreihe angegeben.

Zeitreihen, die der Rekursion (2.25) genügen, heißen autoregressiv mit bedingter Heteroskedastizität (autoregressive with conditional heteroscedasticity (ARCH)) und wurden 1982 von Robert Engle eingeführt, um finanzielle Zeitreihen, wie etwa die Renditen eines Finanzgutes (vgl. Abbildung 1.7 für den DAX) angemessen modellieren zu können. Hierfür wurde Engle 2003 mit dem Preis der schwedischen Nationalbank zu Ehren Alfred Nobels ausgezeichnet. Wir werden in Kapitel 14 genauer auf diese Modellklasse und Erweiterungen hiervon eingehen und beschließen die Zusammenstellung von Beispielen für Zeitreihenmodelle an dieser Stelle zunächst. \square

2.3 Empirische Autokovarianzen und Autokorrelationen

Bereits im Abschnitt 2.1 haben wir mit $\frac{1}{n-h} \sum_{t=1}^{n-h} X_t X_{t+h}$ einen zumindest erwartungstreuen Schätzer für die Autokovarianz $\gamma(h)$ einer reellen stationären und zentrierten Zeitreihe $X = (X_t : t \in \mathbb{Z})$ kennen gelernt. Im Falle einer nicht zentrierten reellen stationären Zeitreihe bietet sich die empirische Zentrierung mit $\bar{X} := \bar{X}_n := \frac{1}{n} \sum_{t=1}^n X_t$ an, wobei der Querstrich hier die Durchschnittsbildung kennzeichnet und nicht „konjugiert komplex“. In der tatsächlichen Anwendung benutzt man dann auch

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n) (X_t - \bar{X}_n), \quad h = 0, 1, 2, \dots, n-1, \quad (2.29)$$

und $\hat{\gamma}(h) := 0$ für $h \geq n$ sowie (falls $\hat{\gamma}(0) > 0$)

$$\hat{\varrho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h = 0, 1, 2, \dots, \quad (2.30)$$

als Schätzer für γ bzw. ϱ und nennt $\hat{\gamma}$ die *empirische oder Stichproben-Autokovarianzfunktion* und $\hat{\varrho}$ entsprechend die *empirische oder Stichproben-Autokorrelationsfunktion*.

Sicherlich verwundert zunächst der für alle Lags h identische Vorfaktor $\frac{1}{n}$ anstelle von $\frac{1}{n-h}$. Die Begründung hierfür wird im Folgenden klar werden.

Für stationäre autoregressive Zeitreihen erster Ordnung haben wir mit den Yule–Walker–Gleichungen (vgl. (2.21)) einen einfachen Zusammenhang zwischen der Autokovarianzfunktion und den Parametern a und σ_e^2 angegeben. Es liegt nun nahe, mit diesen Gleichungen und den empirischen Autokovarianzen $\hat{\gamma}(0)$ und $\hat{\gamma}(1)$ Parameterschätzer \hat{a} für a und $\hat{\sigma}_e^2$ für σ_e^2 zu erklären. Dieses Vorgehen ist in der Tat äußerst sinnvoll und führt zu den sog. *Yule–Walker–Parameterschätzern*. Es ist wünschenswert, dass auch der Parameterschätzer \hat{a} wie der tatsächliche Parameter a die Eigenschaft $|\hat{a}| < 1$ besitzt und stets $\hat{\sigma}_e^2 \geq 0$ gilt. Hierzu haben wir das folgende Resultat, welches wesentlich auf der Wahl des einheitlichen Vorfaktors in (2.29) beruht.

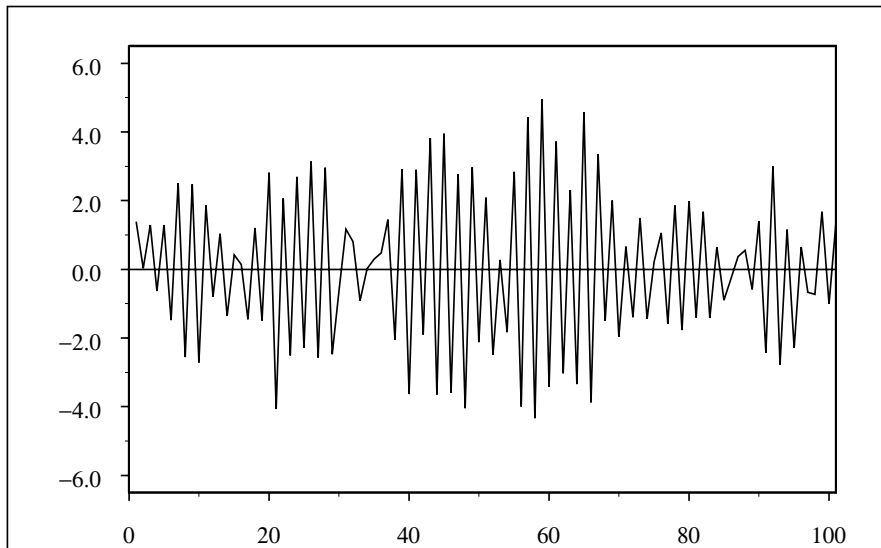


Abb. 2.7 Realisierung einer $AR(1)$ –Zeitreihe mit Parameter $a = 0.95$

Lemma 2.11

Für gegebene reelle Beobachtungen X_1, \dots, X_n einer Zeitreihe seien $\hat{\gamma}(0)$ und

$\hat{\gamma}(1)$ wie in (2.29) gegeben. Dann gilt

$$\hat{\gamma}(0) > 0 \iff \text{nicht alle } X_t \text{ sind gleich.} \quad (2.31)$$

Im Falle $\hat{\gamma}(0) > 0$ gilt für $\hat{a} := -\hat{\gamma}(1)/\hat{\gamma}(0) = -\hat{\varrho}(1)$ stets

$$|\hat{a}| < 1. \quad (2.32)$$

Beweis: Wir haben

$$\begin{aligned} \hat{\gamma}(0) = 0 &\iff \sum_{t=1}^n (X_t - \bar{X}_n)^2 = 0 \\ &\iff X_t = \bar{X}_n \quad \forall t = 1, \dots, n \\ &\iff \text{alle } X_t \text{ sind gleich.} \end{aligned}$$

Weiter folgt für $\hat{\gamma}(0) > 0$ mit der Cauchy–Schwarz Ungleichung

$$\begin{aligned} |\hat{\gamma}(1)| &= \frac{1}{n} \left| \sum_{t=1}^{n-1} (X_{t+1} - \bar{X}_n) (X_t - \bar{X}_n) \right| \\ &\leq \left(\frac{1}{n} \sum_{t=1}^{n-1} (X_{t+1} - \bar{X}_n)^2 \right)^{\frac{1}{2}} \left(\frac{1}{n} \sum_{t=1}^{n-1} (X_t - \bar{X}_n)^2 \right)^{\frac{1}{2}} \\ &\leq \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2 = \hat{\gamma}(0). \end{aligned}$$

Hieraus erhalten wir zunächst $|\hat{a}| \leq 1$. Falls $|\hat{a}| = 1$, d.h. $|\hat{\gamma}(1)| = \hat{\gamma}(0)$, folgt hieraus, dass die letzte Ungleichung eine Gleichung ist. Dies ist aber offenbar nur dann der Fall, wenn $X_1 = X_n = \bar{X}_n$ gilt, woraus weiter folgt

$$\begin{aligned} |n\hat{\gamma}(1)| &= \left| \sum_{t=2}^{n-2} (X_{t+1} - \bar{X}_n) (X_t - \bar{X}_n) \right| \\ &\leq \left(\sum_{t=2}^{n-2} (X_{t+1} - \bar{X}_n)^2 \right)^{\frac{1}{2}} \left(\sum_{t=2}^{n-2} (X_t - \bar{X}_n)^2 \right)^{\frac{1}{2}} \\ &\leq \sum_{t=2}^{n-1} (X_t - \bar{X}_n)^2 = n\hat{\gamma}(0). \end{aligned}$$

Hieraus erhält man wegen $|\hat{\gamma}(1)| = \hat{\gamma}(0)$ auf die gleiche Weise wie oben $X_2 = X_{n-1} = \bar{X}_n$. Sukzessive fährt man fort und erreicht schließlich $X_1 = \dots = X_n = \bar{X}_n$ und damit, dass aus $|\hat{a}| = 1$ die Aussage $\hat{\gamma}(0) = 0$, also ein Widerspruch folgt. Für ungerade n beachte man, dass aus der Tatsache, dass alle bis auf eine Beobachtung mit \bar{X}_n übereinstimmen sofort folgt, dass alle Beobachtungen mit \bar{X}_n übereinstimmen müssen. \square

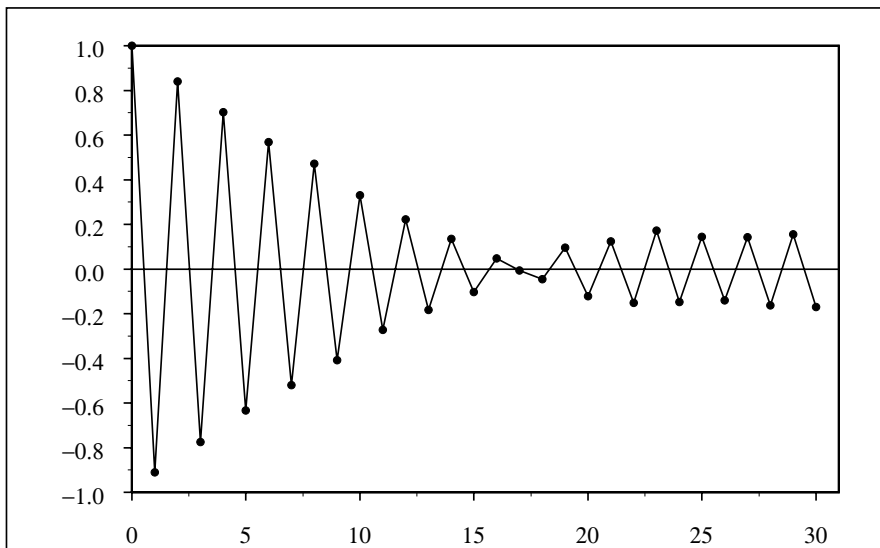


Abb. 2.8 Empirisches Korrelogramm der $AR(1)$ -Zeitreihe aus Abb. 2.7

Bemerkung 2.12

In Satz 8.7 werden wir sehen, dass die Aussage von Lemma 2.11, dass nämlich der Yule–Walker–Schätzer stets zu einem geschätzten Modell führt, welches stationär ist, auf allgemeine autoregressive Modelle übertragen werden kann.

□

Wir wollen diesen Abschnitt mit zwei konkreten Beispielen beschließen.

Beispiel 2.13

Man betrachte den folgenden $AR(1)$ -Prozess

$$X_t + 0.95 \cdot X_{t-1} = e_t, \quad \forall t \in \mathbb{Z},$$

mit einem standardnormalverteilten weißen Rauschen. Abbildung 2.7 zeigt für $n = 100$ eine Realisierung dieser Zeitreihe. Aus diesem Datensatz wurde ein empirisches Korrelogramm erstellt (vgl. Abbildung 2.8), welches im Vergleich zum tatsächlichen Korrelogramm aus Abbildung 2.5 zu sehen ist. In der Tat wurde in Abbildung 2.5 auch der Parameter $a = 0.95$ gewählt. Der Yule–Walker–Parameterschätzer \hat{a} für a ergibt sich aus den Yule–Walker–Gleichungen (2.21) für diesen Datensatz zu

$$\hat{a} = -\hat{\rho}(1) = 0.911.$$

Um die Güte des Yule–Walker–Schätzers im $AR(1)$ -Modell beurteilen zu können, reicht es natürlich nicht aus, eine einzige Realisierung zu betrachten. In Abbildung 2.9 ist für 1000 $AR(1)$ -Zeitreihen der Länge $n = 100$ mit

Parameter $a = 0.95$ ein Histogramm der 1000 resultierenden Yule–Walker–Schätzer wiedergegeben. Man erkennt, dass der Yule–Walker–Schätzer in diesem einfachen $AR(1)$ -Modell und beim Stichprobenumfang $n = 100$ sehr respektable Performance zeigt. Deutlich zu erkennen ist eine Linksschiefe des Histogramms, die zeigt, dass wir den tatsächlichen Parameter $a = 0.95$ in der Regel unterschätzen.

Es wird auch deutlich sichtbar, dass kein einziger der Yule–Walker–Schätzer

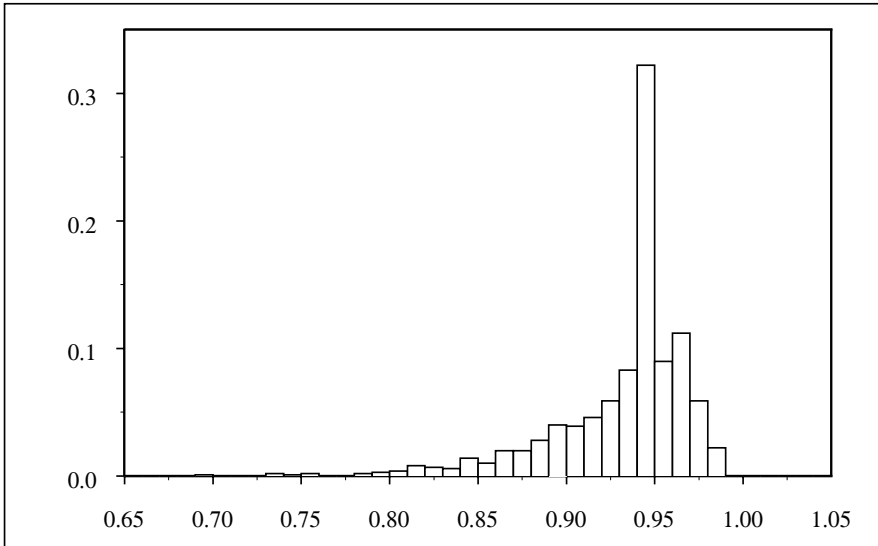


Abb. 2.9 Simuliertes Histogramm des Yule–Walker–Schätzers \hat{a} im $AR(1)$ -Modell mit Parameter $a = 0.95$ und Stichprobenumfang $n = 100$

größer oder gleich 1 ist, was in Übereinstimmung mit Lemma 2.11 steht. \square

Das folgende Beispiel ist eines der klassischen Beispiele der Zeitreihenanalyse und darf auch in dieser Einführung nicht fehlen.

Beispiel 2.14 (Sonnenfleckenzahlen)

Die Züricher Zeitreihe der relativen Sonnenfleckenzahlen (Zurich series of sunspot relative numbers) kann mit Fug und Recht als eine klassische Datenreihe der Zeitreihenanalyse angesehen werden. Die Sonnenflecken wurden 1610 entdeckt und insbesondere die Epochen der Maxima und der Minima können bis zu diesem Datum angegeben werden. Jährliche Mittelwerte der Sonnenfleckenzahlen gehen bis zum Beginn des 18. Jahrhunderts zurück. Die systematische Sammlung dieser Daten verbunden mit statistischen Untersuchungen ist *Johann Rudolf Wolf (1816–1893)* zu verdanken. Wolfs Interesse

an Sonnenfleckenbeobachtungen begann, als er 1847 Direktor des Observatoriums in Bern in der Schweiz wurde. Laut Izenman (1983) begann er genau am 4. Dezember 1847 mit regelmäßigen täglichen Beobachtungen und setzte diese auch später als Professor am Polytechnikum Zürich (der heutigen ETH Zürich) fort. 1848 hat Wolf als Maß für die Sonnenfleckenaktivität die sogenannten relativen Sonnenfleckenzahlen eingeführt. In seinem Testament

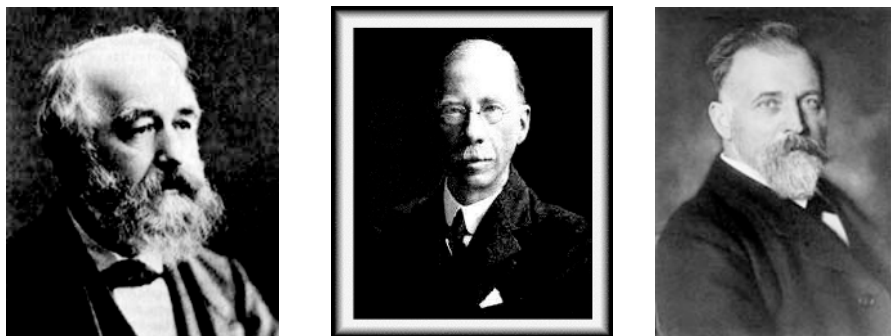


Abb. 2.10 Johann Rudolf Wolf (1816–1893), George Udny Yule (1871–1951) und Alfred Wolfer (1854–1931) (von links nach rechts)

hat Rudolf Wolf festgelegt, dass jährlich eine Ausgabe der *Astronomischen Mitteilungen* zu statistischen Arbeiten über die Fleckenhäufigkeit der Sonne herausgegeben werden solle. Der seinerzeitige Schüler und Assistent von Rudolf Wolf und spätere Direktor der Eidgenössischen Sternwarte in Zürich *Alfred Wolfer* (1854–1931) übernahm diesen Auftrag und hat 1925 eine komplette Revision der relativen Sonnenfleckenzahlen für den Zeitraum 1749–1924 publiziert (vgl. Wolfer (1925)). Diese Arbeit von Wolfer hat nun *Sir George Udny Yule* (1871–1951) zum Anlass genommen, um zum ersten Mal ein autoregressives Zeitreihenmodell anzuwenden (Yule (1927)). Yule war es auch, der den Namen *Wolfer's Sunspot Series* prägte und damit (sicher unbewusst) das ganz bestimmt weit größere Verdienst des akademischen Lehrers von Wolfer, nämlich von Johann Rudolf Wolf, für lange Zeit in den Hintergrund rückte. Yule erhielt damals aus den jährlichen Mittelwerten von 1749–1924 unter anderem das folgende $AR(2)$ -Zeitreihenmodell für die mit dem arithmetischen Mittel $\bar{X}_n = 44.3328$ zentrierten Sonnenfleckenzahlen X_t

$$X_t - 1.34254 \cdot X_{t-1} + 0.65504 \cdot X_{t-2} = e_t, \quad (2.33)$$

mit $E e_t^2 = 15.41^2$.

Eine wichtige Fragestellung bei der Untersuchung der Sunspot Numbers ist, ob eine Periodizität der Sonnenfleckenaktivität nachzuweisen ist. Bereits 1844

publizierte Schwabe (1844) die Vermutung, dass die Sonnenflecken eine gewisse Periodizität von etwa 10 Jahren aufwiesen. Wolf (1852) gelangte aufgrund ausführlicher Untersuchungen zu einer Periode von 11.11 Jahren und wies damals einen Standardfehler von 0.038 Jahren (also knapp 14 Tagen) ohne weitere Begründung für die von ihm angegebene Periode aus.

1961 war es der Direktor des Schweizer Observatoriums in Zürich, Professor *Waldmeier*, der eine vollständig überarbeitete und revidierte Liste der täglichen relativen Sonnenfleckenzahlen, sowie monatliche und jährliche Mittelwerte für den Zeitraum 1700–1960 veröffentlichte (Waldmeier (1961)).

Dem interessierten Leser sei der historische Artikel von Izenman (1983) und die Einleitung der Veröffentlichung von Waldmeier (1961) zur weiteren Lektüre empfohlen.

Heutzutage kann man die vollständigen Sonnenfleckenzahlen von der Internetseite des Solar Influences Data Analysis Center (SIDC), RWC Belgium, World Data Center for the Sunspot Index, Royal Observatory of Belgium (<http://sidc.oma.be>) erhalten. Wir geben an dieser Stelle (vgl. Abbildung 2.11) die jährlichen arithmetischen Mittelwerte der täglichen relativen Sonnenfleckenzahlen für den Zeitraum 1700–2004 aus der genannten Quelle wieder. In der Tat ist in Abbildung 2.11 in erster Näherung eine etwa 11-jährige

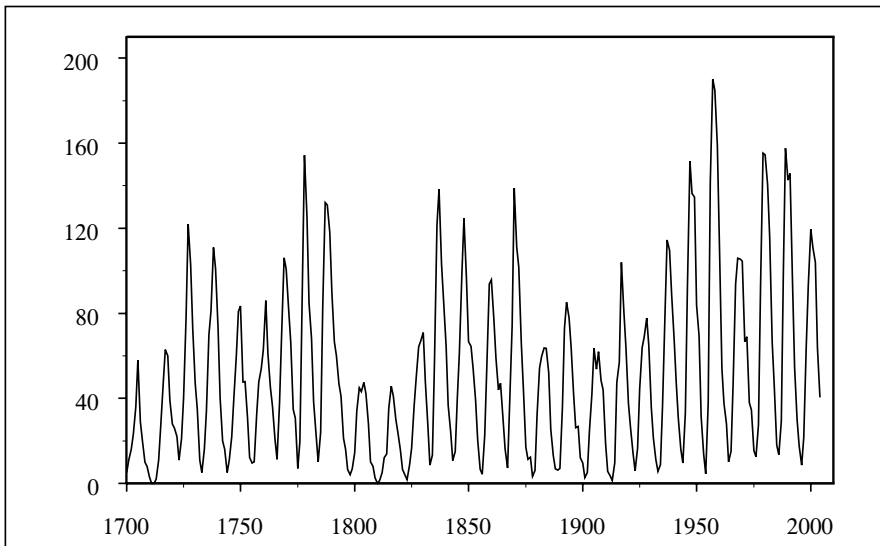


Abb. 2.11 Sunspot Relative Numbers, SIDC, RWC Belgium, World Data Center for the Sunspot Index, Royal Observatory of Belgium, 1700–2004

Periodizität zu erkennen. Aus den Daten erhalten wir gemäß (2.29) und (2.30) die folgenden empirischen Autokorrelationen

$$\hat{\varrho}(1) = 0.82043, \quad \hat{\varrho}(2) = 0.45021, \quad \hat{\varrho}(3) = 0.03861, \quad (2.34)$$

sowie den arithmetischen Mittelwert $\bar{X}_n = 50.22$ und die empirische Autokovarianz $\hat{\gamma}(0) = 1634.08$. Für die mit dem arithmetischen Mittel zentrierten Sonnenfleckendaten ergibt sich hieraus mit Hilfe von später zu behandelnden Parameterschätzverfahren für autoregressive Zeitreihenmodelle das folgende an die Daten angepasste $AR(2)$ -Modell

$$X_t - 1.380 \cdot X_{t-1} + 0.682 \cdot X_{t-2} = e_t \quad (2.35)$$

mit $E e_t^2 = 16.90^2$.

Wir werden später sehen, dass ein $AR(2)$ -Modell grundsätzlich in der Lage ist eine gewisse Periodizität darzustellen. Für das in (2.35) genannte Modell ergibt sich ein Zyklus von etwa 11.36 Jahren. Eine genauere Beschreibung der hier verwendeten Methoden zur Parameterschätzung geben wir in Kapitel 11. \square

2.4 Gaußsche Zeitreihen

In den bisherigen Überlegungen haben Annahmen an die Verteilungsstruktur von Zeitreihen noch keine Rolle gespielt. Dies wollen wir jetzt nachholen. Genau wie in der allgemeinen Stochastik liegt es nahe, an die Möglichkeit einer Normalverteilungsannahme zu denken. In der Tat findet man insbesondere in der klassischen Zeitreihenanalyse derartige Verteilungsannahmen, die zu sogenannten *Gaußschen* Zeitreihen führen.

Definition 2.15 (Gaußsche Zeitreihen)

Wir nennen eine reellwertige Zeitreihe $(X_t : t \in \mathbb{Z})$ eine *Gaußsche Zeitreihe*, falls für jedes $n \in \mathbb{N}$ und alle $t_1, \dots, t_n \in \mathbb{Z}$ gilt, dass die gemeinsame Verteilung von $(X_{t_1}, \dots, X_{t_n})$ eine n -dimensionale Normalverteilung mit Erwartungswertvektor $\mu(t_1, \dots, t_n) = (EX_{t_1}, \dots, EX_{t_n})^T$ und Kovarianzmatrix $\Sigma(t_1, \dots, t_n) = (\text{Cov}(X_{t_j}, X_{t_k}) : j, k = 1, \dots, n)$ besitzt. Wir benutzen für diese Aussage auch die folgende Schreibweise:

$$\mathcal{L}(X_{t_1}, \dots, X_{t_n}) = \mathcal{N}(\mu(t_1, \dots, t_n), \Sigma(t_1, \dots, t_n)).$$

Für stationäre reelle Zeitreihen ergibt sich, dass diese Verteilung vollständig durch den Erwartungswert $\mu = EX_t$ und die Autokovarianzfunktion $\gamma(h) = \text{Cov}(X_{t+h}, X_t)$, $h \in \mathbb{Z}$, beschrieben ist.

Wie gerade in der Definition 2.15 gesagt, legt der Erwartungswert und die Autokovarianzfunktion einer Gaußschen Zeitreihe diese vollständig fest. Wir werden später sehen (vgl. Satz 3.1), dass zu jeder möglichen Autokovarianzfunktion und zu jedem Erwartungswert stets eine stationäre Gaußsche Zeitreihe existiert, die genau diesen Erwartungswert und diese Autokovarianzfunktion besitzt.

Für ein *Gaußsches weißes Rauschen* $e = (e_t : t \in \mathbb{Z})$ erhalten wir aus bekannten Resultaten der Stochastik, dass die Zufallsvariablen e_s und e_t für $s \neq t$ nicht nur wie für ein weißes Rauschen vorausgesetzt unkorreliert, sondern stets unabhängig sind. Diese stochastische Unabhängigkeit überträgt sich auch auf Tupel $(e_{s_1}, \dots, e_{s_m})$ und $(e_{t_1}, \dots, e_{t_n})$, solange nur die Indexmengen $\{s_1, \dots, s_m\}$ und $\{t_1, \dots, t_n\}$ disjunkt sind.

Ferner wird als bekannt vorausgesetzt, dass eine lineare Transformation eines Zufallsvektors eine für diesen bestehende Normalverteilungsannahme grundsätzlich erhält. Genauer folgt für einen n -dimensionalen Zufallsvektor Z mit $\mathcal{L}(Z) = \mathcal{N}(\mu, \Sigma)$ unter der linearen Transformation $Az + b$ mit einer $m \times n$ -Matrix A und einem m -dimensionalen Vektor b die Aussage $\mathcal{L}(AZ + b) = \mathcal{N}(A\mu + b, A\Sigma A^T)$. Aus dieser Tatsache erhalten wir das folgende Resultat.

Bemerkung 2.16

Reelle Moving Average Zeitreihen der Ordnung q mit Gaußschem weißen Rauschen stellen selbst stets Gaußsche Zeitreihen dar.

Sogar für die $AR(1)$ -Zeitreihe aus Beispiel 2.7 gilt wegen der Darstellung $X_t = \sum_{\nu=0}^{\infty} (-a)^\nu e_{t-\nu}$ und der L_2 -Konvergenz der unendlichen Reihe (vgl. Beispiel 2.7), dass $X = (X_t : t \in \mathbb{Z})$ unter der Voraussetzung, dass ein Gaußsches weißes Rauschen vorliegt, selbst eine Gaußsche Zeitreihe ist, vgl. Aufgabe 11.1. \square

Es sei erwähnt, dass Gaußsche Zeitreihen im Rahmen der Zeitreihenanalyse keine überragende Rolle spielen. Was nicht bedeutet, dass eine Reihe von relevanten Kenngrößen (etwa die Autokovarianz) und gewisse später einzuführende Methoden von einer normalverteilten Situation herrühren oder in diesem Bereich eine besondere Bedeutung haben. Wir werden auf diesen Aspekt im Rahmen dieser Einführung in einigen Abschnitten wieder zu sprechen kommen.

2.5 Die partielle Autokorrelation

Neben der Autokovarianz- und der Autokorrelationsfunktion von stationären Zeitreihen wird zusätzlich in vielen Fällen auch die im Folgenden definierte *partielle Autokorrelationsfunktion* betrachtet. Im Gegensatz zur Autokorrelation $\varrho(h)$, die ein Maß für den linearen Zusammenhang von X_t und X_{t+h} darstellt, gibt die partielle Autokorrelation eine Maßzahl für den verbleibenden (linearen) Zusammenhang von X_t und X_{t+h} an, der über den Einfluss der zwischen diesen beiden Größen liegenden Zufallsvariablen $X_{t+1}, \dots, X_{t+h-1}$ hinausgeht. Betrachtet man unter diesem Aspekt stationäre $AR(1)$ -Zeitreihen, so besitzen diese zunächst eine nicht verschwindende Autokorrelation für alle Lags $h \geq 0$ (vgl. Beispiel 2.7), d.h. die Zufallsvariable X_t hat im Prinzip Einfluss auf *alle* zukünftigen Beobachtungen X_{t+h} . Aufgrund der Struktur

$X_t + aX_{t-1} = e_t$ einer $AR(1)$ -Zeitreihe und der obigen (zunächst nur verbalen) Erläuterung vermutet man aber vielleicht, dass X_t keinen über denjenigen von X_{t+h-1} hinausgehenden Einfluss auf X_{t+h} für $h \geq 2$ besitzt. Wir werden weiter unten sehen, dass dies tatsächlich der Fall ist. Vorher ist aber eine präzise Definition der partiellen Autokorrelation notwendig. Hierzu ist als Erstes festzulegen, wie der Einfluss der Zufallsvariablen $X_{t+1}, \dots, X_{t+h-1}$ auf X_t bzw. auf X_{t+h} zu quantifizieren ist. Dazu nehmen wir vereinfachend an, dass die vorliegende Zeitreihe reell und zentriert ist. Da wir mit Korrelationen grundsätzlich nur lineare Zusammenhänge erfassen können, soll dieser Einfluss mittels der Projektion $P_{\text{span}\{X_{t+1}, \dots, X_{t+h-1}\}}$ von X_t bzw. X_{t+h} auf den von $X_{t+1}, \dots, X_{t+h-1}$ aufgespannten linearen Teilraum $\text{span}\{X_{t+1}, \dots, X_{t+h-1}\}$ des $L_2^{\mathbb{R}}$ – also ebenfalls linear – gemessen werden. Bekanntlich können ja alle Elemente dieses linearen Teilraumes durch Linearkombinationen der Zufallsvariablen $X_{t+1}, \dots, X_{t+h-1}$ beschrieben werden. Eigentlich könnten die Zufallsvariablen auch komplex sein. Da die partielle Autokorrelation in diesem Fall in der Theorie und der Praxis bedeutungslos ist, verzichten wir hier auf diese Verallgemeinerung. Bevor wir uns Gedanken über die Berechnung der Projektionen machen, soll die partielle Autokorrelation aber erst einmal definiert werden.

Definition 2.17 (*Partielle Autokorrelation*)

$X = (X_t : t \in \mathbb{Z})$ sei eine reellwertige, stationäre und zentrierte Zeitreihe. Mit dem linearen Teilraum $H = \text{span}\{X_{t+1}, \dots, X_{t+h-1}\}$ von $L_2^{\mathbb{R}}$, wobei $H := \{0\}$ für $h = 1$ gesetzt werde, und der Projektion P_H auf H nennen wir die Größe

$$\pi(h) = \text{Corr}(X_{t+h} - P_H(X_{t+h}), X_t - P_H(X_t)), \quad h = 1, 2, \dots, \quad (2.36)$$

partielle Autokorrelation der Zeitreihe $X = (X_t : t \in \mathbb{Z})$ zum Lag h . Dass diese Größe für stationäre Zeitreihen in der Tat unabhängig von t ist, werden wir erst später beweisen. Für nicht zentrierte reelle und stationäre Zeitreihen definieren wir die partielle Autokorrelation als die partielle Autokorrelation der zugehörigen zentrierten Zeitreihe.

Wegen $P_{\{0\}}(X_s) = 0$ gilt zunächst

$$\pi(1) = \text{Corr}(X_{t+1}, X_t) = \varrho(1). \quad (2.37)$$

Im Kapitel 4 über Vorhersagen von Zeitreihen leiten wir eine Berechnungsmethode für partielle Autokorrelationen her, die wir schon hier benutzen wollen. Und zwar gilt unter der Voraussetzung, dass die Korrelationsmatrix $(\varrho(r-s) : r, s = 1, \dots, h)$ nicht singulär (also invertierbar) ist und somit die Lösung des linearen Gleichungssystems

$$\begin{pmatrix} \varrho(r-s) \\ r, s = 1, \dots, h \end{pmatrix} \begin{pmatrix} d_{h,1} \\ \vdots \\ d_{h,h} \end{pmatrix} = \begin{pmatrix} \varrho(1) \\ \vdots \\ \varrho(h) \end{pmatrix} \quad (2.38)$$

eindeutig bestimmt ist, für dieses h die Aussage

$$\pi(h) = d_{h,h}. \quad (2.39)$$

Es sei darauf hingewiesen, dass (2.38) nichts anderes darstellt als die Normalgleichungen für die Projektion von X_{t+h} auf den von den Zufallsvariablen X_t, \dots, X_{t+h-1} aufgespannten linearen Teilraum. Der Koeffizient $d_{h,h}$ von X_t in der Darstellung der Projektion, also die partielle Autokorrelation, gibt den linearen Einfluss von X_t (auf X_{t+h}) wieder (zumindest dann, wenn die Koeffizienten der Projektion eindeutig bestimmt sind). Diese Tatsache ist eine weitere Interpretationshilfe für die partielle Autokorrelation.

Die Darstellung (2.39) zusammen mit (2.38) und unter der Voraussetzung der Nicht-Singularität der empirischen Autokorrelationsmatrix $(\widehat{\varrho}(r-s) = \widehat{\gamma}(r-s)/\widehat{\gamma}(0) : r, s = 1, \dots, h)$ erlaubt darüberhinaus eine eingängige Definition der *empirischen partiellen Autokorrelation* zum Lag h gemäß

$$\widehat{\pi}(h) := \widehat{d}_{h,h} \quad (2.40)$$

mit $\widehat{d}_{h,h}$ aus

$$\begin{pmatrix} \widehat{\varrho}(r-s) \\ r, s = 1, \dots, h \end{pmatrix} \begin{pmatrix} \widehat{d}_{h,1} \\ \vdots \\ \widehat{d}_{h,h} \end{pmatrix} = \begin{pmatrix} \widehat{\varrho}(1) \\ \vdots \\ \widehat{\varrho}(h) \end{pmatrix}, \quad (2.41)$$

wobei $\widehat{\varrho}$ in (2.30) definiert wurde.

Nun kehren wir zum Abschluss des Abschnittes 2.5 und auch des Kapitels 2 noch einmal zum autoregressiven Prozess erster Ordnung zurück.

Beispiel 2.18 (Partielle Autokorrelation in $AR(1)$ -Zeitreihen)

Gegeben sei eine reellwertige, stationäre $AR(1)$ -Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit $X_t + aX_{t-1} = e_t$, wobei $e = (e_t : t \in \mathbb{Z})$ ein reelles, weißes Rauschen bezeichnet und $a \in \mathbb{R}$ mit $|a| < 1$ ist. In Abbildung 2.12 ist für eine Realisierung dieser $AR(1)$ -Zeitreihe ein sog. $X_{t-1} - X_t$ -Plot gegeben, d.h. eine Darstellung der Punkte (X_{t-1}, X_t) für $t = 2, \dots, n$. Man erkennt sehr gut, dass die eingezeichnete Gerade $y = a \cdot x$ mit $a = 0.8$ eine gute lineare Beschreibung dieser Daten liefert. Dies steht im Einklang mit der üblichen Regressionsanalyse. Die Steigung der aus dem $X_{t-1} - X_t$ -Plot gewonnenen Geraden liefert wegen $\text{Var}(X_{t-1}) = \text{Var}(X_t)$ einen Schätzer für die Autokorrelation zum Lag 1.

Wegen (2.37) und (2.18) gilt zunächst $\pi(1) = \varrho(1) = -a$.

Für höhere Lags $h \geq 2$ gehen wir von der Definition in Gleichung (2.36) aus und bestimmen eine Darstellung der Form $\sum_{j=1}^{h-1} c_j X_{t+h-j}$ für $P_H(X_{t+h})$ mit $H = \text{span}\{X_{t+1}, \dots, X_{t+h-1}\}$. Die Koeffizienten $c_j, j = 1, \dots, h-1$, stellen eine Lösung des linearen Gleichungssystems

$$\begin{pmatrix} \varrho(r-s) \\ r, s = 1, \dots, h-1 \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_{h-1} \end{pmatrix} = \begin{pmatrix} \varrho(1) \\ \vdots \\ \varrho(h-1) \end{pmatrix} \quad (2.42)$$

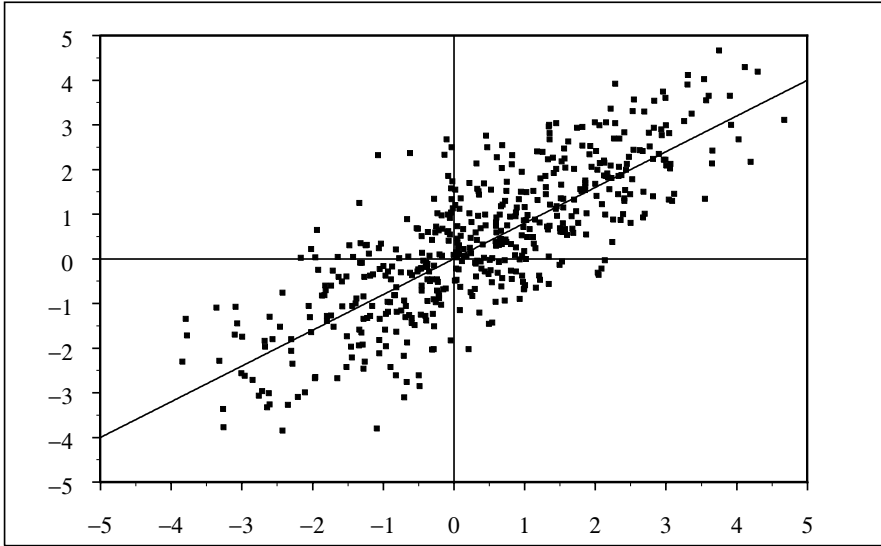


Abb. 2.12 $X_{t-1} - X_t$ -Plot einer Realisierung der Länge $n = 500$ einer $AR(1)$ -Zeitreihe mit Parameter $a = -0.80$ und die Gerade $y = 0.8x$

dar (vgl. (2.38)). Mit der Darstellung der Autokorrelationsfunktion einer $AR(1)$ -Zeitreihe aus (2.18) erhalten wir mit leichter Rechnung, dass $c_1 = -a$ und $c_2 = \dots = c_{h-1} = 0$ eine Lösung von (2.42) darstellt. Damit gilt also für $h \geq 2$

$$P_{\text{span}\{X_{t+1}, \dots, X_{t+h-1}\}}(X_{t+h}) = -aX_{t+h-1}. \quad (2.43)$$

Da $X_{t+h} + aX_{t+h-1} = e_{t+h}$ gilt und $P_{\text{span}\{X_{t+1}, \dots, X_{t+h-1}\}}(X_t)$ auf jeden Fall eine Linearkombination von $X_{t+1}, \dots, X_{t+h-1}$ ist, die wegen der $MA(\infty)$ -Darstellung der $AR(1)$ -Zeitreihe (vgl. (2.16)) linear von $e_s, s \leq t+h-1$, abhängt, folgt aufgrund der Unkorreliertheit des weißen Rauschens direkt $\pi(h) = \text{Corr}(e_{t+h}, P_{\text{span}\{X_{t+1}, \dots, X_{t+h-1}\}}(X_t)) = 0$.

Bereinigt man die Werte X_{t+2} bzw. X_t um die jeweiligen Projektionen auf den von X_{t+1} aufgespannten Teilraum (mit aus den Daten geschätztem Koeffizienten), so erhält man Abbildung 2.13. Man erkennt sehr deutlich, dass es sich hier um eine mehr oder weniger unkorrelierte Situation handelt (im Gegensatz zu Abbildung 2.12).

Die Eigenschaft, dass die partielle Autokorrelation für einen autoregressiven Prozess erster Ordnung ab Lag 2 verschwindet, können wir vorteilhaft zur Identifikation eines solchen Modells verwenden, indem wir die empirische partielle Autokorrelation betrachten und schauen, ob sich nur die erste partielle Autokorrelation signifikant von 0 unterscheidet.

Wir werden später sehen, dass sich diese Eigenschaft der partiellen Autokorrelation auf autoregressive Prozesse höherer Ordnung ausdehnt. Es gilt nämlich,

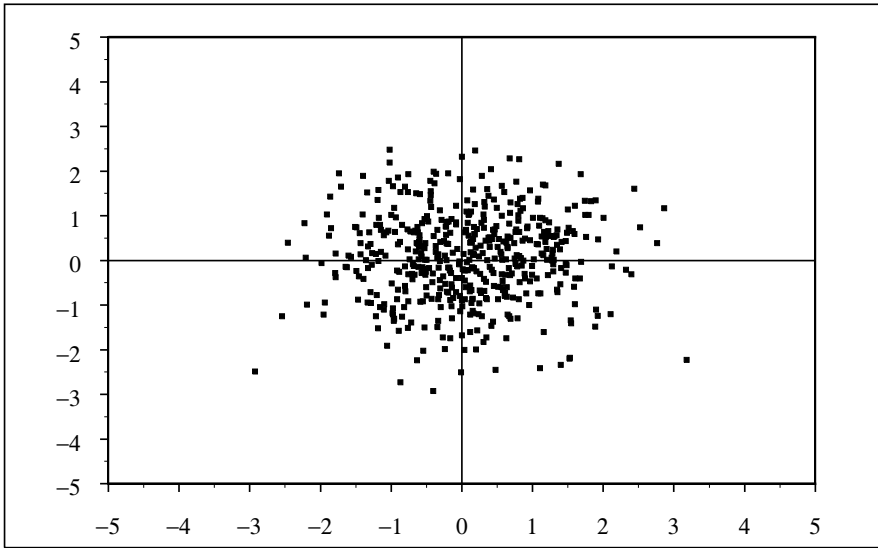


Abb. 2.13 Plot der Punkte $(X_t + \hat{a}X_{t+1}, X_{t+2} + \hat{a}X_{t+1})$

dass für einen stationären $AR(p)$ -Prozess die partiellen Autokorrelationen für Lags $h \geq p + 1$ verschwinden. Die partielle Autokorrelation wird in den Anwendungen dann auch tatsächlich zur Selektion autoregressiver Zeitreihen eingesetzt.

Wir wollen dies an dieser Stelle für die in Beispiel 2.14 betrachteten Sonnenfleckenzahlen tun. Aus den zentrierten Sonnenfleckenzahlen erhalten wir gemäß (2.40) die folgenden empirischen partiellen Autokorrelationen

$$\begin{aligned}\hat{\pi}(1) &= 0.82085, \hat{\pi}(2) = -0.68188, \\ \hat{\pi}(3) &= -0.13517 \text{ und } \hat{\pi}(4) = 0.05506.\end{aligned}\tag{2.44}$$

Genau wie Yule (1927) ist man auf dieser Basis in der Tat geneigt, eine $AR(2)$ - oder eine $AR(3)$ -Zeitreihe an die Sonnenfleckendaten anzupassen. \square

Aufgaben

Aufgabe 2.1 Man zeige in Beispiel 2.7, dass $\{\sum_{\nu=0}^q (-a)^\nu e_{t-\nu} : q \in \mathbb{N}\}$ für jedes $t \in \mathbb{Z}$ eine Cauchy-Folge ist, falls $|a| < 1$ gilt.

Aufgabe 2.2 Gegeben sei eine zyklische Zeitreihe

$$X_t = e_1 \cos(\omega t) + e_2 \sin(\omega t)\tag{2.45}$$

bzw. in der äquivalenten Darstellung $X_t = A \sin(\omega t + \phi)$ mit dem Zusammenhang

$$e_1 = A \sin \phi, \quad e_2 = A \cos \phi.$$

Man zeige: Falls e_1 und e_2 stochastisch unabhängige $\mathcal{N}(0, 1)$ -verteilte Zufallsvariable sind, so sind A und ϕ stochastisch unabhängig mit

$$\phi \sim \mathcal{R}(-\pi, \pi), \quad A^2 \sim \chi_2^2.$$

Dabei bezeichnet $\mathcal{R}(a, b)$ die Rechteck- oder Gleichverteilung auf dem Intervall (a, b) und χ_k^2 die Chi-Quadrat-Verteilung mit k Freiheitsgraden.

Hinweis: Transformationssatz für Dichten.

Aufgabe 2.3 Man untersuche den folgenden Prozess auf schwache Stationarität

$$\begin{aligned} X_t &:= a \cdot X_{t-1} + e_t, & t = 1, 2, \dots \\ X_0 &:= e_0 \end{aligned}$$

mit $a \in \mathbb{R}$ und einem reellen weißen Rauschen $e = (e_t : t \in \mathbb{Z})$, $e_t \sim (0, \sigma_e^2)$. Man diskutiere das Ergebnis.

Aufgabe 2.4 Sei $e = (e_t : t \in \mathbb{Z})$ ein weißes Rauschen mit $e_t \sim (0, \sigma_e^2)$. Man zeige, dass $X_t = \sum_{j=0}^{\infty} (-a)^j e_{t-2j}$, $t \in \mathbb{Z}$, eine Lösung der linearen Differenzgleichung zweiter Ordnung, vgl. Beispiel 2.7,

$$X_t + aX_{t-2} = e_t, \quad \forall t \in \mathbb{Z},$$

ist, wobei $a \in \mathbb{C}$ mit $|a| < 1$ gelte. Man zeige weiter, dass die Autokovarianzfunktion gleich

$$\gamma(h) = \sigma_e^2 \frac{(-a)^{h/2}}{1 - |a|^2} \quad \text{für gerade } h \in \mathbb{N}_0$$

und für ungerade $h \in \mathbb{N}_0$ gleich 0 ist. Für negative h gilt natürlich wie stets $\gamma(h) = \overline{\gamma(-h)}$.

Aufgabe 2.5 (Spezielles Kronecker-Lemma)

Seien $c_k \in \mathbb{C}$, $k \in \mathbb{Z}$, absolut summierbar. Man zeige mit Hilfe des Satzes von Lebesgue (siehe Anhang A.2)

$$\frac{1}{n} \sum_{|k| \leq n} |kc_k| \rightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (2.46)$$

Hinweis: Sei μ das abzählende Maß auf \mathbb{Z} . Damit ist die linke Seite in (2.46) gleich $\int f_n d\mu$ mit $f_n(k) = \frac{1}{n} 1(|k| \leq n) |kc_k|$.

Aufgabe 2.6 Nach Bemerkung 2.16 sind $MA(q)$ -Reihen mit Gaußschem weißen Rauschen selbst Gaußsche Zeitreihen.

Man berechne die Ausdrücke $\mu(t_t, \dots, t_n)$ und $\Sigma(t_t, \dots, t_n)$ aus Definition 2.15 für eine $MA(2)$ -Reihe mit Parametern $b_1, b_2 \in \mathbb{R}$ und stochastisch unabhängigen normalverteilten Zufallsvariablen $e_t \sim \mathcal{N}(0, \sigma_e^2)$.

Aufgabe 2.7 Seien $X = (X_t : t \in \mathbb{Z})$ und $Y = (Y_t : t \in \mathbb{Z})$ zwei unkorrelierte, stationäre Zeitreihen, d.h. $Cov(X_s, Y_t) = 0 \ \forall s, t$. Weiter seien die Autokovarianzen von X und Y beide gleich γ .

Man zeige, dass dann auch $(X + Y)/\sqrt{2} := ((X_t + Y_t)/\sqrt{2} : t \in \mathbb{Z})$ und $(X - Y)/\sqrt{2} := ((X_t - Y_t)/\sqrt{2} : t \in \mathbb{Z})$ unkorrelierte, stationäre Zeitreihen sind, die ebenfalls beide die Autokovarianz γ besitzen.

Aufgabe 2.8 Seien die Zufallsvariablen U und V gemeinsam normalverteilt mit Erwartungswert 0, gleicher Varianz σ^2 und Korrelationskoeffizient ϱ .

a) Man zeige

$$P\{U > 0, V > 0\} = \frac{1}{4} + \frac{1}{2\pi} \arcsin(\varrho).$$

b) Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, Gaußsche, zentrierte, stationäre Zeitreihe mit Autokorrelationsfunktion ϱ_X . Man definiere die Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ gemäß $Y_t = 1(X_t > 0)$ und zeige für ihre Autokorrelationsfunktion ϱ_Y

$$\varrho_Y(h) = \frac{2}{\pi} \arcsin(\varrho_X(h)).$$

Aufgabe 2.9

a) Sei $X = (X_t : t \in \mathbb{Z})$ eine $MA(1)$ -Zeitreihe. Man zeige für die zugehörige Autokorrelationsfunktion ϱ_X die Ungleichung $\varrho_X(1) \leq 0.5$.

b) Sei $X = (X_t : t \in \mathbb{Z})$ eine $MA(q)$ -Zeitreihe. Man zeige entsprechend $\varrho_X(1) \leq \cos(\pi/(2+q))$.

c) Gibt es eine $MA(2)$ -Zeitreihe X mit $\varrho_X(1) = 0.8$?

Aufgabe 2.10 Seien e_1, e_2, \dots stochastisch unabhängig mit $e_t \sim \mathcal{N}(0, \sigma_e^2) \ \forall t \in \mathbb{N}$. Der rekursiv definierte Prozess $(X_t : t \geq 1)$ gemäß $X_t = X_{t-1} + e_t$ mit Startzufallsvariable X_0 , die unkorreliert mit den e_t sei, heißt *Random Walk*.

a) Man berechne EX_t und $VarX_t$ und folgere daraus, dass (X_t) kein stationärer Prozess ist.

b) Angenommen, man plottet simulierte Realisierungen eines Random Walk X_0, X_1, \dots, X_{200} mit $\sigma_e^2 = 1$. Wie sollte man den Bildbereich für (X_t) wählen, damit möglichst alle Pfade vollständig zu sehen sind?

Die Autokovarianz und die Autokorrelation

In diesem Kapitel wollen wir uns näher mit der Autokovarianz- und Autokorrelationsfunktion von Zeitreihen auseinander setzen. In Kapitel 2 haben wir bereits gesehen, dass diese Funktionen wesentliche Kenngrößen für Zeitreihen sind, und wir haben dort auch bereits für eine ganze Reihe von Beispielen die Autokovarianzfunktion und damit die Autokorrelationsfunktion berechnet. Im Folgenden sollen nun charakterisierende Eigenschaften dieser wichtigen Kenngrößen hergeleitet werden. Wir werden im Abschnitt 3.1 zunächst aus der positiven Semidefinitheit einer jeden Autokovarianzfunktion grundlegende Eigenschaften ableiten, bevor wir im Abschnitt 3.2 den Begriff des Spektralmaßes und der Spektraldichte einer Zeitreihe, sowie deren Zusammenhang mit der Autokovarianz bzw. Autokorrelation kennen lernen. Um in nachfolgenden Kapiteln keine Probleme mit der Gültigkeit entsprechender Resultate auch für komplexwertige Zeitreihen zu bekommen, wird hier grundsätzlich eine komplexwertige stationäre Zeitreihe unterstellt. Der Mehraufwand ist gering und erspart späteres Nachbessern.

3.1 Grundlegende Eigenschaften

Sei $X = (X_t : t \in \mathbb{Z})$ eine komplexwertige und (schwach) stationäre Zeitreihe mit $EX_t = \mu$ und Autokovarianzfunktion

$$\gamma(h) = \gamma_X(h) = E(X_{t+h} - \mu) \overline{(X_t - \mu)}, \quad \forall h, t \in \mathbb{Z}, \quad (3.1)$$

bzw. Autokorrelationsfunktion

$$\varrho(h) = \varrho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)}, \quad \forall h \in \mathbb{Z}, \quad (3.2)$$

wobei, wie schon früher erwähnt, der Erwartungswert einer komplexwertigen Zufallsvariablen $Y = U + iV$ mit Realteil U und Imaginärteil V gemäß $EY = EU + iEV$ definiert ist, siehe Anhang A.2.

Man spricht auch einfach von der Kovarianz- bzw. Korrelationsfunktion. Fasst man X_t als Element des komplexen Hilbertraums $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ der komplexwertigen und quadratintegrierbaren Zufallsvariablen auf (Ω, \mathcal{A}, P) mit dem inneren Produkt $\langle X, Y \rangle = EX\bar{Y}$ auf, siehe Anhang A.2, so gilt

$$\gamma(h) = \langle X_{t+h} - \mu, X_t - \mu \rangle$$

sowie

$$\varrho(h) = \langle X_{t+h} - \mu, X_t - \mu \rangle / \|X_t - \mu\|^2.$$

Jede Kovarianzfunktion γ ist (komplex) *positiv semidefinit*, d.h. es gilt

$$\sum_{i,j=1}^r \lambda_i \bar{\lambda}_j \gamma(t_i - t_j) \geq 0 \quad (3.3)$$

für alle $\lambda_1, \dots, \lambda_r \in \mathbb{C}$, $t_1, \dots, t_r \in \mathbb{Z}$, $r \in \mathbb{N}$. Gilt in (3.3) das strikte Größerzeichen sobald nicht alle λ_j verschwinden, so heißt die Autokovarianzfunktion γ (komplex) *positiv definit*. Zum Nachweis von (3.3) nehme man o.B.d.A. $\mu = 0$ an. Wegen $\gamma(t_i - t_j) = EX_{t_i} \bar{X}_{t_j}$ ist die linke Seite von (3.3) gleich $E \left| \sum_{i=1}^r \lambda_i X_{t_i} \right|^2$ und damit stets reell und größer gleich Null.

Satz 3.1 (*Charakterisierung von Autokovarianzfunktionen*)

Zu jeder (komplex) positiv semidefiniten Funktion $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ gibt es einen stationären, zentrierten stochastischen Prozess mit Autokovarianzfunktion γ . Dieser Prozess kann reell gewählt werden, falls die Autokovarianzfunktion γ reellwertig ist.

Beweis: Siehe Ash/Gardner (1975), Theorem 1.2.3. □

Unter Verwendung dieses Satzes erhält man leicht die im folgenden Lemma aufgeführten Eigenschaften von Kovarianzfunktionen, die wir aber direkt aus der positiven Semidefinitheit herleiten wollen.

Lemma 3.2 (*Grundeigenschaften von Autokovarianzfunktionen*)

Sei $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ eine (komplex) positiv semidefinite Funktion. Dann gilt für jedes $h \in \mathbb{Z}$

$$\gamma(0) \geq 0, \quad (3.4)$$

$$\gamma(-h) = \overline{\gamma(h)} \quad (\text{Hermite-}) \text{ Symmetrie} \quad (3.5)$$

$$|\gamma(h)| \leq \gamma(0). \quad (3.6)$$

Beweis: Zum Nachweis von (3.4) setze man in (3.3) $r = 1$, $\lambda_1 = 1$ und $t_1 = 0$ ein.

Für (3.5) verwende man in (3.3) $r = 2$, $\lambda_1 = \lambda_2 = 1$, $t_1 = 0$ und $t_2 = h$, woraus zunächst $2\gamma(0) + \gamma(-h) + \gamma(h) \geq 0$ folgt. Da $\gamma(0)$ reell ist, folgt

dass auch $\gamma(-h) + \gamma(h)$ reell sein muss. Setzt man sodann in (3.3) $r = 2, \lambda_1 = i, \lambda_2 = 1, t_1 = 0$ sowie $t_2 = h$ ein, so folgt ganz entsprechend $2\gamma(0) + i(\gamma(-h) - \gamma(h)) \geq 0$. Also ist $\gamma(-h) - \gamma(h)$ rein imaginär. Zusammen folgt (3.5).

Es bleibt (3.6) zu zeigen. Sei o.B.d.A. $\gamma(h) \neq 0$ angenommen. Mit $r = 2, \lambda_1 = 1, \lambda_2 = -x/\overline{\gamma(h)}, t_1 = h, t_2 = 0$ und $x \in \mathbb{R}$, folgt aus (3.3)

$$\gamma(0) - (x/\gamma(h))\gamma(h) - (x/\overline{\gamma(h)})\gamma(-h) + (x^2/|\overline{\gamma(h)}|^2)\gamma(0) \geq 0,$$

d.h. für alle $x \in \mathbb{R}$

$$\begin{aligned} 0 &\leq \gamma(0) - 2x + x^2\gamma(0)/|\gamma(h)|^2 \\ &= \frac{\gamma(0)}{|\gamma(h)|^2} \left\{ \left(x - \frac{|\gamma(h)|^2}{\gamma(0)} \right)^2 + |\gamma(h)|^2 \left(1 - \frac{|\gamma(h)|^2}{\gamma(0)^2} \right) \right\}. \end{aligned}$$

Für $x = |\gamma(h)|^2/\gamma(0)$ ergibt sich schließlich $1 - |\gamma(h)|^2/\gamma(0)^2 \geq 0$, woraus sofort die Behauptung folgt. \square

Bemerkung 3.3 (Komplexe und reelle positive Semidefinitheit)

In der Definition der positiven Semidefinitheit von γ in (3.3) ist es wesentlich, dass komplexe λ_i zugelassen werden, um die (Hermite-) Symmetrie folgern zu können. Würde man die Ungleichung (3.3) nur für reelle λ_i verlangen (diese Eigenschaft nennen wir reelle positive Semidefinitheit), so könnte daraus die Symmetrie nicht mehr gefolgert werden.

Ist allerdings $\gamma: \mathbb{Z} \rightarrow \mathbb{R}$ reell positiv semidefinit bzw. reell positiv definit *und* symmetrisch um 0, so ist γ auch komplex positiv semidefinit bzw. komplex positiv definit wie eine leichte Rechnung zeigt. \square

3.2 Spektralmaß und Spektraldichte

Wir kommen nun zu einer weiteren Charakterisierung von Kovarianzfunktionen, die ein erster Schritt in Richtung auf die Spektraltheorie stationärer Zeitreihen ist.

Satz 3.4 (Herglotz-Lemma und Spektralmaß)

Sei $\gamma: \mathbb{Z} \rightarrow \mathbb{C}$ eine positiv semidefinite Abbildung. Dann gibt es ein eindeutig bestimmtes, endliches Maß μ auf $((-\pi, \pi], (-\pi, \pi] \cap \mathbb{B})$ mit

$$\gamma(h) = \int_{(-\pi, \pi]} e^{ih\omega} d\mu(\omega), \quad \forall h \in \mathbb{Z}. \quad (3.7)$$

Umgekehrt ist jede Funktion $\gamma: \mathbb{Z} \rightarrow \mathbb{C}$ mit einer Darstellung (3.7) positiv semidefinit. Das Maß μ nennt man das Spektralmaß von γ .

Besitzt das Spektralmaß μ eine Lebesgue-Dichte f auf $(-\pi, \pi]$, d.h. es gilt mit dem Lebesgue-Maß λ auf \mathbb{R} für alle $B \in (-\pi, \pi] \cap \mathcal{B}$ die Gleichung $\mu(B) = \int_B f \, d\lambda$, so ist f bekanntlich λ -fast sicher eindeutig auf $(-\pi, \pi]$ bestimmt und heißt *Spektraldichte* von γ . Durch eine beliebige Festlegung von $f(-\pi)$ wird keine Punktmasse in $-\pi$ erzeugt und μ kann im Falle einer existierenden Spektraldichte auch als Maß auf $[-\pi, \pi]$ und f als Dichte auf $[-\pi, \pi]$ betrachtet werden. Aus Symmetriegründen betrachtet man bei existierenden Spektraldichten f diese dann meist auf $[-\pi, \pi]$.

Im Falle, dass μ einen diskreten Anteil mit Punktmasse in π besitzt, ist für die Eindeutigkeit des Spektralmaßes die Einschränkung auf das halboffene Intervall $(-\pi, \pi]$ jedoch wesentlich, wie wir im Beweis zu Satz 3.4 sehen werden!

Bemerkung 3.5

Wir verwenden das Symbol μ in diesem Buch zum einen als Erwartungswert einer Zeitreihe (vgl. (3.1)) und zum anderen als Spektralmaß (vgl. (3.7)). Da diese Bezeichnungen in der Literatur üblich sind, wollen wir an ihnen festhalten und denken, dass aus dem Zusammenhang stets deutlich wird, ob μ den Erwartungswert oder das Spektralmaß einer Zeitreihe bezeichnet. Mit dem Spektralmaß einer stationären Zeitreihe meinen wir dabei das Spektralmaß der zur Zeitreihe gehörenden Autokovarianzfunktion. \square

Beweis: (zu Satz 3.4)

Es gelte (3.7). Wegen $\gamma(k-j) = \int_{(-\pi, \pi]} e^{i(k-j)\omega} d\mu(\omega)$ erhalten wir die positive Semidefinitheit wie folgt:

$$\sum_{k,j=1}^r \lambda_k \overline{\lambda_j} \gamma(t_k - t_j) = \int_{(-\pi, \pi]} \left| \sum_{k=1}^r \lambda_k e^{it_k \omega} \right|^2 d\mu(\omega) \geq 0,$$

$$\forall \lambda_1, \dots, \lambda_r \in \mathbb{C}, \quad \forall t_1, \dots, t_r \in \mathbb{Z}, \quad \forall r \in \mathbb{N}.$$

Für die umgekehrte Richtung können wir annehmen, dass $\gamma(0) > 0$ ist, da für $\gamma(0) = 0$ die Aussage (3.7) mit dem Maß $\mu \equiv 0$ erfüllt ist. Mit $\varrho(k) := \gamma(k)/\gamma(0)$ definieren wir nun reelle Funktionen $f_n : [-\pi, \pi] \rightarrow \mathbb{R}$, $n \geq 1$, durch

$$\begin{aligned} f_n(\omega) &:= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \varrho(k) e^{-ik\omega} \\ &= \frac{1}{2\pi n} \sum_{j=1}^n \sum_{m=1}^n \varrho(j-m) e^{-ij\omega} e^{im\omega} \geq 0, \end{aligned} \quad (3.8)$$

wobei die letzte Gleichheit aus der Summationsformel (A.11) des Anhangs folgt und die Nichtnegativität der Funktionen f_n eine direkte Konsequenz der positiven Semidefinitheit von γ und damit von ϱ ist. Man erhält für $h \in \mathbb{Z}$

$$\begin{aligned}
\int_{-\pi}^{\pi} f_n(\omega) e^{ih\omega} d\omega &= \frac{1}{2\pi} \sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \varrho(k) \int_{-\pi}^{\pi} e^{i(h-k)\omega} d\omega \\
&= \left(1 - \frac{|h|}{n}\right) \varrho(h) 1(|h| < n),
\end{aligned} \tag{3.9}$$

da das letzte Integral für $k \neq h$ verschwindet und für $k = h$ den Wert 2π besitzt. Insbesondere erhalten wir also für $h = 0$ die Gleichheit

$$\int_{-\pi}^{\pi} f_n(\omega) d\omega = \varrho(0) = 1.$$

Mit dem Lebesgue-Maß λ definieren wir nun Wahrscheinlichkeitsmaße μ_n auf $[-\pi, \pi]$ durch $\mu_n(B) = \int_B f_n d\lambda \quad \forall B \in [-\pi, \pi] \cap \mathbb{B}$. Nach dem Satz von Prohorov (siehe Anhang A.5) gibt es stets eine Teilfolge $\{\mu_{n(k)} : k \in \mathbb{N}\}$ und ein Wahrscheinlichkeitsmaß μ_0 auf $[-\pi, \pi]$ mit $\mu_{n(k)} \xrightarrow{\mathcal{D}} \mu_0$ für $k \rightarrow \infty$, wobei $\xrightarrow{\mathcal{D}}$ die Verteilungskonvergenz bezeichnet, siehe Anhang A.5. Es folgt aus (3.9) und der Verteilungskonvergenz

$$\begin{aligned}
\frac{\gamma(h)}{\gamma(0)} &= \lim_{k \rightarrow \infty} \int_{[-\pi, \pi]} e^{ih\omega} d\mu_{n(k)}(\omega) = \int_{[-\pi, \pi]} e^{ih\omega} d\mu_0(\omega) \\
&= \int_{(-\pi, \pi]} e^{ih\omega} d\mu_0(\omega) + e^{-ih\pi} \mu_0(\{-\pi\}) \\
&= \frac{1}{\gamma(0)} \int_{(-\pi, \pi]} e^{ih\omega} d\mu(\omega),
\end{aligned}$$

wobei $\mu(B) := (\mu_0(B \cap (-\pi, \pi]) + \mu_0(\{-\pi\}) 1_B(\pi)) \gamma(0)$ gesetzt wurde. Damit gilt (3.7). Es bleibt die Eindeutigkeit von μ zu zeigen. Dazu verwenden wir, dass sich jede stetige Funktion $f : [-\pi, \pi] \rightarrow \mathbb{C}$ mit $f(-\pi) = f(\pi)$ gleichmäßig auf $[-\pi, \pi]$ durch ein trigonometrisches Polynom $P(t) = \sum_{n=-N}^N c_n e^{int}$ approximieren lässt, siehe etwa Rudin (1970), Theorem 4.25. Für zwei Maße μ_1, μ_2 mit (3.7) folgt aus $\int_{(-\pi, \pi]} P d\mu_1 = \int_{(-\pi, \pi]} P d\mu_2$ also auch $\int_{(-\pi, \pi]} f d\mu_1 = \int_{(-\pi, \pi]} f d\mu_2$. Da man jede Indikatorfunktion $1_{(a, b]}$ mit $-\pi < a < b < \pi$ durch Folgen stetiger Funktionen $f_r, r \geq 1$, mit $|f_r| \leq 1$ und $f_r(-\pi) = f_r(\pi)$ approximieren kann gemäß $f_r \rightarrow 1_{(a, b]}$, folgt mit dem Satz von Lebesgue $\mu_1((a, b]) = \mu_2((a, b])$. Der Eindeutigkeitsatz für Maße liefert dann $\mu_1 = \mu_2$. \square

Reelle zyklische Zeitreihen, die wir bereits in Definition 2.9 kennen gelernt haben, eignen sich besonders, um die Bedeutung des Spektralmaßes besser zu verstehen, wie das folgende Beispiel zeigt.

Beispiel 3.6 (Zyklische Zeitreihen)

Sei

$$X_t = e_1 \cos(\omega t) + e_2 \sin(\omega t), \quad \forall t \in \mathbb{Z}, \quad (3.10)$$

eine zyklische Zeitreihe mit reellen, unkorrelierten, zentrierten Zufallsvariablen e_1, e_2 , beide mit Varianz $\sigma^2 > 0$, vgl. Definition 2.9. Da sich X_t nicht ändert, wenn man ω um ein Vielfaches von 2π verschiebt und Änderung des Vorzeichens von ω nur zu einer Vorzeichenänderung von e_2 führt, nehmen wir o.B.d.A. an, dass $\omega \in (0, \pi]$ gilt. Jeder Pfad des Prozesses $X = (X_t : t \in \mathbb{Z})$ ist eine periodische Funktion, deren Perioden- oder Wellenlänge wir λ nennen wollen. Dann ist die Anzahl der Perioden, die in eine Zeiteinheit fallen, also die Frequenz, gleich $\nu := 1/\lambda$, und somit ist $\omega = 2\pi\nu$ die Frequenz in Kreisläufen (Radian) gerechnet, woraus sich der Name Kreisfrequenz für ω erklärt. Während für mathematische Darstellungen die Kreisfrequenz besser geeignet ist, lässt sich die Frequenz einfacher interpretieren. Ist die Periodenlänge z.B. $\lambda = 12$ Monate bei monatlichen Erhebungen, so ist die zugehörige Kreisfrequenz $\omega = 2\pi/12 = \pi/6$, während die Frequenz einfach $\nu = 1/12$ ist. In der Sprechweise folgen wir vielen Autoren und bezeichnen, wenn der Zusammenhang klar ist, ω selbst als Frequenz.

Nach (2.24) ist X stationär mit Kovarianzfunktion

$$\begin{aligned} \gamma(h) &= \sigma^2 \cos(\omega h) \\ &= \frac{\sigma^2}{2} (e^{i\omega h} + e^{-i\omega h}). \end{aligned}$$

Folglich ist das zu X gehörige Spektralmaß μ ein zum Nullpunkt symmetrisches Zweipunktmaß mit Trägerpunkten $\pm\omega$ gemäß

$$\mu(B) = \frac{\sigma^2}{2} (1_B(\omega) + 1_B(-\omega)),$$

falls $\omega \in (0, \pi)$ und das Einpunktmaß $\mu(B) = \sigma^2 1_B(\pi)$, falls $\omega = \pi$ gilt.

Die Tatsache, dass die Kreisfrequenz im Herglotz-Lemma und im vorliegenden Beispiel maximal den Wert π annimmt, ist eine Konsequenz der Diskrettheit der Beobachtungszeitpunkte. Offenbar können dann nur Periodenlängen von minimal $\lambda = 2$ Zeiteinheiten erkannt werden, was einer Kreisfrequenz von $\omega = 2\pi/\lambda = \pi$ entspricht. Würde man die zyklische Zeitreihe (3.10) an den Zeitpunkten $t\Delta$ beobachten für ein $\Delta > 0$, so wären Periodenlängen von minimal $\lambda = 2\Delta$ zu entdecken, was zur maximal erkennbaren Kreisfrequenz (*Nyquist-Frequenz*) $\omega = \pi/\Delta$ führt. Bei zeitlich immer dichteren Beobachtungspunkten, also $\Delta \rightarrow 0$, wächst die Nyquist-Frequenz gegen ∞ . Tatsächlich gilt ein Analogon zum Herglotz-Lemma bei zeitkontinuierlicher Beobachtung mit Integrationsbereich $(-\infty, \infty)$ (Satz von Bochner, siehe z.B. Ash/Gardner (1975), Theorem 1.2.7).

Wie wir in Kapitel 5 sehen werden, kann man beliebige stationäre Zeitreihen

als Überlagerung (entweder als Summe oder im allgemeinen Fall als Integral) unkorrelierter zyklischer Zeitreihen darstellen, wobei im reellen Fall die Varianz jeder zyklischen Komponente zur Frequenz ω das Gewicht der Kreisfrequenzen $\pm\omega$ im zugehörigen Spektralmaß ist. \square

Die Symmetrie des Spektralmaßes gilt bei reellen Zeitreihen allgemein, wie die folgende Bemerkung zeigt.

Bemerkung 3.7 (Symmetrie des Spektralmaßes)

Ist $\gamma: \mathbb{Z} \rightarrow \mathbb{R}$ eine reelle Kovarianzfunktion, so gilt für das zugehörige Spektralmaß μ

$$\mu(B) = \mu(-B) \quad (3.11)$$

für alle Borelmengen $B \subset (-\pi, \pi)$ (der Punkt $+\pi$ ist hier ausgeschlossen!). Dabei ist $-B := \{-\omega : \omega \in B\}$. Besitzt μ eine Spektraldichte f , so gilt für diese

$$f(\omega) = f(-\omega) \quad \lambda - f.s. \text{ auf } (-\pi, \pi). \quad (3.12)$$

Diese Symmetrie des Spektralmaßes und (bei Existenz) der Spektraldichte für reelle Autokovarianzen sieht man wie folgt ein:

Mit der Spiegelungsabbildung $T: (-\pi, \pi] \rightarrow (-\pi, \pi]$ gemäß $T(\omega) = -\omega$ für $-\pi < \omega < \pi$ und $T(\pi) = \pi$ ist (3.11) äquivalent zu $\mu = \mu^T$. Da γ reell ist, gilt $\gamma(h) = \gamma(-h)$, also $\int_{(-\pi, \pi]} e^{ih \cdot} d\mu = \int_{(-\pi, \pi]} e^{-ih \cdot} d\mu = \int_{(-\pi, \pi]} e^{ihT(\cdot)} d\mu = \int_{(-\pi, \pi]} e^{ih \cdot} d\mu^T \quad \forall h \in \mathbb{Z}$. Wegen der Eindeutigkeit des Spektralmaßes folgt $\mu = \mu^T$. Besitzt $\mu = \mu^T$ eine Spektraldichte f , so gilt wegen der Spiegelungssymmetrie $\lambda = \lambda^T$ und $T \circ T(\omega) = \omega$

$$\int_B f d\lambda = \mu(B) = \mu^T(B) = \int_{T^{-1}(B)} f d\lambda = \int_{T^{-1}(B)} f \circ T \circ T d\lambda = \int_B f \circ T d\lambda$$

für alle B wie oben, also $f = f \circ T \quad \lambda - f.s. \text{ auf } (-\pi, \pi)$. \square

Wir werden im folgenden Satz zeigen, dass das Spektralmaß einer absolut summierbaren Kovarianzfunktion stets eine stetige Lebesgue-Dichte besitzt, die mit Hilfe der sogenannten Inversionsformel berechnet werden kann.

Satz 3.8 (Inversionsformel)

Sei $\gamma: \mathbb{Z} \rightarrow \mathbb{C}$ eine positiv semidefinite Funktion, die absolut summierbar ist, d.h. es gilt

$$\sum_{h=-\infty}^{+\infty} |\gamma(h)| < \infty. \quad (3.13)$$

Dann ist die gemäß

$$f(\omega) := \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}, \quad \forall \omega \in [-\pi, \pi], \quad (3.14)$$

definierte Funktion $f : [-\pi, \pi] \rightarrow \mathbb{R}$ nichtnegativ und stetig. Sie erfüllt ferner die Gleichungen

$$\int_{-\pi}^{\pi} f(\omega) d\omega = \gamma(0), \quad (3.15)$$

sowie

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\omega} f(\omega) d\omega, \quad \forall h \in \mathbb{Z}. \quad (3.16)$$

Das Spektralmaß von γ besitzt also die stetige Spektraldichte f .

Beweis: Wegen (3.13) konvergieren die offensichtlich stetigen Partialsummen in (3.14) gleichmäßig in ω gegen die Funktion f , die deshalb ihrerseits auch stetig ist. Weiter gilt für die stetigen, nichtnegativen Funktionen $\gamma(0) \cdot f_n$ mit f_n aus (3.8) nach dem Kronecker-Lemma, siehe Beispiel A.2 in Anhang A.2, die Aussage $\gamma(0) \cdot f_n(\omega) \rightarrow f(\omega) \quad \forall \omega \in [-\pi, \pi]$. Insbesondere ist f reell und sogar nichtnegativ. Wegen $0 \leq f_n(\omega) \leq (1/2\pi) \sum_{h \in \mathbb{Z}} |\varrho(h)| < \infty \quad \forall \omega \in [-\pi, \pi] \quad \forall n \in \mathbb{N}$, ergibt die Anwendung des Satzes von Lebesgue auf (3.9) die Gleichung (3.16). \square

Bemerkung 3.9

Ist γ in der Inversionsformel reell, so sieht man direkt, dass dann $f(-\omega) = f(\omega) \quad \forall \omega \in [-\pi, \pi]$ gilt (vgl. (3.12)). Offenbar gilt wegen dieser Symmetrie $\int_{-\pi}^{\pi} \sin(h\omega) f(\omega) d\omega = 0$ und deshalb

$$\gamma(h) = \int_{-\pi}^{\pi} \cos(h\omega) f(\omega) d\omega = 2 \int_0^{\pi} \cos(h\omega) f(\omega) d\omega, \quad \forall h \in \mathbb{Z}. \quad (3.17)$$

\square

Bemerkung 3.10 (Autokorrelation als Maß für lineare Abhängigkeit)

Wegen $\varrho(h) = \gamma(h) / \gamma(0)$ gelten die Sätze 3.4 und 3.8, sowie die Bemerkungen 3.7 und 3.9 auch für die Korrelationsfunktion ϱ . Das zugehörige Spektralmaß ist dann sogar ein Wahrscheinlichkeitsmaß und die Spektraldichte ist eine Wahrscheinlichkeitsdichte. Ist ϱ die Korrelationsfunktion einer reellen Zeitreihe $X = (X_t : t \in \mathbb{Z})$, so ist $\varrho(h)$ ein Maß für den linearen Zusammenhang von X_{t+h} und X_t . Bekanntlich gilt nämlich (siehe Behnen/Neuhaus (2003), Lemma 7.15) für $h \in \mathbb{Z}$

$$\begin{aligned} & \min \left\{ E \left(X_t - (a + bX_{t+h}) \right)^2 : a, b \in \mathbb{R} \right\} \\ &= E \left((X_t - \mu) - \varrho(h) (X_{t+h} - \mu) \right)^2 = \gamma(0) (1 - \varrho^2(h)). \end{aligned} \quad (3.18)$$

Beispiele für den Graphen der Funktion $\varrho(h)$, $h = 0, 1, 2, \dots$, den man wie bereits erwähnt als *Korrelogramm* bezeichnet, findet man in den Abbildungen 2.5 und 2.6 in Kapitel 2. Aus der speziellen Gestalt eines Korrelogramms lassen sich Eigenschaften der Zeitreihe ablesen. In der Praxis muss man natürlich aus Beobachtungen einer Zeitreihe die nicht bekannte Korrelation $\varrho(h)$ schätzen, um so Einblicke in die Abhängigkeitsstruktur der zugrunde liegenden Zeitreihe zu gewinnen. \square

Nun wollen wir an einigen konkreten Beispielen die zugehörigen Spektraldichten explizit bestimmen.

Beispiel 3.11 (Weißes Rauschen)

Für die Korrelationsfunktion $\varrho(\cdot)$ eines weißen Rauschens $e = (e_t : t \in \mathbb{Z})$, also einer orthogonalen Folge von Elementen aus $L_2^{\mathbb{C}}$ mit $\sigma_e^2 = \|e_t\|^2 > 0$ für alle $t \in \mathbb{Z}$, gilt

$$\varrho(0) = 1 \text{ und } \varrho(h) = 0, \quad \text{für } |h| = 1, 2, \dots \quad (3.19)$$

Aufgrund der Inversionsformel (3.14) besitzt $\varrho(\cdot)$ die Spektraldichte

$$f(\omega) = \frac{1}{2\pi}, \quad \forall \omega \in [-\pi, \pi]. \quad (3.20)$$

Entsprechend ergibt sich die Spektraldichte f_γ zur Autokovarianzfunktion γ mit $\gamma(0) = \sigma_e^2$ und $\gamma(h) = 0$ für $h \neq 0$ eines weißen Rauschens e zu $f_\gamma \equiv \sigma_e^2/2\pi$.

Im Sinne von Beispiel 3.6 gibt die Spektraldichte das Gewicht der Frequenzen der einzelnen Schwingungen an, die sich bei der Darstellung der Zeitreihe überlagern. Insofern bedeutet die Tatsache, dass die Spektraldichte des weißen Rauschens konstant ist, dass ein weißes Rauschen als eine Überlagerung von Schwingungen angesehen werden kann, in die alle möglichen Frequenzen mit dem gleichen Gewicht eingehen. Dies geschieht in gleicher Weise beim sogenannten weißen Licht, was für die Namensgebung *weißes Rauschen* gesorgt hat (im Gegensatz zu einem sogenannten farbigen Rauschen). \square

Beispiel 3.12 (AR(1)-Reihen)

Sei $X_t + aX_{t-1} = e_t$, $t \in \mathbb{Z}$, eine reelle oder auch komplexe AR(1)-Reihe mit reellem Parameter a , $|a| < 1$. Sie besitzt dann gemäß (2.18) die Korrelationsfunktion

$$\varrho(h) = (-a)^{|h|}, \quad \forall h \in \mathbb{Z}. \quad (3.21)$$

Aufgrund der Inversionsformel (3.14) besitzt ϱ die Spektraldichte f gemäß

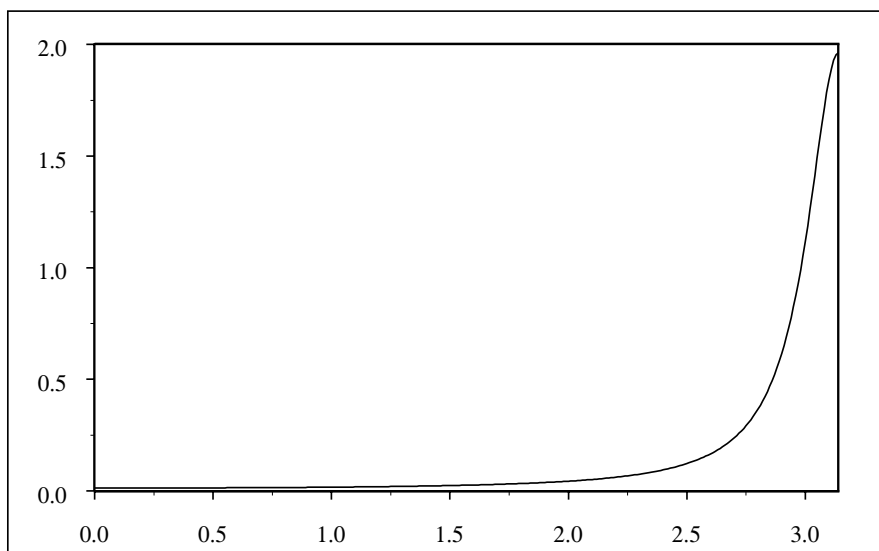


Abb. 3.1 Spektraldichte einer $AR(1)$ Zeitreihe $X_t + aX_{t-1} = e_t$ mit $a = 0.85$

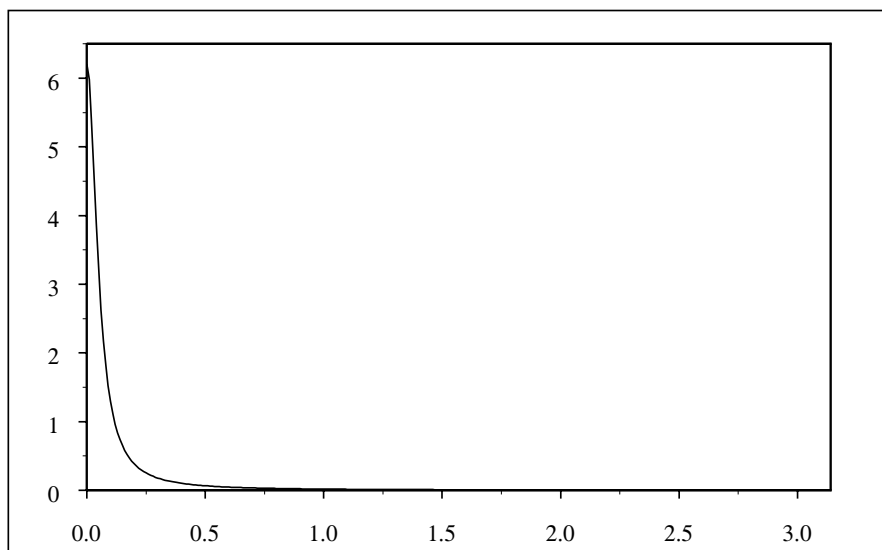


Abb. 3.2 Spektraldichte einer $AR(1)$ Zeitreihe $X_t + aX_{t-1} = e_t$ mit $a = -0.95$

$$\begin{aligned}
2\pi f(\omega) &= \sum_{h=-\infty}^{+\infty} (-a)^{|h|} e^{-ih\omega} \\
&= 1 + \sum_{h=1}^{\infty} (-a \cdot e^{i\omega})^h + \sum_{h=1}^{\infty} (-a \cdot e^{-i\omega})^h \\
&= \frac{1}{1 + ae^{i\omega}} + \frac{1}{1 + ae^{-i\omega}} - 1 \\
&= \frac{(1 + ae^{-i\omega}) + (1 + ae^{i\omega}) - (1 + ae^{i\omega}) \cdot (1 + ae^{-i\omega})}{(1 + ae^{i\omega})(1 + ae^{-i\omega})} \\
&= \frac{1 - a^2}{(1 + 2a \cos \omega + a^2)},
\end{aligned}$$

also

$$f(\omega) = \frac{1}{2\pi} \frac{1 - a^2}{1 + 2a \cos \omega + a^2}, \quad \forall \omega \in [-\pi, \pi]. \quad (3.22)$$

Für die zugehörige Kovarianzfunktion und die hierzu gehörende Spektraldichte multipliziert man (3.21) und (3.22) mit $\gamma(0) = \sigma_e^2 / (1 - |a|^2)$. Die Abbildungen 3.1 und 3.2 zeigen jeweils die Spektraldichte einer $AR(1)$ -Zeitreihe. In Abbildung 3.1 ist der Parameter $a = 0.85$ gewählt worden. Die zugehörige Spektraldichte steigt monoton und wird maximal für $\omega = \pi$, d.h. sie legt ein stärkeres Gewicht auf die größeren Frequenzen ω , also auf die kleineren Wellenlängen. Unter Verwendung der bereits bei zyklischen Reihen und im Rahmen des weißen Rauschens angesprochenen Interpretation müsste sich in einer Realisierung einer $AR(1)$ -Zeitreihe mit Parameter $a = 0.85$ also zeigen, dass höherfrequente, oder mit anderen Worten kurzwellige, Schwingungen ausgeprägter auftreten sollten. Abbildung 3.3 zeigt eine simulierte Realisation einer derartigen $AR(1)$ -Zeitreihe und man kann das erwartete Phänomen einer stark fluktuierenden Zeitreihe in der Tat beobachten. Genau umgekehrt ist die Situation im Falle des $AR(1)$ -Modells mit Parameter $a = -0.95$. Die Abbildung 3.2 der Spektraldichte legt diesmal eher niederfrequente, d.h. verstärkt langwellige Schwingungsanteile nahe und in der Tat zeigt die Realisation einer $AR(1)$ -Zeitreihe mit Parameter $a = -0.95$ (vgl. Abbildung 3.4) im Gegensatz zu Abbildung 3.3 einen deutlich ruhigeren Pfad.

□

Besonders gut lässt sich die Bedeutung der Spektraldichte an einer speziellen $AR(2)$ -Zeitreihe demonstrieren.

Beispiel 3.13 (Spezielle $AR(2)$ -Reihen)

Für die $AR(2)$ -Zeitreihe $X_t + aX_{t-2} = e_t$, $t \in \mathbb{Z}$, mit $a \in \mathbb{R}$, $|a| < 1$, berechnet sich die Autokorrelationsfunktion zu

$$\varrho(h) = \begin{cases} (-a)^{|h|/2}, & h \text{ gerade} \\ 0, & h \text{ ungerade} \end{cases}$$

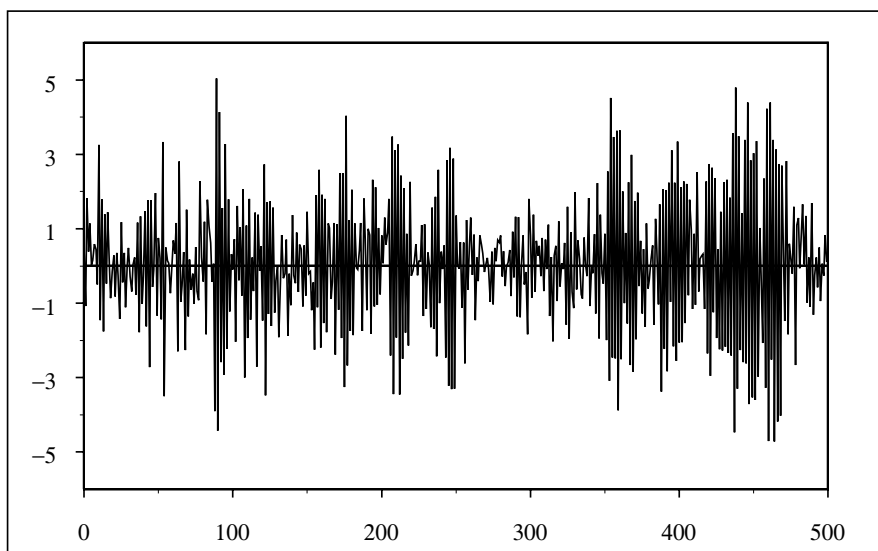


Abb. 3.3 Simulation einer $AR(1)$ -Zeitreihe mit Parameter $a = 0.85$ der Länge $n = 500$

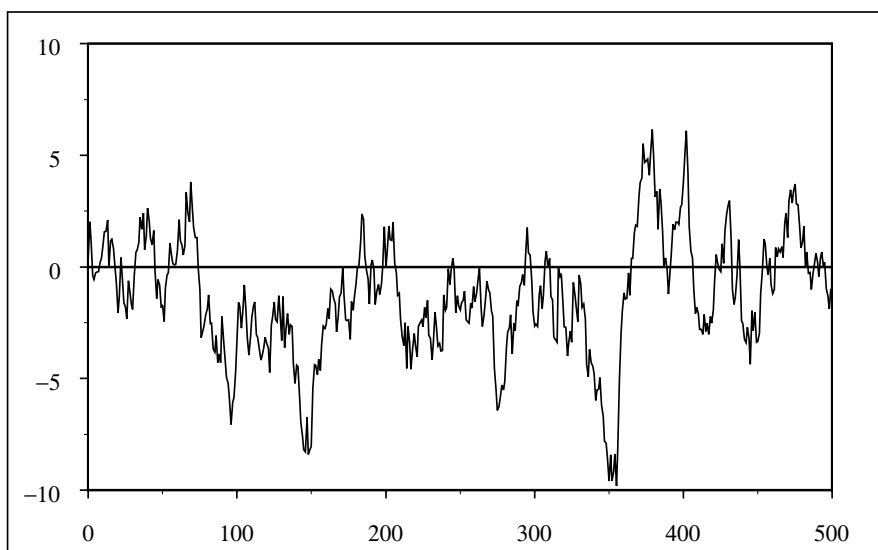


Abb. 3.4 Simulation einer $AR(1)$ -Zeitreihe mit Parameter $a = -0.95$ der Länge $n = 500$

siehe Aufgabe 3.3. Für $a > 0$ haben wir also über zwei Lags (Zeitdifferenz von 2 Einheiten) die negative Korrelation $-a$ und über vier Lags die positive Korrelation a^2 . Je dichter wir nun a bei $+1$ wählen, desto ausgeprägter sind diese Korrelationen. Deshalb erwarten wir aufgrund dieser speziellen Struktur ein mehr oder weniger deutliches periodisches Verhalten der Pfade der zugehörigen $AR(2)$ -Zeitreihe und zwar mit Periode vier. Stimmt die oben angesprochene Heuristik der Spektraldichte, dann müsste diese ein Maximum bei der Frequenz $\omega_0 = 2\pi/4 = \pi/2$ besitzen.

Letzteres können wir wie folgt einsehen: Mit derselben Rechnung wie in Bei-

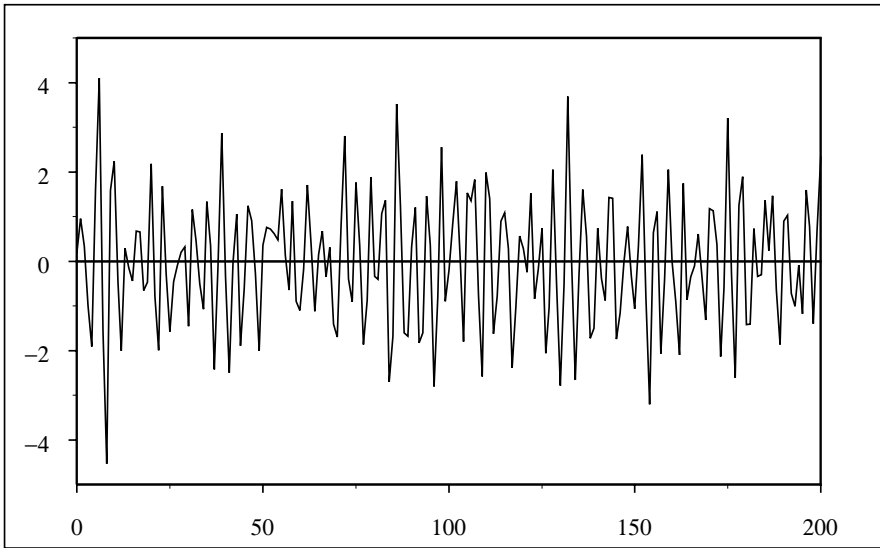


Abb. 3.5 Simulation der $AR(2)$ -Zeitreihe $X_t + 0.70 \cdot X_{t-2} = e_t$ der Länge $n = 200$

spiel 3.12 erhält man nämlich für die Spektraldichte der Autokorrelation

$$f(\omega) = \frac{1}{2\pi} \frac{1 - a^2}{1 + 2a \cos(2\omega) + a^2}, \quad \forall \omega \in [-\pi, \pi], \quad (3.23)$$

d.h. bei $0 < a < 1$ ein Maximum an der Stelle $\omega = \pi/2$ mit dem Maximalwert $f(\pi/2) = (1 + a)/((1 - a)2\pi)$, der für $a \rightarrow 1$ gegen ∞ strebt.

Die in den Abbildungen 3.5 und 3.6 dargestellten Simulationen von Pfaden des betrachteten $AR(2)$ -Prozesses mit Parameter $a = 0.70$ bzw. $a = 0.95$ zeigen zudem recht deutlich das erwartete periodische Verhalten.

Die zugehörigen Spektraldichten sind in den Abbildungen 3.7 bzw. 3.8 dargestellt und man erkennt, dass je ausgeprägter der peak der Spektraldichte an

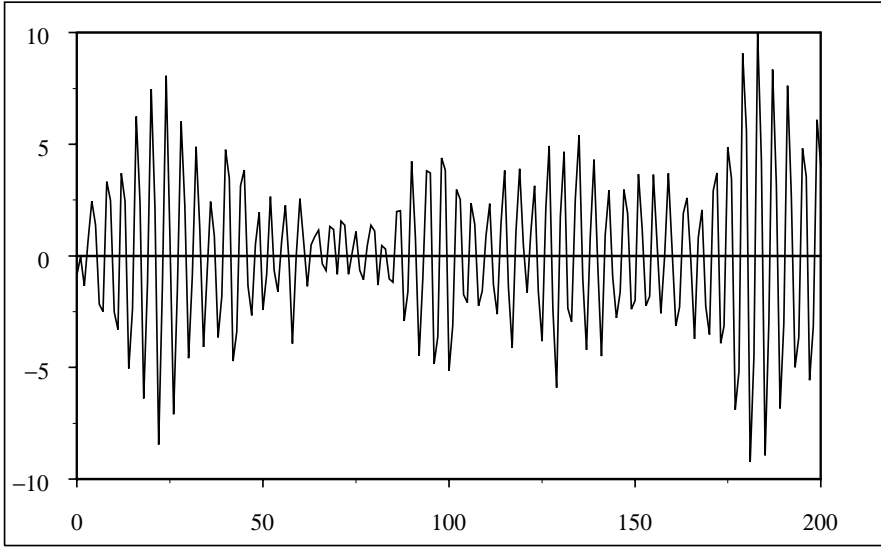


Abb. 3.6 Simulation der $AR(2)$ -Zeitreihe $X_t + 0.95 \cdot X_{t-2} = e_t$ der Länge $n = 200$

der Stelle $\omega_0 = \pi/2$ ist, desto deutlicher fällt das periodische Aussehen der Realisation der speziellen $AR(2)$ -Zeitreihe ins Auge. \square

Abschließend wollen wir noch die Klasse der Moving Average Zeitreihen näher betrachten.

Beispiel 3.14 (MA-Reihen)

Die Autokovarianzstruktur von Moving Average Zeitreihen haben wir bereits in Beispiel 2.6 bestimmt. Besonders charakteristisch an dieser Autokovarianzstruktur ist, dass für $MA(q)$ -Zeitreihen stets $\gamma(h) = \varrho(h) = 0$ für $|h| > q$ gilt. Damit ergibt die Inversionsformel (vgl. Satz 3.8) die folgende Spektraldichte für $MA(q)$ -Zeitreihen, wobei wir der Einfachheit halber den reellen Fall, d.h. ein reelles weißes Rauschen und reelle Koeffizienten b_j betrachten

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{|h| \leq q} \varrho(h) e^{-i\omega h} \\ &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{h=1}^q \varrho(h) \cos(h\omega), \quad \forall \omega \in [-\pi, \pi], \end{aligned} \quad (3.24)$$

mit

$$\varrho(h) = \begin{cases} \sum_{j=0}^{q-h} b_j b_{j+h} / \sum_{j=0}^q b_j^2, & h = 0, \dots, q \\ 0, & h > q \end{cases}.$$

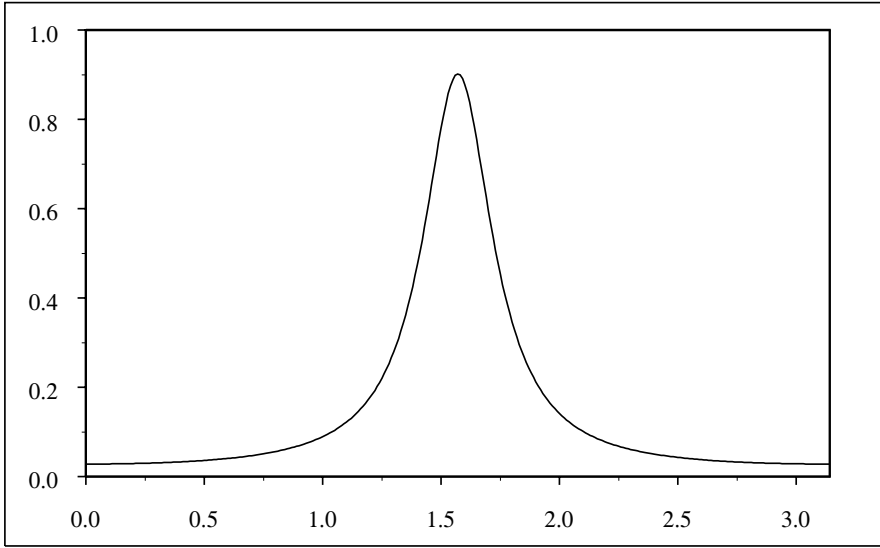


Abb. 3.7 Spektraldichte der $AR(2)$ -Zeitreihe $X_t + 0.70 \cdot X_{t-2} = e_t$

Für den $MA(1)$ -Prozess $X_t = e_t + b \cdot e_{t-1}$, $t \in \mathbb{Z}$, erhalten wir sofort die Spektraldichte

$$f(\omega) = \frac{1}{2\pi} + \frac{b}{\pi(1+b^2)} \cos(\omega), \quad \forall \omega \in [-\pi, \pi].$$

Für den speziellen $MA(2)$ -Prozess $X_t = e_t + b \cdot e_{t-2}$ gilt wegen $\varrho(1) = 0$ und $\varrho(2) = b/(1+b^2)$ aufgrund von (3.24)

$$\begin{aligned} 2\pi \cdot f(\omega) &= 1 + \varrho(2) (e^{2i\omega} + e^{-2i\omega}) \\ &= \frac{1}{1+b^2} (1 + 2b \cos(2\omega) + b^2). \end{aligned} \quad (3.25)$$

Für $0 < b < 1$ nimmt f in $\omega = \pm\pi/2$ den Minimalwert $f(\pm\pi/2) = (1-b)^2 / ((1+b^2)2\pi)$ an, der umso kleiner wird, je dichter b bei 1 liegt (vgl. Abbildung 3.9).

Das Ergebnis ist dual zu dem in Beispiel 3.13 behandelten speziellen $AR(2)$ -Prozess, dessen Spektraldichte an diesen Frequenzen ein Maximum besaß. Wir werden in Kapitel 6 über Filter sehen, dass diese Eigenschaft von Moving Average Zeitreihen ausgenutzt werden kann, um bestehende periodische Anteile in realen Zeitreihen (hier mit der Frequenz $\pi/2$ bzw. mit der Wellenlänge 4) auszublenden. \square

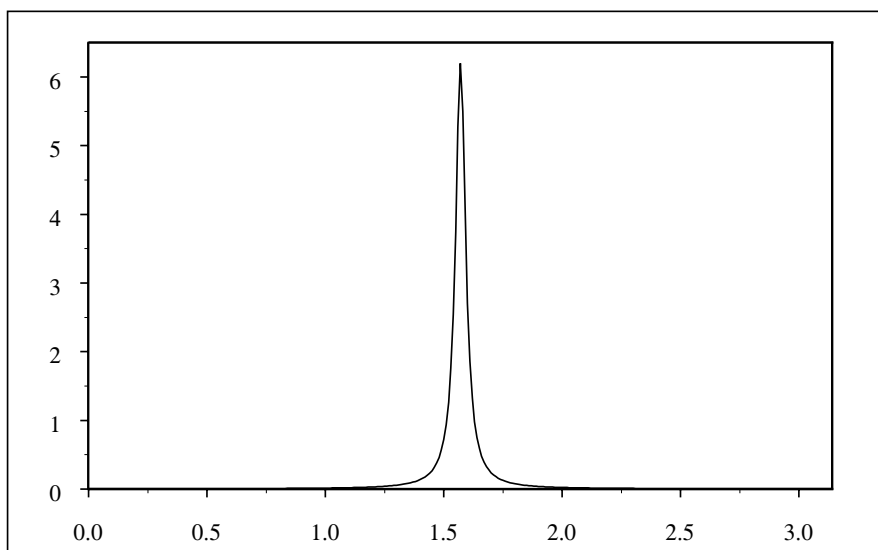


Abb. 3.8 Spektraldichte der $AR(2)$ -Zeitreihe $X_t + 0.95 \cdot X_{t-2} = e_t$

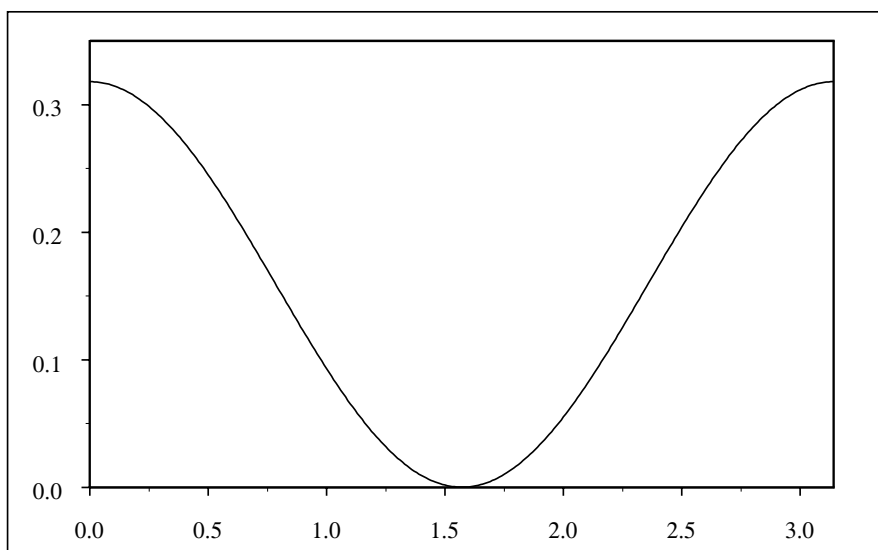


Abb. 3.9 Spektraldichte der $MA(2)$ -Zeitreihe $X_t = e_t + 0.95 \cdot e_{t-2}$

Aufgaben

Aufgabe 3.1 Sei $\varrho : \mathbb{Z} \rightarrow \mathbb{R}$ gegeben durch $\varrho(h) = 1(h=0) + \alpha 1(|h|=1)$.

Man zeige, dass ϱ genau dann die Autokovarianzfunktion einer reellen, stationären Zeitreihe ist, wenn $|\alpha| \leq 1/2$ gilt.

Hinweise: Für eine Richtung benutze man die Ergebnisse über $MA(1)$ -Reihen. Für die andere Richtung beachte man die positive Semidefinitheit von Autokovarianzfunktionen und benutze die Vektoren $(1, -1, 1, \dots)$ und $(1, 1, 1, \dots, 1)$.

Aufgabe 3.2 Sei $\gamma : \mathbb{Z} \rightarrow \mathbb{R}$ gegeben durch

$$\gamma(h) = 1(h=0) - 0.5 \cdot 1(|h|=2) - 0.25 \cdot 1(|h|=3).$$

Ist dies eine Autokovarianzfunktion?

Aufgabe 3.3 Man berechne die Autokorrelationsfunktion ϱ für die $AR(2)$ -Reihe in Beispiel 3.13

Aufgabe 3.4 a) Man berechne für eine $\mathcal{R}(0,1)$ -verteilte ZV U die Kovarianzfunktion K der stochastischen Prozesse $(X_t := 1_{[0,t]}(U) : t \in [0,1])$ und $(Y_t := 1_{[0,t]}(U) - t : t \in [0,1])$.

b) Man zeige: Mit reellen Kovarianzfunktionen K_1, K_2 ist auch $K = K_1 \cdot K_2$ eine Kovarianzfunktion.

Aufgabe 3.5 Für einen quadratintegrierbaren, komplexwertigen, nicht notwendig (schwach) stationären Prozess $X = (X_t : t \in T)$, $T \subset \mathbb{R}$, definiere man seine Kovarianzfunktion $K(\cdot, \cdot)$ auf $T \times T$ durch

$$K(s, t) = E(X_s - EX_s)(\overline{X_t} - E\overline{X_t}).$$

a) Man zeige, dass K positiv semidefinit ist, d.h. es gilt (3.3) mit $\gamma(t_i - t_j)$ ersetzt durch $K(t_i, t_j)$ und $t_i, t_j \in T$ und folgere allein aus dieser Eigenschaft, dass K dann auch Hermite-symmetrisch ist.

Solche Funktionen K heißen *Kovarianzfunktionen*, denn man kann zeigen, dass es zu jeder positiv semidefiniten Funktion K einen Gaußprozess gibt, der diese als seine Kovarianzfunktion hat, vgl. Satz 3.1.

b) Man zeige für alle $s, t \in T$ die Aussagen $K(s, s) \geq 0$ und

$$|K(s, t)| \leq (K(s, s)K(t, t))^{1/2},$$

und weise nach, dass der Realteil einer Kovarianzfunktion stets eine solche ist, der Imaginärteil aber nur, wenn er identisch 0 ist.

Aufgabe 3.6 Seien $g_k(t), t \in \mathbb{Z}$, beliebige komplexwertige Funktionen für $k = 1, 2, \dots$. Sei

$$K(s, t) := \sum_{k=1}^{\infty} g_k(s) \overline{g_k(t)}.$$

Man zeige: Falls die Reihe absolut konvergent ist $\forall s, t \in \mathbb{Z}$, so ist K positiv semidefinit.

Aufgabe 3.7 Sei jetzt $T = (a, b)$ mit $a < b$, wobei $a = -\infty$ und $b = \infty$ sein darf. Der Prozess X in Aufgabe 3.5 heißt L_2 -stetig im Punkt $t_0 \in T$, wenn aus $t \rightarrow t_0$ die Aussage $X_t \rightarrow X_{t_0}$ im quadratischen Mittel folgt.

Man zeige: Falls EX_t stetig auf T ist, so ist X genau dann L_2 -stetig in allen Punkten von T , wenn seine Kovarianzfunktion $K(s, t)$ in allen Punkten (t, t) der Diagonalen stetig ist. Man zeige außerdem, dass hieraus schon die Stetigkeit von K auf $T \times T$ folgt.

Aufgabe 3.8 Man gebe einen stationären L_2 -Prozess $X = (X_t : t \in \mathbb{R})$ an, der nicht L_2 -stetig ist.

Aufgabe 3.9 Es sollen dieselben Voraussetzungen wie in Aufgabe 3.5 gelten. Man zeige:

Genau dann existiert ein $Y \in L_2^{\mathbb{C}}$ mit $X_t \rightarrow Y$ falls $t \rightarrow t_0$, wenn es eine komplexe Zahl L gibt derart, dass für alle Folgen $t_n \rightarrow t_0$ und $s_m \rightarrow t_0$ gilt

$$E(X_{t_n} \overline{X_{s_m}}) \rightarrow L \text{ falls } n, m \rightarrow \infty.$$

Aufgabe 3.10 Sei $X = (X(t) : a \leq t \leq b)$, $-\infty < a < b < \infty$, ein zentrierter $L_2^{\mathbb{C}}$ -Prozess. Konvergiert für jede Zerlegungsfolge $D_n : a = t_0^{(n)} < \dots < t_n^{(n)} := b$ mit $\max_{1 \leq k \leq n} t_k^{(n)} - t_{k-1}^{(n)} \rightarrow 0$ für $n \rightarrow \infty$ und jede Wahl der Zwischenpunkte $\bar{t}_k^{(n)}$ aus $[t_{k-1}^{(n)}, t_k^{(n)}]$ die Folge der Zufallsvariablen

$$X_{D_n}(\omega) := \sum_{k=1}^n X(\bar{t}_k^{(n)}, \omega)(t_k^{(n)} - t_{k-1}^{(n)})$$

im quadratischen Mittel gegen einunddasselbe $Y \in L_2^{\mathbb{C}}$, so definiert man $\int_a^b X(t) dt := Y$.

a) Man zeige: $\int_a^b X(t) dt$ existiert genau dann, wenn $\int_a^b \int_a^b K(s, t) ds dt$ existiert mit $K(s, t) := EX(s) \overline{X(t)}$. Letzteres ist insbesondere dann erfüllt, wenn X auf $[a, b]$ im quadratischen Mittel stetig ist.

b) Man bestimme eine allgemeine Formel für die Varianz von $I := \int_a^b X(t) dt$ und berechne damit $\text{Var } I$ für $a = 0, b = 1$ im Fall des Prozesses $(X_t = 1_{[0, t]}(U) - t : 0 \leq t \leq 1)$ mit $U \sim \mathcal{R}(0, 1)$. Außerdem berechne man für diesen speziellen Fall I pfadweise und anschließend damit $\text{Var } I$ erneut.

Lineare Vorhersage bei endlicher Vergangenheit

In diesem Kapitel behandeln wir die Frage, wie man für eine stationäre, reelle und zentrierte Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit gegebener Autokovarianzfunktion $\gamma(\cdot)$ bei bekannten Werten X_1, \dots, X_n einen zukünftigen Wert X_{n+h} mit $h \geq 1$ durch eine Linearkombination von X_1, \dots, X_n möglichst gut approximiert oder mit anderen Worten *linear vorhersagt*.

Dazu betrachten wir im Raum $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P) =: L_2^{\mathbb{R}}$ den linearen Unterraum $\text{span}\{X_1, \dots, X_n\} =: H_n$. Unter der besten linearen Vorhersage von X_{n+h} verstehen wir dann denjenigen Punkt aus H_n , der den kleinsten Abstand von X_{n+h} in der $L_2^{\mathbb{R}}$ -Norm hat, also die *Projektion* $\hat{X}_{n+h}^{(h)}$ von X_{n+h} auf H_n , die wir auch als *h-Schritt-Vorhersage* von X_{n+h} bezeichnen. Da H_n ein abgeschlossener, linearer Teilraum von $L_2^{\mathbb{R}}$ ist, existiert bekanntlich genau ein solches Element $\hat{X}_{n+h}^{(h)}$ aus H_n und ist eindeutig bestimmt als eine Linearkombination

$$\hat{X}_{n+h}^{(h)} = X_1 d_{n,n}^{(h)} + X_2 d_{n,n-1}^{(h)} + \dots + X_n d_{n,1}^{(h)} \quad (4.1)$$

mit

$$\hat{X}_{n+h}^{(h)} - X_{n+h} \perp H_n, \quad (4.2)$$

siehe z.B. Anhang A.3. Die Bedingung (4.2) ist gleichbedeutend mit

$$\langle \hat{X}_{n+h}^{(h)} - X_{n+h}, X_{n+1-j} \rangle = 0 \quad \text{für} \quad j = 1, \dots, n,$$

d.h. mit den sog. Normalgleichungen (siehe Anhang A.3)

$$\sum_{i=1}^n \gamma(j-i) d_{n,i}^{(h)} = \gamma(h+j-1) \quad \text{für} \quad j = 1, \dots, n,$$

oder in Matrizenschreibweise unter Verwendung der Bezeichnungen

$$\begin{aligned}\Gamma_n &:= (\gamma(j-i))_{i,j=1,\dots,n} \\ \gamma_n^{(h)} &:= (\gamma(h), \gamma(h+1), \dots, \gamma(h+n-1))^T \\ d_n^{(h)} &:= \left(d_{n,1}^{(h)}, \dots, d_{n,n}^{(h)}\right)^T\end{aligned}$$

mit der Gleichung

$$\Gamma_n d_n^{(h)} = \gamma_n^{(h)}. \quad (4.3)$$

Da die Projektion $\widehat{X}_{n+h}^{(h)}$ immer existiert, besitzt (4.3) stets eine Lösung $d_n^{(h)}$. Diese ist eindeutig bestimmt, wenn Γ_n nicht ausgeartet ist, nämlich gleich $d_n^{(h)} = \Gamma_n^{-1} \gamma_n^{(h)}$. Falls Γ_n ausgeartet ist, gibt es unendlich viele Lösungen $d_n^{(h)}$, die jedoch gemäß (4.1) alle zum selben Wert von \widehat{X}_{n+h} führen.

Bemerkung 4.1 (Zeitinvarianz der Koeffizienten und des Vorhersagefehlers)

Wegen der Annahme einer stationären Zeitreihe ändert eine Parallelverschiebung aller Beobachtungszeitpunkte um einen festen Wert nichts an den Normalgleichungen, da die Kovarianzen ja nur von Zeitdifferenzen abhängen. Ist die Matrix Γ_n nicht ausgeartet, so sind die eindeutig bestimmten Koeffizienten $d_{n,i}$ deshalb zeitinvariant, d.h. sie sind für jedes $t \in \mathbb{Z}$ die Entwicklungskoeffizienten der Projektion von X_{t+h} auf den von den n Vektoren X_j , $t-n+1 \leq j \leq t$, aufgespannten linearen Teilraum. Auch der quadratische Abstand der Projektion vom projizierten Vektor, also

$$\|X_{n+h} - \widehat{X}_{n+h}^{(h)}\|^2, \quad (4.4)$$

den wir als quadratischen h -Schritt-Vorhersagefehler bezeichnen, bleibt aus demselben Grund bei einer Parallelverschiebung aller Beobachtungszeitpunkte derselbe. \square

In den folgenden beiden Abschnitten wollen wir besonders einfache rekursive Methoden, die zur Bestimmung von Projektionen entwickelt wurden, auf die spezielle Situation der Zeitreihenanalyse zugeschnitten darstellen. Dazu seien X_1, X_2, \dots Elemente aus einem reellen Hilbertraum H mit innerem Produkt $\langle \cdot, \cdot \rangle$ und Norm $\|\cdot\|$. In unseren Anwendungen wird H stets gleich $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ sein mit dem üblichen Skalarprodukt und die Elemente X_1, \dots, X_n werden aus einer reellen, stationären und zentrierten Zeitreihe stammen. Im Moment ist dies jedoch nicht gefordert. Wir betrachten zunächst nur die Einschnitt-Vorhersage ($h = 1$) und schreiben dann zur Abkürzung \widehat{X}_{n+1} statt $\widehat{X}_{n+1}^{(1)}$. Den allgemeinen Fall $h > 1$ werden wir später mit Hilfe der Iterationseigenschaft von Projektionsoperatoren auf die Einschnitt-Vorhersage zurückführen.

4.1 Die rekursive Gram-Schmidt-Orthogonalisierung

Zu jeder Folge X_1, X_2, \dots linear unabhängiger Elemente von H wird nach *Gram-Schmidt* durch

$$Y_1 = X_1, Y_2 = X_2 - \widehat{X}_2, \dots, Y_n = X_n - \widehat{X}_n, \dots \quad (4.5)$$

mit $\widehat{X}_{j+1} := P_j X_{j+1}$, $j \geq 1$, und den Orthogonalprojektionen $P_j : H \rightarrow H$ auf $H_j := \text{span}\{X_1, \dots, X_j\}$ eine Folge orthogonaler Vektoren Y_j , $j \geq 1$, definiert mit der Eigenschaft

$$H_j = \text{span}\{X_1, \dots, X_j\} = \text{span}\{Y_1, \dots, Y_j\}, \quad \forall j \geq 1. \quad (4.6)$$

Die *Einschritt–Vorhersage* \widehat{X}_{n+1} hat wegen (4.6) mit geeigneten Koeffizienten $c_{n,j}$ die Darstellung

$$\widehat{X}_{n+1} = \sum_{k=1}^n c_{n,k} Y_{n+1-k}, \quad \forall n \geq 1. \quad (4.7)$$

Skalarmultiplikation von (4.7) von rechts mit Y_{n+1-j} liefert unter Verwendung von $\langle \widehat{X}_{n+1}, Y_{n+1-j} \rangle = \langle X_{n+1}, Y_{n+1-j} \rangle$ die Lösungen

$$c_{n,j} = \langle X_{n+1}, Y_{n+1-j} \rangle \langle Y_{n+1-j}, Y_{n+1-j} \rangle^{-1}, \quad 1 \leq j \leq n. \quad (4.8)$$

Wegen der linearen Unabhängigkeit der X_1, X_2, \dots sind auch die Orthogonalvektoren Y_1, Y_2, \dots linear unabhängig, insbesondere also ungleich 0. Die Größen

$$\begin{aligned} v_n &:= \|X_{n+1} - \widehat{X}_{n+1}\|^2 = \langle Y_{n+1}, Y_{n+1} \rangle, \quad \forall n \geq 1, \\ v_0 &:= \langle X_1, X_1 \rangle \end{aligned} \quad (4.9)$$

nennt man quadratische (*Einschritt–*)*Vorhersagefehler*, vgl. (4.4) (hier ist allerdings keine Stationarität vorausgesetzt).

Man beachte, dass die Koeffizienten $c_{n,j}$ gemäß Gleichung (4.7) eine Linearkombination der orthogonalen Vektoren Y_1, \dots, Y_n definieren. Deshalb stimmen sie im Allgemeinen nicht mit den Koeffizienten $d_{n,j}^{(1)}$ aus (4.1) überein, können aber natürlich linear ineinander transformiert werden. Die orthogonalen Vektoren $Y_1 = X_1, Y_2 = X_2 - \widehat{X}_2, Y_3 = X_3 - \widehat{X}_3, \dots$ werden auch als *Innovationen* bezeichnet. Y_{n+1} gibt den orthogonalen Zuwachs an, der zum Raum $\text{span}\{X_1, \dots, X_n\}$ hinzuzufügen ist, um $\text{span}\{X_1, \dots, X_{n+1}\}$ zu erzeugen.

Die Koeffizienten $c_{n,j}$, $1 \leq j \leq n$, der Einschritt–Vorhersage \widehat{X}_{n+1} sowie v_n erhält man rekursiv aus den folgenden Gleichungen. Es wird dabei stets vorausgesetzt, dass $\langle X_i, X_j \rangle$ für alle vorkommenden i, j bekannt ist. Zunächst folgt aus (4.8) direkt

$$c_{n,n} = \langle X_{n+1}, X_1 \rangle \cdot \langle X_1, X_1 \rangle^{-1}, \quad \forall n \geq 1. \quad (4.10)$$

Die Rekursion selbst verwendet für $n \geq 2$ und $1 \leq k \leq n-1$ die unten verifizierten Beziehungen

$$\begin{aligned}
c_{n,n-k} &= \left(\langle X_{n+1}, X_{k+1} \rangle - \sum_{j=0}^{k-1} c_{n,n-j} \cdot v_j \cdot c_{k,k-j}^T \right) v_k^{-1} \\
v_n &= \langle X_{n+1}, X_{n+1} \rangle - \sum_{j=0}^{n-1} c_{n,n-j} \cdot v_j \cdot c_{n,n-j}^T.
\end{aligned} \tag{4.11}$$

Man erhält dann aus (4.10) und (4.11) nacheinander die Werte

$$v_0; \quad c_{1,1}, v_1; \quad c_{2,2}, c_{2,1}, v_2; \quad c_{3,3}, c_{3,2}, c_{3,1}, v_3; \quad \dots \text{ usw.}$$

Bemerkung 4.2 (Zur multivariaten Schreibweise)

Obwohl die Koeffizienten $c_{n,j}$ hier reelle Zahlen sind, haben wir ein Transponiertzeichen eingefügt. Der Grund ist, dass eine analoge Überlegung in Kapitel 15 für multivariate Zeitreihen durchgeführt wird. Dann werden die $c_{n,j}$ Matrizen sein. Auch den folgenden Nachweis haben wir so geschrieben, dass er im multivariaten Fall ohne Änderung gültig bleibt. \square

Zum Nachweis von (4.11) verwenden wir (4.8) für $1 \leq k \leq n-1$ und $n \geq 2$ (mit $j = n-k$). Es gilt

$$\begin{aligned}
c_{n,n-k} v_k &= \langle X_{n+1}, Y_{k+1} \rangle \\
&= \langle X_{n+1}, X_{k+1} \rangle - \langle X_{n+1}, \widehat{X}_{k+1} \rangle \\
&= \langle X_{n+1}, X_{k+1} \rangle - \langle X_{n+1}, \sum_{j=1}^k c_{k,j} Y_{k+1-j} \rangle \\
&= \langle X_{n+1}, X_{k+1} \rangle - \sum_{j=1}^k \langle X_{n+1}, Y_{k+1-j} \rangle \cdot c_{k,j}^T \\
&= \langle X_{n+1}, X_{k+1} \rangle - \sum_{j=0}^{k-1} \langle X_{n+1}, Y_{j+1} \rangle \cdot c_{k,k-j}^T \\
&= \langle X_{n+1}, X_{k+1} \rangle - \sum_{j=0}^{k-1} c_{n,n-j} \cdot v_j \cdot c_{k,k-j}^T,
\end{aligned}$$

wobei für die letzte Gleichung die Aussage (4.8) mit j ersetzt durch $n-j$ angewendet wurde. Die obige Rechnung liefert die erste Gleichung in (4.11).

Aus der Orthogonalität von \widehat{X}_{n+1} und Y_{n+1} und $X_{n+1} = \widehat{X}_{n+1} + Y_{n+1}$ folgt die Gleichung $v_n = \langle X_{n+1}, X_{n+1} \rangle - \langle \widehat{X}_{n+1}, \widehat{X}_{n+1} \rangle$ und schließlich mit (4.7) die zweite Gleichung in (4.11).

Die h-Schritt-Vorhersage $\widehat{X}_{n+h}^{(h)}$

Bezeichnet P_n den Projektionsoperator auf H_n , so folgt wegen $P_n P_{n+h-1} = P_n$ (die Iterationseigenschaft von Projektionen)

$$\begin{aligned}
\widehat{X}_{n+h}^{(h)} &= P_n X_{n+h} \\
&= P_n P_{n+h-1} X_{n+h} = P_n \widehat{X}_{n+h} \\
&= P_n \left(\sum_{j=1}^{n+h-1} c_{n+h-1,j} Y_{n+h-j} \right) \\
&= \sum_{j=h}^{n+h-1} c_{n+h-1,j} Y_{n+h-j},
\end{aligned}$$

wegen $Y_k \perp H_n$ für $k > n$, also $P_n Y_k = 0$ für $k > n$. Es folgt

$$\widehat{X}_{n+h}^{(h)} = \sum_{j=h}^{n+h-1} c_{n+h-1,j} \left(X_{n+h-j} - \widehat{X}_{n+h-j} \right) \quad (4.12)$$

und somit für den quadratischen h -Schritt-Vorhersagefehler

$$\begin{aligned}
\|X_{n+h} - \widehat{X}_{n+h}^{(h)}\|^2 &= \|X_{n+h}\|^2 - \|\widehat{X}_{n+h}^{(h)}\|^2 \\
&= \|X_{n+h}\|^2 - \sum_{j=h}^{n+h-1} c_{n+h-1,j}^2 \cdot v_{n+h-j}. \quad (4.13)
\end{aligned}$$

Während es für die gerade behandelte Gram–Schmidt–Orthogonalisierung keine Rolle spielt, ob die Zeitreihe $X = (X_t : t \in \mathbb{Z})$ stationär ist, wird bei der nun zu betrachtenden Rekursion nach Levinson (1947) ausgiebig von der Stationarität der zugrunde liegenden Zeitreihe Gebrauch gemacht werden.

4.2 Die Levinson–Rekursion

Wir verwenden dieselben Bezeichnungen wie bei der Gram–Schmidt–Rekursion und machen zusätzlich die Stationaritätsannahme, dass das Skalarprodukt $\langle X_i, X_j \rangle = EX_i X_j = \gamma(|i - j|)$ nur vom Abstand $|i - j|$ abhängt. Die Vektoren X_1, X_2, \dots seien linear unabhängig im reellen Hilbertraum H . P_n sei der Projektionsoperator auf $H_n = \text{span} \{X_1, \dots, X_n\}$.

Wir wollen nun die Einschnitt–Vorhersage $\widehat{X}_{n+1} = P_n X_{n+1}$ als Linearkombination der X_1, \dots, X_n darstellen gemäß

$$\widehat{X}_{n+1} = d_{n,1} X_n + \dots + d_{n,n} X_1, \quad \forall n \geq 1. \quad (4.14)$$

Den quadratischen (Einschnitt–) Vorhersagefehler bezeichnen wir wie oben mit

$$\begin{aligned}
v_n &= \|X_{n+1} - \hat{X}_{n+1}\|^2 \\
&= E \left(X_{n+1} - \sum_{j=1}^n d_{n,j} X_{n+1-j} \right) X_{n+1} \\
&\quad - \sum_{h=1}^n d_{n,h} E \left(X_{n+1} - \hat{X}_{n+1} \right) X_{n+1-h} \\
&= \gamma(0) - \sum_{j=1}^n d_{n,j} \gamma(j), \quad \forall n \geq 1,
\end{aligned} \tag{4.15}$$

da $X_{n+1-h} \perp (X_{n+1} - \hat{X}_{n+1}) \forall h = 1, \dots, n$. Mit $d_{1,1} = \gamma(1)/\gamma(0)$ und $v_0 = \gamma(0)$ gilt die *Levinson-Rekursion*

$$d_{n,n} = \left(\gamma(n) - \sum_{j=1}^{n-1} d_{n-1,j} \gamma(n-j) \right) \cdot v_{n-1}^{-1}, \tag{4.16}$$

$$d_{n,j} = d_{n-1,j} - d_{n,n} \cdot d_{n-1,n-j}, \quad 1 \leq j \leq n-1, \tag{4.17}$$

und

$$v_n = v_{n-1} (1 - d_{n,n}^2). \tag{4.18}$$

Man berechnet damit rekursiv

$$v_0; \quad d_{1,1}, v_1; \quad d_{2,2}, d_{2,1}, v_2; \quad d_{3,3}, d_{3,2}, d_{3,1}, v_3; \quad \dots \text{ usw.}$$

Um die Gültigkeit der Levinson-Rekursion einzusehen, folgen wir der Darstellung von Brockwell/Davis (1991) und zerlegen $H_n = \text{span}\{X_1, \dots, X_n\}$ in die zueinander orthogonalen Teilräume $K_1 = \text{span}\{X_2, \dots, X_n\}$ und $K_2 = \text{span}\{X_1 - P_{K_1} X_1\}$, also $H_n = K_1 \oplus K_2$, wobei P_{K_i} die Projektion auf K_i bedeutet. Es gilt dann

$$\begin{aligned}
\hat{X}_{n+1} &= P_{K_1} X_{n+1} + P_{K_2} X_{n+1} \\
&= P_{K_1} X_{n+1} + a (X_1 - P_{K_1} X_1)
\end{aligned} \tag{4.19}$$

mit

$$a = \frac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_1 - P_{K_1} X_1\|^2}. \tag{4.20}$$

Für die Vektoren $(Z_1, \dots, Z_n) := (X_2, \dots, X_{n+1})$ gilt wegen der Stationaritätsannahme $\langle Z_i, Z_j \rangle = \langle X_{i+1}, X_{j+1} \rangle = \langle X_i, X_j \rangle, 1 \leq i, j \leq n$, und deshalb

$$P_{K_1} X_{n+1} = \sum_{j=1}^{n-1} d_{n-1,j} Z_{n-j} = \sum_{j=1}^{n-1} d_{n-1,j} X_{n+1-j}. \tag{4.21}$$

Entsprechend gilt für $(\tilde{Z}_1, \dots, \tilde{Z}_n) := (X_n, \dots, X_1)$ die Beziehung $\langle \tilde{Z}_i, \tilde{Z}_j \rangle = \langle X_{n+1-i}, X_{n+1-j} \rangle = \langle X_i, X_j \rangle$ und folglich

$$P_{K_1} X_1 = \sum_{j=1}^{n-1} d_{n-1,j} \tilde{Z}_{n-j} = \sum_{j=1}^{n-1} d_{n-1,j} X_{j+1}. \quad (4.22)$$

Weiter gilt offenbar

$$\begin{aligned} v_{n-1} &= \|X_1 - P_{K_1} X_1\|^2 = \|X_{n+1} - P_{K_1} X_{n+1}\|^2 \\ &= \|X_n - \hat{X}_n\|^2. \end{aligned} \quad (4.23)$$

Einsetzen von (4.21) und (4.22) in (4.19) liefert

$$\hat{X}_{n+1} = a X_1 + \sum_{j=1}^{n-1} (d_{n-1,j} - a d_{n-1,n-j}) X_{n-j+1}. \quad (4.24)$$

Aus (4.20), (4.22) und (4.23) ergibt sich weiter

$$\begin{aligned} a &= \left(\langle X_{n+1}, X_1 \rangle - \sum_{j=1}^{n-1} d_{n-1,j} \langle X_{n+1}, X_{j+1} \rangle \right) \cdot \frac{1}{v_{n-1}} \\ &= \left(\gamma(n) - \sum_{j=1}^{n-1} d_{n-1,j} \gamma(n-j) \right) \cdot \frac{1}{v_{n-1}}. \end{aligned} \quad (4.25)$$

Da die Vektoren X_1, \dots, X_n linear unabhängig sind, sind die Koeffizienten $d_{n,j}$ in (4.14) eindeutig bestimmt. Es folgt also aus (4.24) und (4.14) durch Koeffizientenvergleich $d_{n,n} = a$ und damit sowohl (4.16) als auch (4.17).

Nun zu (4.18):

$$\begin{aligned} v_n &= \|X_{n+1} - P_n X_{n+1}\|^2 = \|X_{n+1} - P_{K_1} X_{n+1} - P_{K_2} X_{n+1}\|^2 \\ &= \|X_{n+1} - P_{K_1} X_{n+1}\|^2 + \|P_{K_2} X_{n+1}\|^2 - 2 \langle X_{n+1} - P_{K_1} X_{n+1}, P_{K_2} X_{n+1} \rangle \\ &= v_{n-1} - \|P_{K_2} X_{n+1}\|^2 \\ &= v_{n-1} - \|(X_1 - P_{K_1} X_1) \cdot a\|^2 \\ &= v_{n-1} - v_{n-1} a^2 = v_{n-1} (1 - d_{n,n}^2). \end{aligned}$$

Mit (4.16)–(4.18) steht dem Anwender eine sehr effiziente rekursive Methode zur Berechnung linearer Vorhersagen bei stationären Zeitreihen zur Verfügung.

Aus (4.20), (4.23) und $P_{K_1} X_{n+1} \perp X_1 - P_{K_1} X_1$ folgt

$$\begin{aligned} d_{n,n} &= \frac{\langle X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_1 - P_{K_1} X_1\|^2} \\ &= \frac{\langle X_{n+1} - P_{K_1} X_{n+1}, X_1 - P_{K_1} X_1 \rangle}{\|X_{n+1} - P_{K_1} X_{n+1}\| \cdot \|X_1 - P_{K_1} X_1\|}. \end{aligned} \quad (4.26)$$

Im Fall $H = L_2^{\mathbb{R}}$ und einer stationären, reellen und zentrierten Zeitreihe $X = (X_t : t \in \mathbb{Z})$ ist der letzte Ausdruck in (4.26) gerade die Korrelation von $X_{n+1} - P_{K_1} X_{n+1}$ und $X_1 - P_{K_1} X_1$, also die Korrelation der Anteile von X_{n+1} und X_1 , die senkrecht auf $K_1 = \text{span}\{X_2, \dots, X_n\}$ stehen. Man spricht auch von der Korrelation von X_1 und X_{n+1} nach (linearer) Beseitigung des Einflusses der zwischenzeitlichen Beobachtungen. Nach Definition 2.17 ist dies aber gerade die partielle Autokorrelation $\pi(n)$, sodass wir als Nebenprodukt der Levinson–Rekursion nunmehr den zurückgestellten Beweis für (2.39), also für

$$\pi(n) = d_{n,n}, \quad \forall n \geq 2, \quad (4.27)$$

erbracht haben. Wie schon in Kapitel 2 besprochen, erhält man eine *Stichprobenversion* $\hat{\pi}(n)$ von $\pi(n)$, die *empirische partielle Autokorrelation* dadurch, dass man in der Levinson–Rekursion $\gamma(h)$ durch die Stichprobenautokovarianz $\hat{\gamma}(h)$, siehe (2.29), ersetzt und $\hat{\pi}(n)$ dann gleich dem berechneten Wert $\hat{d}_{n,n}$ setzt.

Aufgaben

Aufgabe 4.1 Gegeben sei der Prozess $X = (X_t : t \in \mathbb{N}_0)$ mit

$$X_t = X_{t-1} + e_t + (\alpha - 1)e_{t-1}, \quad t \geq 1, \quad X_0 := e_0,$$

wobei $\alpha \in \mathbb{R}$ und $e_t, t \in \mathbb{N}_0$, stochastisch unabhängige, reelle, zentrierte, quadratintegrierbare Zufallsvariable seien. Man berechne für $s > t$ die optimale Vorhersage $\hat{P}_t X_s := E(X_s | \mathcal{B}_t)$ (d.h. den bedingten Erwartungswert von X_s gegeben \mathcal{B}_t) für X_s mit $\mathcal{B}_t := \sigma(X_u, u \leq t)$ und zeige für

$$\hat{X}_s = \hat{P}_{s-1} X_s$$

die Darstellungen

$$\hat{X}_{t+1} = \alpha X_t + (1 - \alpha) \hat{X}_t, \quad \forall t \geq 1,$$

$$\hat{X}_{t+1} = \alpha \sum_{j=0}^t (1 - \alpha)^j X_{t-j}.$$

Die Methode heißt *exponentielles Glätten*, vgl. auch (1.6) und das spätere Beispiel 6.9.

Aufgabe 4.2 Sei (Ω, \mathcal{A}, Q) ein W–Raum, \mathcal{F} eine Teil- σ –Algebra von \mathcal{A} . Ist X eine \mathcal{A} – \mathbb{B} –messbare, reelle Funktion, so bezeichne \tilde{X} die Äquivalenzklasse aller \mathcal{A} –messbaren Funktionen Y mit $X = Y[Q]$. Dann ist der Raum $L_2^{\mathbb{R}}$ der

quadratintegrierbaren Funktionen $L_2^{\mathbb{R}} = \{\tilde{X} : EX^2 < \infty\}$ mit dem Skalarprodukt $\langle \tilde{X}_1, \tilde{X}_2 \rangle := EX_1X_2$ ein reeller Hilbertraum mit Norm $\|\cdot\| := \sqrt{\langle \cdot, \cdot \rangle}$. Auf $L_2^{\mathbb{R}}$ werde ein Operator P erklärt durch $P\tilde{X} := \tilde{Y}, Y = E(X|\mathcal{F})$. Man zeige (genau zwischen Äquivalenzklassen und Repräsentanten unterscheiden):

- a) $P(c\tilde{X}_1 + d\tilde{X}_2) = c \cdot P\tilde{X}_1 + d \cdot P\tilde{X}_2, c, d \in \mathbb{R}$.
- b) $P(L_2^{\mathbb{R}}) =: L$ ist ein linearer, abgeschlossener Teilraum in $L_2^{\mathbb{R}}$; es gilt außerdem $\|P\tilde{X}\| \leq \|\tilde{X}\|$.
- c) $\langle P\tilde{X}_1, P\tilde{X}_2 \rangle = \langle \tilde{X}_1, P\tilde{X}_2 \rangle$
- d) $PP\tilde{X} = P\tilde{X}$
- e) $\langle P\tilde{X}_1, \tilde{X}_2 - P\tilde{X}_2 \rangle = 0$
- f) Für $\tilde{Y} \in L$ gilt $\|\tilde{X} - \tilde{Y}\| = \inf\{\|\tilde{X} - \tilde{Z}\| : \tilde{Z} \in L\}$ genau dann, wenn $\tilde{Y} = P\tilde{X}$.
- g) $P\tilde{X} = \tilde{X}$ gilt genau dann, wenn $\tilde{X} \in L$.

Aufgabe 4.3 Sei $X = (X(t) : t \in T), T \neq \emptyset$, ein reeller, zentrierter Gauß-Prozess. Für eine Zufallsvariable Y bezeichne \tilde{Y} wieder die zugehörige Äquivalenzklasse aller mit Y f.s. übereinstimmenden ZV.

Man zeige

$$E(X(t)|\mathcal{B}_S)^{\sim} = E_2(\tilde{X}(t)|L_S), t \in T, S \subset T.$$

Dabei ist \mathcal{B}_S die von den $X(s), s \in S$, erzeugte σ -Algebra und $L_S = \overline{\text{span}}\{\tilde{X}_s : s \in S\}$ im reellen Hilbertraum $L_2^{\mathbb{R}} = L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ und $E_2(\tilde{Y}|L_S)$ die Projektion von \tilde{Y} auf L_S .

Hinweis (ist auch zu zeigen): Bei gemeinsam normalverteilten ZV X, Y mit $EX = EY = 0$ gilt: $EXY = 0 \Leftrightarrow X$ und Y sind stochastisch unabhängig.

Aufgabe 4.4 (Die lineare Vorhersage ist nicht optimal)

Seien X, Z i.i.d. $\sim \mathcal{N}(0, 1)$. Man zeige für $Y := X^2 + Z$ und den bedingten Erwartungswert $E(Y|X)$ die Gleichung

$$E(Y|X) = X^2. \quad (4.28)$$

Weiter zeige man für $V = \text{span}\{1, X\}$

$$P_V(Y) =: Y^* = 1, \quad (4.29)$$

wobei $P_V(Y)$ die Projektion von Y auf den Teilraum V in $L_2^{\mathbb{R}}$ bezeichne. Man zeige außerdem für die Vorhersagefehler

$$E(Y - E(Y|X))^2 = 1 \quad (4.30)$$

und

$$E(Y - Y^*)^2 = 3. \quad (4.31)$$

Aufgabe 4.5 Gegeben seien drei quadratintegrierbare, zentrierte, reelle Zufallsvariable U_1, U_2, U_3 auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$(\text{Cov}(U_i, U_j))_{i,j=1,\dots,3} = \begin{pmatrix} 1 & 4 & 5 \\ 4 & 17 & 23 \\ 5 & 23 & 35 \end{pmatrix}.$$

Man bestimme mit dem Gram–Schmidt–Orthogonalisierungsverfahren unkorrelierte Zufallsvariable $\widehat{U}_1, \widehat{U}_2, \widehat{U}_3$ mit $\text{span}\{\widehat{U}_1\} = \text{span}\{U_1\}$, $\text{span}\{\widehat{U}_1, \widehat{U}_2\} = \text{span}\{U_1, U_2\}$ und $\text{span}\{\widehat{U}_1, \widehat{U}_2, \widehat{U}_3\} = \text{span}\{U_1, U_2, U_3\}$.

Der Spektralsatz für stationäre Zeitreihen

In diesem Kapitel werden wir den Spektralsatz für stationäre Zeitreihen formulieren und beweisen. Für zyklische Zeitreihen, die wir in Definition 2.9 bereits eingeführt haben und im ersten Abschnitt dieses Kapitels erneut aufgreifen, können wir die Aussage des Spektralsatzes, nämlich die Darstellung einer stationären Zeitreihe als Überlagerung von Schwingungen mit stochastischen Amplituden, besonders gut deutlich machen. Die wesentliche Idee der Formalisierung einer solchen Darstellung für beliebige stationäre Zeitreihen ist die Herstellung einer Isometrie, d.h. einer linearen, bijektiven und Skalarprodukt erhaltenden Abbildung zwischen dem Raum der quadratintegrierbaren Funktionen auf $(-\pi, \pi]$ (versehen mit dem Spektralmaß der gegebenen stationären Zeitreihe) und dem von der Zeitreihe selbst aufgespannten Teilraum der quadratintegrierbaren Zufallsvariablen. Wir werden in Lemma 5.5 sehen, dass unabhängig vom Maß der Raum der diesbezüglich quadratintegrierbaren Funktionen auf $(-\pi, \pi]$ durch die Abbildungen $e^{it\cdot} : \omega \mapsto e^{it\omega}$, $t \in \mathbb{Z}$, aufgespannt wird. Die angesprochene Isometrie kann dann durch die Zuordnung $e^{it\cdot} \mapsto X_t$, $t \in \mathbb{Z}$, realisiert werden. Es wird sich zeigen, dass die Zuordnungsvorschrift die Eigenschaften eines Integrals besitzt, so dass wir X_t in gewisser Weise als Integral über $e^{it\omega}$ in der Variablen ω auffassen können. Allerdings sind die Werte dieser *stochastischen Integration* Zufallsvariable, so dass wir in Abschnitt 5.2 noch eine Reihe von Schritten zu präzisieren haben. Der Spektralsatz selbst wird dann der Inhalt von Abschnitt 5.3 sein und eine für später sehr wichtige Substitutionsregel für stochastische Integrale (Abschnitt 5.4) schließt dieses Kapitel ab.

Man mag sich an dieser Stelle fragen, warum hier grundsätzlich nur von Frequenzen ω aus dem Intervall $(-\pi, \pi]$ die Rede ist. Der Grund hierfür liegt in der 2π -Periodizität der Funktionen $\omega \mapsto e^{it\omega}$ für jedes $t \in \mathbb{Z}$.

5.1 Die Spektraldarstellung zyklischer Zeitreihen

In Kapitel 2 hatten wir das zyklische Modell

$$X_t = \sum_{j=1}^M (e_{1j} \cos \omega_j t + e_{2j} \sin \omega_j t), \quad \forall t \in \mathbb{Z}, \quad (5.1)$$

kennen gelernt. Dort wurden e_{1j} und e_{2j} , $j = 1, 2, \dots, M$, als zentrierte, quadratintegrierbare und unkorrelierte reelle Zufallsvariable angenommen und ω_j bezeichneten feste Kreisfrequenzen, die aufgrund der 2π -Periodizität von $\cos(\omega t)$ und $\sin(\omega t)$ für jedes $t \in \mathbb{Z}$ aus $(-\pi, \pi]$ angenommen werden können, ohne den Wert von X_t zu verändern.

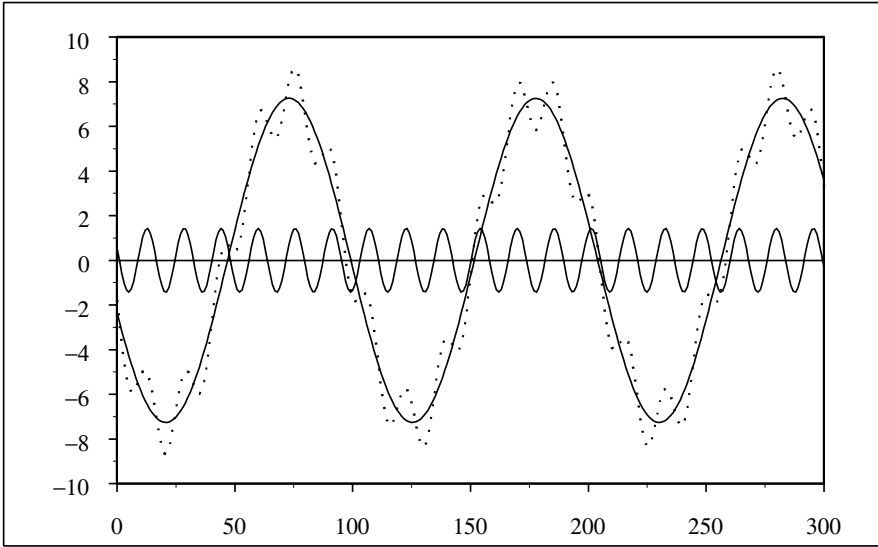


Abb. 5.1 Simulation einer zyklischen Zeitreihe (gepunktet) mit $M = 2$ und Frequenzen $\omega_1 = 0.06$, $\omega_2 = 0.4$ sowie Varianzen der zugehörigen Amplituden $\sigma_1^2 = 9$ bzw. $\sigma_2^2 = 1$ (vgl. (5.1)) und die beiden Schwingungsanteile getrennt (durchgezogen)

Beispiel 5.1 (Zyklische Zeitreihe)

Mit $M = 2$ und den Frequenzen $\omega_1 = 0.06$ und $\omega_2 = 0.4$ werde die zyklische Zeitreihe (5.1) betrachtet. Die zufälligen Amplituden e_{kj} seien unabhängig und normalverteilt mit Erwartungswert Null und Varianzen $Ee_{11}^2 = Ee_{21}^2 = 9$ sowie $Ee_{12}^2 = Ee_{22}^2 = 1$. Eine Realisierung dieser speziellen zyklischen Zeitreihe ist in Abbildung 5.1 gegeben. Man erkennt sehr deutlich die Überlagerung der beiden unterschiedlichen Schwingungsanteile und auch deren unterschiedliche Intensität, die proportional zur Varianz der zufälligen Amplituden ist.

□

Für die weitere Diskussion der allgemeinen zyklischen Zeitreihe lässt sich (5.1) in eine wesentlich elegantere komplexe Form umschreiben. Unter Beachtung

der Identität

$$e_1 \cos \alpha + e_2 \sin \alpha = \frac{1}{2} (e_1 - ie_2) e^{i\alpha} + \frac{1}{2} (e_1 + ie_2) e^{-i\alpha} \quad (5.2)$$

erhält man nämlich mit $Z_j := (e_{1j} - ie_{2j})/2$, $j = 1, 2, \dots, M$, $Z_{-j} := \overline{Z_j}$, $Z_0 := 0$ sowie $\omega_{-j} := -\omega_j$ und $\omega_0 := 0$ die zu (5.1) äquivalente Darstellung

$$X_t = \sum_{j \in J} Z_j e^{i\omega_j t}, \quad \forall t \in \mathbb{Z}, \quad (5.3)$$

mit $J = \{-M, -M+1, \dots, M\}$. Man rechnet leicht nach, dass die Zufallsvariablen Z_j , $j \in J$, aufgrund der Voraussetzungen an die e_{1j}, e_{2j} , $j \in J$, ebenfalls zentrierte und unkorrelierte, nun aber komplexwertige, Zufallsvariable sind (siehe Aufgabe 5.1).

Fasst man die endliche Summe in (5.3) in formaler Analogie zur üblichen Integration bzgl. reellwertiger Maße nun als „Integral“ bzgl. eines „diskreten Maßes“ Z über $(-\pi, \pi]$ mit Trägerpunkten ω_j , $j \in J$, und „zufallswertiger Masse“ Z_j auf ω_j auf, so kann man (5.3) – zunächst rein formal – als Integral

$$X_t = \int_{(-\pi, \pi]} e^{it\omega} Z(d\omega) \quad (5.4)$$

schreiben.

Für eine Borel-Menge $B \subset (-\pi, \pi]$ wird man dann ihr „Maß“ unter Z als $Z(B) = \sum_{j \in J} 1_B(\omega_j) Z_j$ setzen. $Z(B)$ ist dann offensichtlich eine zentrierte Zufallsvariable, genauer ein Element des Teilraumes $\text{span}\{Z_j : j \in J\}$ von $L_2^{\mathbb{C}}$. Das „Maß“ Z nimmt also Werte in $L_2^{\mathbb{C}}$ an. Die Unkorreliertheit der Z_j impliziert, dass für disjunkte Mengen B_1 und B_2 die Bildwerte $Z(B_1)$ und $Z(B_2)$ auch unkorreliert sind, was zusammen mit ihrer Zentriertheit gerade bedeutet, dass sie im Raum $L_2^{\mathbb{C}}$ zueinander orthogonal sind.

Die Darstellung (5.3) bzw. in formaler Analogie auch (5.4) bezeichnet man als Spektraldarstellung der zyklischen Zeitreihe $X = (X_t : t \in \mathbb{Z})$.

Da alle Zufallsvariablen Z_j , wie auch alle $Z(B)$, zentriert sind, quantifizieren wir das Gewicht der Schwingung mit Frequenz ω_j in der Darstellung (5.3) durch die Varianz der Zufallsvariablen Z_j . Ersetzen wir nun in der Darstellung $Z(B) = \sum_{j \in J} 1_B(\omega_j) Z_j$ die Zufallsvariablen Z_j durch die Varianzen dieser Zufallsvariablen, so erhalten wir das diskrete Spektralmaß μ der zyklischen Zeitreihe (vgl. Beispiel 3.6 für den Fall $M = 1$ und das spätere Beispiel 5.9). Auf diesen Zusammenhang zwischen Spektraldarstellung einer zyklischen Zeitreihe als Überlagerung von Schwingungen mit stochastischen Amplituden (repräsentiert durch das zufallswertige „Maß“ Z) und dem Spektralmaß als Varianz des „Maßes“ Z sei besonders hingewiesen.

Unser Ziel im Folgenden ist es, eine Darstellung (5.4) (den sog. Spektralsatz)

für alle zentrierten, stationären (komplex- oder reellwertigen) Zeitreihen herzuleiten, wobei dann überabzählbar viele Kreisfrequenzen ω zugelassen werden müssen. Der Übergang von Integralen bzgl. diskreter „Maße“ zu allgemeinen „Maßen“ ist in unserem Zusammenhang ähnlich demjenigen in der üblichen Integrationstheorie für reellwertige Maße. Auch werden wir den Zusammenhang der Spektraldarstellung mit dem Spektralmaß für die allgemeine Situation herstellen. Bevor wir soweit sind, müssen wir aber im folgenden Abschnitt einige Begriffsbildungen präzisieren.

5.2 Maße mit orthogonalen Werten und ein stochastisches Integral

In diesem Abschnitt wollen wir präzisieren, was wir unter Maßen mit Zufallsvariablen als Werten verstehen wollen und wie bezüglich dieser Maße integriert werden kann. In der folgenden allgemeinen Definition 5.2 erlauben wir für die Maße sogar Werte in einem beliebigen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$. Der uns hauptsächlich interessierende Fall wird aber der Raum $H = L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ der quadratintegrierbaren Zufallsvariablen sein.

Definition 5.2 (Maß mit orthogonalen Werten (MOW))

Ein Maß mit orthogonalen Werten (MOW) Z über einem Messraum $(\mathcal{X}, \mathcal{B})$ ist eine Abbildung $Z : \mathcal{B} \rightarrow H$ in einen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ mit der Eigenschaft, dass die durch

$$\mu_Z(B) := \|Z(B)\|^2, \quad \forall B \in \mathcal{B}, \quad (5.5)$$

definierte Mengenfunktion μ_Z ein (notwendigerweise endliches) Maß über $(\mathcal{X}, \mathcal{B})$ ist mit der Eigenschaft

$$\langle Z(B_1), Z(B_2) \rangle = \mu_Z(B_1 \cap B_2), \quad \forall B_1, B_2 \in \mathcal{B}. \quad (5.6)$$

Offenbar folgt aus (5.6)

$$Z(B_1) \perp Z(B_2) \quad \text{falls} \quad \mu_Z(B_1 \cap B_2) = 0, \quad (5.7)$$

was den Zusatz „mit orthogonalen Werten“ erklärt.

Von einem Maß erwartet man eine Art von σ -Additivität, also für paarweise disjunkte Mengen $B_j \in \mathcal{B}$ die Gültigkeit der Gleichheit

$$Z\left(\sum_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} Z(B_j), \quad (5.8)$$

wobei, wie üblich, die Summenzeichen „ \sum “ und „+“ bei Mengen deren Vereinigung bedeuten, falls diese paarweise disjunkt sind.

Tatsächlich folgt (5.8) schon aus (5.6). Es gilt nämlich mit $Y := Z(B_1 + B_2) - Z(B_1) - Z(B_2)$ unter Verwendung von (5.6) die Gleichung $\langle Y, Y \rangle = 0$, also $Y = 0$. Dies ist (5.8) für zwei Mengen, woraus die Aussage dann auch für endlich viele Mengen (*endliche Additivität*) mit vollständiger Induktion folgt.

Insbesondere erhält man

$$Z\left(\sum_{j=1}^{\infty} B_j\right) = \sum_{j=1}^n Z(B_j) + Z\left(\sum_{j=n+1}^{\infty} B_j\right).$$

Gemäß

$$\left\|Z\left(\sum_{j=n+1}^{\infty} B_j\right)\right\|^2 = \mu_Z\left(\sum_{j=n+1}^{\infty} B_j\right) \longrightarrow 0 \text{ für } n \rightarrow \infty$$

folgt hieraus (5.8).

Der Spektralsatz für Zeitreihen ergibt sich nun im Wesentlichen aus einem allgemeinen Isometriesatz für Hilberträume. Wir werden diesen Isometriesatz angeben und auch beweisen. Dazu benötigen wir einige Bezeichnungen.

Für einen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ sei $\{u(t) : t \in T\}$ eine Familie von Elementen aus H , indiziert mit einer beliebigen Indexmenge T . $\overline{\text{span}}\{u(t) : t \in T\}$ bezeichne den von $\{u(t) : t \in T\}$ aufgespannten linearen Teil-Hilbertraum, d.h. die abgeschlossene Hülle des erzeugten Vektorraums

$$\text{span}\{u(t) : t \in T\} := \left\{ \sum_{i=1}^n c_i u(t_i) : c_i \in \mathbb{C}, t_i \in T, n \in \mathbb{N} \right\}.$$

Dann gilt das folgende Resultat.

Satz 5.3 (*Isometriesatz*)

Seien $(H_i, \langle \cdot, \cdot \rangle_i)$, $i = 1, 2$, zwei Hilberträume, T eine beliebige Indexmenge, und es gelte $H_1 = \overline{\text{span}}\{u(t) : t \in T\}$, $H_2 = \overline{\text{span}}\{v(t) : t \in T\}$, sowie

$$\langle u(s), u(t) \rangle_1 = \langle v(s), v(t) \rangle_2, \quad \forall s, t \in T. \quad (5.9)$$

Dann gibt es eine lineare, bijektive Abbildung $\psi : H_1 \rightarrow H_2$, die Skalarprodukt erhaltend ist, d.h. mit

$$\langle x, y \rangle_1 = \langle \psi(x), \psi(y) \rangle_2, \quad \forall x, y \in H_1, \quad (5.10)$$

eine sog. Isometrie, die jedes Element $u(t)$ in $v(t)$ überführt, d.h.

$$\psi(u(t)) = v(t), \quad \forall t \in T. \quad (5.11)$$

Die Isometrie ψ ist durch (5.11) eindeutig festgelegt.

Beweis: Man definiere $\psi\left(\sum_{i=1}^n c_i u(t_i)\right) = \sum_{i=1}^n c_i v(t_i)$. Dann ist ψ wohldefiniert auf $\text{span}\{u(t) : t \in T\}$, denn aus $\sum_{i=1}^n c_i u(t_i) = 0$ folgt

$$\begin{aligned} \left\| \sum_i c_i v(t_i) \right\|_2^2 &= \sum_{i,j} c_i \overline{c_j} \langle v(t_i), v(t_j) \rangle_2 = \sum_{i,j} c_i \overline{c_j} \langle u(t_i), u(t_j) \rangle_1 \\ &= \left\| \sum_i c_i u(t_i) \right\|_1^2 = 0, \end{aligned}$$

also

$$\sum_{i=1}^n c_i v(t_i) = 0.$$

Offenbar ist ψ auch linear auf $\text{span}\{u(t) : t \in T\}$ und dort wegen einer analogen Rechnung wie oben ebenfalls Skalarprodukt erhaltend:

$$\langle x, y \rangle_1 = \langle \psi(x), \psi(y) \rangle_2 \quad \forall x, y \in \text{span}\{u(t) : t \in T\}. \quad (5.12)$$

Aus (5.12) folgt, dass jede Cauchy-Folge (u_n) in $\text{span}\{u(t) : t \in T\}$ in eine Cauchy-Folge in $\text{span}\{v(t) : t \in T\}$ transformiert wird, und es gilt

$$\lim_{n \rightarrow \infty} u_n = 0 \Leftrightarrow \lim_{n \rightarrow \infty} \psi(u_n) = 0 \quad (5.13)$$

wegen $\|u_n\|_1 = \|\psi(u_n)\|_2$. Definiert man also für $u = \lim_{n \rightarrow \infty} u_n$ den Wert $\psi(u)$ durch $\psi(u) := \lim_{n \rightarrow \infty} \psi(u_n)$, so ist $\psi(u)$ wegen (5.13) wohldefiniert, und es gilt für $x, y \in H_1$ mit $x_n \rightarrow x, y_n \rightarrow y, x_n, y_n \in \text{span}\{u(t) : t \in T\}$ unter Verwendung der Stetigkeit des inneren Produkts in beiden Komponenten, siehe Satz A.9

$$\langle x, y \rangle_1 = \lim_{n \rightarrow \infty} \langle x_n, y_n \rangle_1 = \lim_{n \rightarrow \infty} \langle \psi(x_n), \psi(y_n) \rangle_2 = \langle \psi(x), \psi(y) \rangle_2.$$

Ebenso überträgt sich die Linearität von ψ unmittelbar von $\text{span}\{u(t) : t \in T\}$ auf $\overline{\text{span}}\{u(t) : t \in T\}$. \square

Wir wollen nun, ausgehend von einem MOW Z (vgl. Definition 5.2), das Integral $\int f dZ$ definieren. Dabei bezeichnet $L_2^{\mathbb{C}}(\mu) := L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu)$ den Raum der komplexen und quadratintegrierbaren Funktionen auf einem Maßraum $(\mathcal{X}, \mathcal{B}, \mu)$.

Lemma 5.4

Sei $(\mathcal{X}, \mathcal{B}, \mu)$ ein endlicher Maßraum und $(H_0, \langle \cdot, \cdot \rangle)$ ein Hilbertraum.

a) Ist ein MOW $Z : \mathcal{B} \rightarrow H_0$ gegeben mit $H_Z := \overline{\text{span}}\{Z(B) : B \in \mathcal{B}\} = H_0$ und $\mu_Z = \mu$, so wird durch

$$Z(B) = \psi(1_B), \quad \forall B \in \mathcal{B}, \quad (5.14)$$

eine Isometrie ψ

$$\psi : L_2^{\mathbb{C}}(\mu) \rightarrow H_0. \quad (5.15)$$

eindeutig definiert.

b) Ist umgekehrt ψ eine Isometrie wie in (5.15), so wird durch die Gleichung (5.14) ein MOW $Z : \mathcal{B} \rightarrow H_0$ definiert mit $H_Z = H_0$ und $\mu_Z = \mu$.

Beweis: a) Ist Z ein MOW mit $\mu_Z = \mu$ und $H_Z = H_0$, so gilt für $B_1, B_2 \in \mathcal{B}$

$$\langle 1_{B_1}, 1_{B_2} \rangle_{\mu} = \int 1_{B_1} \cdot 1_{B_2} d\mu = \mu(B_1 \cap B_2) = \langle Z(B_1), Z(B_2) \rangle. \quad (5.16)$$

Da die Indikatoren $\{1_B : B \in \mathcal{B}\}$ den Hilbertraum $(L_2^{\mathbb{C}}(\mu), \langle \cdot, \cdot \rangle_{\mu})$ aufspannen, bestimmt die Zuordnung $1_B \mapsto Z(B)$ nach Satz 5.2 genau eine Isometrie ψ .

b) Ist nun umgekehrt ψ eine Isometrie und Z durch (5.14) definiert, so ist Z ein MOW mit $\mu_Z = \mu$ und $H_Z = H_0$, denn (5.6) gilt wegen

$$\langle Z(B_1), Z(B_2) \rangle = \langle \psi(1_{B_1}), \psi(1_{B_2}) \rangle = \langle 1_{B_1}, 1_{B_2} \rangle_{\mu} = \mu(B_1 \cap B_2).$$

□

Jede Isometrie $\psi : L_2^{\mathbb{C}}(\mu) \rightarrow H_0$ ist also gemäß (5.14) die „Fortsetzung“ eines MOW's Z von den Mengen B , d.h. von den Indikatorfunktionen 1_B , auf den vollen Raum $L_2^{\mathbb{C}}(\mu)$. Wegen (5.14) gibt es nur ein einziges MOW Z , welches zu ψ gehört. Eine Isometrie ψ mit zugehörigem MOW Z besitzt die Eigenschaften eines Integrals (siehe die unten stehenden Rechenregeln). Deshalb schreibt man intuitiver

$$\psi(f) =: \int_{\mathcal{X}} f(x) Z(dx) =: \int f dZ, \quad \forall f \in L_2^{\mathbb{C}}(\mu_Z), \quad (5.17)$$

und nennt $\psi(\cdot) = \int \cdot dZ$ ein *stochastisches Integral* bzgl. Z .

Zugegebenerweise ist diese Definition von $\int f dZ$ nicht besonders anschaulich. Ein großer Vorteil der Definition über eine Isometrie ψ ist allerdings, dass wir die gesamten Eigenschaften der Isometrien ohne weiteren Aufwand direkt zu Rechenregeln des stochastischen Integrals umformulieren können. So folgen aus der Isometrieeigenschaft von $\int f dZ = \psi(f)$ unmittelbar die nachstehenden *Rechenregeln* für alle $a, b \in \mathbb{C}$, $B \in \mathcal{B}$, $f, g, f_n \in L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu_Z)$ und $n \rightarrow \infty$

$$\int 1_B dZ = Z(B) \quad (5.18)$$

$$\int (af + bg) dZ = a \int f dZ + b \int g dZ \quad (5.19)$$

$$f_n \rightarrow f \text{ in } L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu_Z) \Leftrightarrow \int f_n dZ \rightarrow \int f dZ \text{ in } H_Z \quad (5.20)$$

$$\left\langle \int f dZ, \int g dZ \right\rangle = \int f \bar{g} d\mu_Z \quad (5.21)$$

$$\left\| \int f dZ \right\|^2 = \int |f|^2 d\mu_Z \quad (5.22)$$

$$f = g \Leftrightarrow \int f dZ = \int g dZ. \quad (5.23)$$

Die Rückrichtungen in (5.20) und (5.23) sind eine Spezialität des hier eingeführten Integrals. Sie sind Ausdruck der Isometrieeigenschaft von ψ . Bekanntlich gelten derartige Aussagen für das übliche Maßintegral nicht.

An dieser Stelle wollen wir das stochastische Integral mit der abstrakten, aber für die spätere Anwendung sehr schlagkräftigen, Definition (5.17) zunächst stehen lassen. Im Anschluss an den Spektralsatz werden wir eine Verständnis- und Interpretationshilfe für den uns speziell interessierenden Fall $(\mathcal{X}, \mathcal{B}) = ((-\pi, \pi], (-\pi, \pi] \cap \mathbb{B})$ und $H_0 = \overline{\text{span}}\{X_t : t \in \mathbb{Z}\} \subset L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ in Form einer Art Riemann-Approximation angeben.

Später benötigen wir noch die folgende hilfreiche Aussage.

Lemma 5.5

Ist μ ein endliches Maß über $(\mathcal{X}, \mathcal{B}) = ((-\pi, \pi], (-\pi, \pi] \cap \mathbb{B})$, so gilt

$$L_2^{\mathbb{C}}(\mu) := L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu) = \overline{\text{span}}\{e^{it\cdot} : t \in \mathbb{Z}\} \quad (5.24)$$

mit den Exponentialfunktionen $e^{it\cdot} : (-\pi, \pi] \rightarrow \mathbb{C}$, $\omega \mapsto e^{it\omega}$.

Beweis: Die Inklusion „ \supset “ ist trivial. Für die umgekehrte Richtung benutzen wir die Tatsache, dass sich jede stetige Funktion f auf $(-\pi, \pi]$ mit $f(-\pi) = f(\pi)$ gleichmäßig, also in der sup-Norm, durch ein trigonometrisches Polynom $P(t) = \sum_{n=-N}^N c_n e^{int}$ approximieren lässt, siehe Rudin (1970), Theorem 4.25. Wegen

$$\int_{(-\pi, \pi]} |f - P|^2 d\mu \leq \mu(-\pi, \pi] \cdot \sup^2 \{|f(t) - P(t)| : -\pi \leq t \leq \pi\}$$

ist dies auch eine Approximation in $L_2^{\mathbb{C}}(\mu)$. Weiter gibt es zu jeder stetigen und beschränkten Funktion g auf $(-\pi, \pi]$ eine stetige Funktion f auf $[-\pi, \pi]$ mit $f(-\pi) = f(\pi)$ derart, dass der $L_2^{\mathbb{C}}(\mu)$ -Abstand von g und f beliebig klein wird. Dazu definiere man für $-\pi < s < \pi$ eine lineare Funktion f_s auf

$(-\pi, s]$ mit $f_s(-\pi) = g(\pi)$, $f_s(s) = g(s)$ und setze $f_s = g$ auf $(s, \pi]$. Für $s \rightarrow -\pi$ wird dann wegen der Endlichkeit von μ der $L_2^{\mathbb{C}}(\mu)$ -Abstand von f_s und g beliebig klein.

Berücksichtigt man nun noch, dass die stetigen und beschränkten Funktionen in $L_2^{\mathbb{C}}(\mu)$ dicht liegen, siehe Dunford/Schwartz (1957), Lemma IV 8.19, so folgt daraus die Behauptung. \square

Bemerkung 5.6 Anstelle des Zitats aus Dunford/Schwartz (1957) kann man auch direkt argumentieren. Siehe dazu Aufgabe 5.1. \square

5.3 Der Spektralsatz

Nun können wir den angekündigten Spektralsatz formulieren und beweisen.

Satz 5.7 (Spektralsatz)

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre und zentrierte Zeitreihe mit Kovarianzfunktion $\gamma : \mathbb{Z} \rightarrow \mathbb{C}$ und zugehörigem Spektralmaß μ , d.h.

$$\gamma(t) = \int_{(-\pi, \pi]} e^{it\omega} \mu(d\omega), \quad \forall t \in \mathbb{Z}. \quad (5.25)$$

Dann definiert die Abbildung

$$e^{it\cdot} \mapsto X_t, \quad \forall t \in \mathbb{Z}, \quad (5.26)$$

eine Isometrie $\psi : L_2^{\mathbb{C}}(\mu) \rightarrow \overline{\text{span}}\{X_t : t \in \mathbb{Z}\} \subset L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$, wobei $L_2^{\mathbb{C}}(\mu) = L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu)$ mit $\mathcal{X} = (-\pi, \pi]$ und $\mathcal{B} = \mathcal{X} \cap \mathbb{B}$ ist. Ist Z das gemäß (5.14) zu ψ korrespondierende MOW, also $Z(B) = \psi(1_B)$, so gilt $\mu_Z = \mu$, $H_Z = \overline{\text{span}}\{Z(B) : B \in \mathcal{B}\} = \overline{\text{span}}\{X_t : t \in \mathbb{Z}\}$ sowie

$$\psi(f) = \int_{(-\pi, \pi]} f dZ, \quad \forall f \in L_2^{\mathbb{C}}(\mu). \quad (5.27)$$

Insbesondere gilt also

$$X_t = \int_{(-\pi, \pi]} e^{it\omega} Z(d\omega), \quad \forall t \in \mathbb{Z}. \quad (5.28)$$

Z ist durch (5.28) und die Forderung $\mu_Z = \mu$ eindeutig bestimmt.

Beweis: Mit dem üblichen inneren Produkt $\langle \cdot, \cdot \rangle_\mu$ in $L_2^{\mathbb{C}}(\mu)$ und $\langle \cdot, \cdot \rangle$ in $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ gilt nach (5.25)

$$\begin{aligned}\langle e^{is\cdot}, e^{it\cdot} \rangle_\mu &= \int e^{is\omega} e^{-it\omega} \mu(d\omega) = \gamma(s-t) \\ &= \langle X_s, X_t \rangle, \quad \forall s, t \in \mathbb{Z}.\end{aligned}\tag{5.29}$$

Wegen Lemma 5.5 ist der Isometriesatz auf die Räume $L_2^{\mathbb{C}}(\mu)$ und $\overline{\text{span}}\{X_t : t \in \mathbb{Z}\}$ anwendbar. Damit ist ψ eine Isometrie. Bis auf die letzte Eindeutigkeitsaussage folgen nun alle Behauptungen aus Lemma 5.4.

Zum Nachweis der Eindeutigkeit seien nun Z_1 und Z_2 zwei MOW's, die (5.28) erfüllen. Für die korrespondierenden Isometrien ψ_{Z_1}, ψ_{Z_2} folgt deren Gleichheit zunächst auf $\{e^{it\cdot} : t \in \mathbb{Z}\}$ und dann auch auf $L_2^{\mathbb{C}}(\mu)$, woraus wegen (5.14) sofort $Z_1 = Z_2$ folgt. \square

Bemerkung 5.8

Es sei darauf hingewiesen, dass $\psi(f) = \int_{(-\pi, \pi]} f dZ$ in (5.27) als Element von $\overline{\text{span}}\{X_t : t \in \mathbb{Z}\}$ aufgrund der in Satz 5.7 vorausgesetzten Zentriertheit der Zeitreihe X stets selbst zentrierte Zufallsvariable darstellen. \square

Um das stochastische Integral $\int f dZ$ im Zusammenhang mit dem Spektralsatz besser verstehen und interpretieren zu können, betrachten wir eine Art Riemann-Approximation. Sei dazu Z das zu einer stationären und zentrierten Zeitreihe $X = (X_t : t \in \mathbb{Z})$ gehörende MOW mit zugehörigem Spektralmaß μ . f sei eine zu integrierende Funktion auf $(-\pi, \pi]$ und $\omega_0 = -\pi < \omega_1 < \dots < \omega_n = \pi$ eine Folge von Unterteilungspunkten. Falls $f_n = \sum_{j=1}^n f(\omega_j) 1_{(\omega_{j-1}, \omega_j]}$ im Raum $L_2^{\mathbb{C}}(\mu)$ gegen f konvergiert (was z.B. dann der Fall ist, wenn f gleichmäßig stetig ist und die maximale Differenz der Unterteilungspunkte gegen Null konvergiert), dann wird das stochastische Integral $\int f dZ$ wegen Rechenregel (5.20) – ebenfalls im quadratischen Mittel – approximiert durch die folgende *Riemann-Summe*

$$\int f_n dZ = \sum_{j=1}^n f(\omega_j) Z((\omega_{j-1}, \omega_j]) . \tag{5.30}$$

Dabei sind die Zufallsvariablen $Z_j := Z((\omega_{j-1}, \omega_j])$, $j = 1, \dots, n$, zentriert und wegen (5.7) auch unkorreliert. Wenden wir diese Approximation auf die speziellen (und im Übrigen gleichmäßig stetigen) Funktionen $e^{it\cdot}$ an, so erhalten wir unter Beachtung von (5.28) die folgende Approximation an X_t im Sinne der Konvergenz im quadratischen Mittel

$$X_t \approx \sum_{j=1}^n e^{it\omega_j} Z_j . \tag{5.31}$$

Die rechte Seite in (5.31) stellt demnach eine zyklische Zeitreihe, also eine Überlagerung von Schwingungen mit Frequenzen ω_j , $j = 1, \dots, n$, dar. Die Intensität des Schwingungsanteils mit Frequenz ω_j , gemessen durch die Varianz der Zufallsvariable Z_j , ergibt sich wegen (5.22) zu

$$\begin{aligned}\mathrm{Var} Z_j &= E|Z_j|^2 = \left\| \int 1_{(\omega_{j-1}, \omega_j]} dZ \right\|^2 \\ &= \int |1_{(\omega_{j-1}, \omega_j]}|^2 d\mu = \mu(\omega_{j-1}, \omega_j].\end{aligned}\quad (5.32)$$

Die Intensität des Schwingungsanteils mit Frequenz ω_j wird also durch die Masse des Spektralmaßes in der Nähe dieser Frequenz bestimmt. Diese Überlegung erlaubt das Spektralmaß bzw. die Spektraldichte dahingehend zu interpretieren, dass die Bereiche im Intervall $(-\pi, \pi]$, die eine besonders große Masse bzgl. des Spektralmaßes besitzen, dazu führen, dass in der Spektraldarstellung der Zeitreihe als Überlagerung von Schwingungen den Frequenzen aus diesen Bereichen ein besonders großes Gewicht zukommt. Falls eine Spektraldichte existiert, dann können ausgeprägte Maxima an bestimmten Frequenzen so verstanden werden, dass Realisierungen der Zeitreihe ausgeprägte Schwingungsanteile mit diesen Frequenzen widerspiegeln. Dass dies der Fall ist, haben wir in einem konkreten Simulationsbeispiel in Kapitel 3 (vgl. Beispiel 3.13) schon gesehen.

Ein Wort der Vorsicht zur Approximation in (5.31) ist angebracht. Trotz Stationarität gilt diese nämlich im Allgemeinen nicht gleichmäßig für alle t mit demselben n und denselben Z_j , sondern nur für ein festes t . Deshalb kann man die approximierende zyklische Zeitreihe auf der rechten Seite in (5.31) auch nicht zur Vorhersage künftiger Werte der Zeitreihe verwenden. Tatsächlich können bei zyklischen Zeitreihen, sobald ein genügend langer Anfangsabschnitt bekannt ist, alle zukünftigen Werte exakt berechnet werden. Schon daraus erkennt man, dass sie als Approximation allgemeiner stationärer Zeitreihen nicht geeignet sein können.

Ausgangspunkt der Überlegungen zum Spektralsatz war die Summenformel (5.3) für das zyklische Modell, die wir im folgenden Beispiel als Spezialfall des Spektralsatzes wiedererkennen werden.

Beispiel 5.9 (Zeitreihen mit diskretem Spektralmaß)

a) Sei $X = (X_t : t \in \mathbb{Z})$ eine zentrierte und stationäre Zeitreihe mit diskretem Spektralmaß μ_X . Die Trägerpunkte von μ_X seien $\{\omega_j : j \in J\} \subset (-\pi, \pi]$. Dann gilt in $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ die Gleichheit

$$X_t = \sum_{j \in J} e^{it\omega_j} Z_j, \quad \forall t \in \mathbb{Z}, \quad (5.33)$$

wobei die Elemente $Z_j := Z(\{\omega_j\})$, $j \in J$, in $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ zentriert und paarweise unkorreliert sind mit $\sum_{j \in J} \|Z_j\|^2 < \infty$. Als Folgerung aus (5.33) erhält man eine Zerlegung der Varianz $\mathrm{Var} X_t = \|X_t\|^2$ von X_t in die Summe der Varianzanteile $\mathrm{Var} Z_j = \|Z_j\|^2$ zur Frequenz ω_j gemäß

$$\mathrm{Var} X_t = \sum_{j \in J} \mathrm{Var} Z_j. \quad (5.34)$$

Das zu X gehörige MOW Z ist darstellbar als

$$Z(B) = \sum_{j \in J} Z_j 1_B(\omega_j), \quad \forall B \in \mathcal{B} = (-\pi, \pi] \cap \mathbb{B}, \quad (5.35)$$

und das Spektralmaß μ_X als

$$3\mu_X(B) = \sum_{j \in J} \|Z_j\|^2 1_B(\omega_j), \quad \forall B \in \mathcal{B}. \quad (5.36)$$

b) Wird umgekehrt eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ durch (5.33) definiert mit $\omega_j \in (-\pi, \pi]$, $j \in J$ (abzählbar), und zentrierten, paarweise unkorrelierten Elementen $Z_j \in L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ mit $\sum_{j \in J} \|Z_j\|^2 < \infty$, so ist X zentriert und stationär mit zugehörigem MOW Z gemäß (5.35) und Spektralmaß μ_X gemäß (5.36). \square

Beweis: a) Zunächst gilt für alle $t \in \mathbb{Z}$ die Gleichheit

$$e^{it\cdot} = \sum_{j \in J} e^{it\omega_j} 1_{\{\omega_j\}} \mu_X - f.s. \quad (5.37)$$

Wegen

$$\left| \sum_{j \in J_0} e^{it\omega_j} 1_{\{\omega_j\}} \right|^2 = \sum_{j \in J_0} 1_{\{\omega_j\}} \quad \forall J_0 \subset J,$$

bilden die Partialsummen in (5.37) in $L_2^{\mathbb{C}}(\mu_X)$ eine Cauchy-Folge, konvergieren wegen (5.37) also in $L_2^{\mathbb{C}}(\mu_X)$ gegen $e^{it\cdot}$. Die Rechenregeln (5.18) – (5.22) liefern dann zusammen mit dem Spektralsatz die Darstellung (5.33) mit $Z_j = Z(\{\omega_j\})$. Wegen $Z_j \in \overline{\text{span}}\{X_t : t \in \mathbb{Z}\}$ sind die Z_j zentriert und wegen $\langle Z_j, Z_k \rangle = \langle Z(\{\omega_j\}), Z(\{\omega_k\}) \rangle = \mu_X(\{\omega_j\} \cap \{\omega_k\}) = 0$ für $j \neq k$ auch paarweise unkorreliert.

Analog folgt aus der Gleichheit

$$1_B = \sum_{j \in J} 1_B(\omega_j) 1_{\{\omega_j\}} \mu_X - f.s.$$

die Darstellung (5.35) und damit wegen $\mu_X(B) = \|Z(B)\|^2$ auch (5.36), insbesondere $\sum_{j \in J} \|Z_j\|^2 = \mu_X(-\pi, \pi] < \infty$.

b) Wegen der Zentriertheit und Unkorreliertheit der Z_j ergibt sich aus der Darstellung (5.33) zunächst, dass X stationär ist mit

$$\gamma_X(h) = \sum_{j \in J} \|Z_j\|^2 e^{i\omega_j h} = \int_{(-\pi, \pi]} e^{ih\cdot} d\mu_X,$$

wobei μ_X das in Gleichung (5.36) definierte Maß ist, das sich somit als das eindeutig bestimmte Spektralmaß erweist. Die in (5.35) definierte Mengenfunktion Z ist offenbar ein MOW mit $\|Z(B)\|^2 = \mu_X(B) \forall B \in \mathcal{B}$. Wegen der $L_2^{\mathbb{C}}(\mu_X)$ -Konvergenz in (5.37) und wegen $Z(\{\omega_j\}) = Z_j$ in (5.35) folgt

$$\int e^{it \cdot} dZ = \sum_{j \in J} e^{it\omega_j} Z_j = X_t, \quad \forall t \in \mathbb{Z}.$$

Wegen der Eindeutigkeit der MOW's im Spektralsatz ist Z das zu X gehörige MOW. (5.36) folgt aus (5.35). \square

Der Spektralsatz ist besonders nützlich als technisches Hilfsmittel, z.B. zum Nachweis der Existenz von Lösungen der AR-, MA- oder ARMA-Gleichungen, wie wir in Kapitel 7 sehen werden, und zur Herleitung von Spektraldarstellungen von Kovarianzfunktionen. Letztere werden in der Praxis zur Interpretation vorliegender Zeitreihen bevorzugt herangezogen.

5.4 Eine Substitutionsregel für stochastische Integrale

Abschließend kehren wir noch einmal zu allgemeineren Überlegungen im Zusammenhang mit MOW's zurück und leiten eine Art Ketten- oder Substitutionsregel her. Insbesondere bei der Untersuchung von ARMA-Modellen und bei der Betrachtung von sogenannten Filtern für Zeitreihen werden wir diese Substitutionsregel vorteilhaft einsetzen.

Lemma 5.10

Sei Z ein MOW über $(\mathcal{X}, \mathcal{B})$ mit Werten in einem Hilbertraum H mit innerem Produkt $\langle \cdot, \cdot \rangle$. μ_Z gemäß (5.5) sei das zugehörige Maß.

Für $g \in L_2^{\mathbb{C}}(\mu_Z)$ definiere man über $(\mathcal{X}, \mathcal{B})$ eine H -wertige Mengenfunktion Z_1 und ein endliches Maß μ_1 durch

$$\begin{aligned} Z_1(B) &:= \int_B g dZ := \int g \cdot 1_B dZ \\ \mu_1(B) &:= \int_B |g|^2 d\mu_Z, \quad \forall B \in \mathcal{B}. \end{aligned} \tag{5.38}$$

Dann gilt:

- a) Z_1 ist ein MOW mit $\mu_{Z_1} = \mu_1$.
- b) Für $f \in L_2^{\mathbb{C}}(\mu_1)$ gilt $f \cdot g \in L_2^{\mathbb{C}}(\mu_Z)$ und

$$\int f dZ_1 = \int fg dZ \quad (\text{Substitutionsregel für MOW's}). \tag{5.39}$$

- c) Falls $|g| > 0$ μ_Z -f.s. gilt, so folgt

$$Z(B) = \int_B g^{-1} dZ_1, \quad \forall B \in \mathcal{B}. \tag{5.40}$$

Beweis: a) Für $B_1, B_2 \in \mathcal{B}$ gilt

$$\begin{aligned}\langle Z_1(B_1), Z_1(B_2) \rangle &= \left\langle \int g 1_{B_1} dZ, \int g 1_{B_2} dZ \right\rangle \\ &= \int |g|^2 1_{B_1 \cap B_2} d\mu_Z = \mu_1(B_1 \cap B_2).\end{aligned}\quad (5.41)$$

b) Für primitives f , also für eine messbare Funktion, die nur endlich viele Werte annimmt, gilt die Aussage wegen der Definition von Z_1 . Für beliebiges $f \in L_2^{\mathbb{C}}(\mu_1)$ sei (f_n) eine Folge primitiver Funktionen mit der Eigenschaft $\int |f_n - f|^2 d\mu_1 \rightarrow 0$ für $n \rightarrow \infty$. Dann gilt wegen (5.20) auch $\int f_n dZ_1 \rightarrow \int f dZ_1$ in H . Mit (5.22) erhalten wir weiter

$$\begin{aligned}\left\| \int f_n dZ_1 - \int f \cdot g dZ \right\|^2 &= \left\| \int f_n \cdot g dZ - \int f \cdot g dZ \right\|^2 \\ &= \int |f_n - f|^2 |g|^2 d\mu = \int |f_n - f|^2 d\mu_1 \rightarrow 0.\end{aligned}\quad (5.42)$$

Hieraus folgt (5.39).

c) Man setze $g^{-1} := 0$ auf $\{g = 0\}$. Wegen $|g| > 0$ μ_Z -f.s. folgt die Gleichung $g^{-1} \cdot g = 1$ μ_Z -f.s. Mit

$$\int |g^{-1} \cdot 1_B|^2 d\mu_1 = \int |g|^{-2} \cdot 1_B \cdot |g|^2 d\mu_Z = \mu_Z(B) < \infty$$

folgt $g^{-1} \cdot 1_B \in L_2^{\mathbb{C}}(\mu_1)$. Aufgrund von Teil b) ergibt sich

$$\int g^{-1} \cdot 1_B dZ_1 = \int g^{-1} \cdot 1_B \cdot g dZ = Z(B), \quad \forall B \in \mathcal{B}.$$

□

Aufgaben

Aufgabe 5.1 Man zeige, dass die in (5.2) definierten komplexen Zufallsvariablen Z_j zentriert und unkorreliert sind.

Aufgabe 5.2 μ sei ein endliches Maß auf $(\mathcal{X}, \mathcal{B}) = ((-\pi, \pi], (-\pi, \pi] \cap \mathbb{B})$. Man zeige direkt, dass die stetigen und beschränkten Funktionen $f : \mathcal{X} \rightarrow \mathbb{C}$ dicht in $L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu)$ liegen.

Hinweis: Die primitiven Funktionen $\sum_{n=1}^N c_n 1_{B_n}$, $c_n \in \mathbb{C}$, $B_n \in \mathcal{B}$, liegen bekanntlich dicht in $L_2^{\mathbb{C}}(\mu)$. Es reicht also $f = 1_B$, $B \in \mathcal{B}$, zu approximieren. Nach dem üblichen Beweis des Fortsetzungssatzes von Maßen (vgl. Behnen/Neuhaus (2003), Satz 39.1) gibt es aus einer erzeugenden Algebra \mathcal{B}_0 von \mathcal{B} Mengen B_0 , so dass $\int |1_B - 1_{B_0}| d\mu$ beliebig klein wird. Man kann demnach $B \in \mathcal{B}_0$ voraussetzen. Wählt man als \mathcal{B}_0 die Familie der Summen endlich vieler

Intervalle $(a, b]$, so reicht es sogar $f = 1_{(a, b]}$, $-\pi \leq a < b \leq \pi$ anzunehmen. Es ist nun leicht zu sehen, dass derartige Indikatorfunktionen in $L_2^{\mathbb{C}}(\mu)$ durch z.B. stückweise lineare stetige Funktionen beliebig gut approximiert werden können.

Aufgabe 5.3 Sei $(\mathcal{X}, \mathcal{B}, \mu)$ ein endlicher Maßraum, H ein Hilbertraum und $Z : \mathcal{B} \rightarrow H$ ein MOW. Man zeige:

- a) $\|Z(B_1) - Z(B_2)\|^2 = \mu(B_1 \triangle B_2)$ mit $B_1 \triangle B_2 = B_1 \setminus B_2 + B_2 \setminus B_1$
- b) Aus $B_2 \subset B_1$ folgt $\|Z(B_2)\| \leq \|Z(B_1)\|$
- c) $Z(B_1 \cup B_2) = Z(B_1) + Z(B_2) - Z(B_1 \cap B_2)$
- d) $Z(B_1 \setminus B_2) = Z(B_1) - Z(B_1 \cap B_2)$.

Aufgabe 5.4 Man betrachte für eine zentrierte, quadratintegrierbare Zufallsvariable e die Zeitreihe $X_t = (-1)^t e$, $t \in \mathbb{Z}$, und zeige, dass sie schwach stationär ist. Man berechne das zugehörige Spektralmaß und das MOW.

Aufgabe 5.5 Seien $(\mathcal{X}, \mathcal{B})$ und $(\mathcal{Y}, \mathcal{C})$ zwei Messräume, $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und $Z : \mathcal{B} \rightarrow H$ ein MOW mit zugehörigem Spektralmaß μ . Ist $T : \mathcal{X} \rightarrow \mathcal{Y}$ eine $\mathcal{B} - \mathcal{C}$ -messbare Abbildung, so sei die Abbildung $Z^T : \mathcal{C} \rightarrow H$ durch $Z^T(C) := Z(T^{-1}(C))$, $C \in \mathcal{C}$, definiert. Man zeige

- a) Z^T ist ein MOW (das *induzierte MOW*) mit zugehörigem Maß μ^T .
- b) Ist $g \in L_2^{\mathbb{C}}(\mathcal{Y}, \mathcal{C}, \mu^T)$, so gilt die Transformationsformel

$$\int g dZ^T = \int g \circ T dZ. \quad (5.43)$$

Aufgabe 5.6 Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, zentrierte, stationäre Zeitreihe mit Spektralmaß μ und MOW Z . Weiter sei $g \in L_2^{\mathbb{R}}(\mu)$. Ist $T(\omega) = -\omega$ für $\omega \in (-\pi, \pi)$ und $T(\pi) = \pi$ die Spiegelungsabbildung, so zeige man

- a) Falls $g \circ T = g$ μ -fast sicher gilt, so ist $\int g dZ$ reell.
- b) Falls $g \circ T = -g$ μ -fast sicher gilt, so ist $\int g dZ$ rein imaginär.

Aufgabe 5.7 Sei $\omega_0 \in (0, \pi)$.

Man zeige für eine quadratintegrierbare, zentrierte Zufallsvariable $Z \neq 0$, dass die Zeitreihe $X = (X_t := Z e^{it\omega_0} : t \in \mathbb{Z})$ ein um 0 unsymmetrisches Spektralmaß besitzt.

Filterung stationärer Zeitreihen

In diesem Kapitel werden wir Zeitreihen mit Hilfe von sogenannten Filtern transformieren und die Eigenschaften der resultierenden Zeitreihen studieren. Filter werden in der Praxis hauptsächlich verwendet, um bestimmte Frequenzen zu unterdrücken oder zu verstärken. Beispiele sind sogenannte *low pass*- und *high pass*-Filter, die entweder im Wesentlichen nur niederfrequente oder nur hochfrequente Anteile der Zeitreihe passieren lassen. Darüberhinaus hat die hier vorgestellte Filtertheorie aber einen entscheidenden Nutzen in der Darstellung von Lösungen gewisser Differenzengleichungen (ARMA-Gleichungen), in Kapitel 7. Der Einsatz der hier entwickelten Methodik wird es uns erlauben, die für die Anwendungen so wichtige Klasse der ARMA-Zeitreihen äußerst effizient und vollständig zu untersuchen.

6.1 Grundbegriffe und einfache Eigenschaften von Filtern

Zunächst wollen wir den Begriff des Filters bei Zeitreihen definieren.

Definition 6.1 (*Filter*)

Sei $Y = (Y_t : t \in \mathbb{Z})$ eine komplexe, stationäre und zentrierte Zeitreihe mit MOW Z_Y und Spektralmaß μ_Y . Mit Hilfe einer Transferfunktion $\varphi \in L_2^{\mathbb{C}}(\mu_Y)$ definieren wir die gefilterte Zeitreihe $X = (X_t : t \in \mathbb{Z})$ gemäß

$$X_t = \int_{(-\pi, \pi]} \varphi(\omega) e^{it\omega} Z_Y(d\omega), \quad \forall t \in \mathbb{Z}. \quad (6.1)$$

Statt von der Transferfunktion sprechen wir auch von einem Filter φ , während $|\varphi|^2$ die zugehörige Filterfunktion genannt wird.

Jede zyklische „Input“-Komponente $e^{it\omega} Z_Y(d\omega)$ wird also mit dem Faktor $\varphi(\omega) = |\varphi(\omega)| e^{i\vartheta(\omega)}$ multipliziert, wobei die Amplitude mit $|\varphi(\omega)|$ multipliziert und die Phase um $\vartheta(\omega)$ verschoben wird. Insbesondere werden Input-Komponenten je nach dem Wert von $|\varphi(\omega)|$ verstärkt oder geschwächt, also

„gefiltert“.

Die grundlegenden Eigenschaften gefilterter Zeitreihen können wir mit Hilfe des Spektralsatzes recht einfach ableiten.

Satz 6.2 (*Kovarianzen und Spektraldichten bei Filterung*)

Die gemäß (6.1) definierte Zeitreihe $X = (X_t : t \in \mathbb{Z})$ ist eine komplexe, stationäre und zentrierte Zeitreihe mit Kovarianzfunktion

$$\gamma_X(h) = \int_{(-\pi, \pi]} e^{ih\omega} |\varphi(\omega)|^2 \mu_Y(d\omega), \quad \forall h \in \mathbb{Z}, \quad (6.2)$$

mit zugehörigem MOW

$$Z_X(B) = \int_B \varphi dZ_Y, \quad \forall B \in (-\pi, \pi] \cap \mathbb{B}. \quad (6.3)$$

Sie besitzt das Spektralmaß

$$\mu_X(B) = \int_B |\varphi|^2 d\mu_Y, \quad \forall B \in (-\pi, \pi] \cap \mathbb{B}, \quad (6.4)$$

und, falls Y eine Spektraldichte f_Y besitzt, die Spektraldichte

$$f_X = |\varphi|^2 f_Y. \quad (6.5)$$

Die Filterfunktion $|\varphi|^2$ ist also für die Änderung des Spektralmaßes verantwortlich.

Beweis: Wegen $e^{it\cdot}\varphi \in L_2^{\mathbb{C}}(\mu_Y)$ ergibt sich die Existenz und Zentriertheit von X_t aus der Definition des stochastischen Integrals. Die Stationarität von X folgt aus der Rechenregel (5.21); diese ergibt sogar (6.2), denn es gilt

$$\begin{aligned} \gamma_X(h) &= \langle X_{t+h}, X_t \rangle = \left\langle \int e^{i(t+h)\cdot} \varphi dZ_Y, \int e^{it\cdot} \varphi dZ_Y \right\rangle \\ &= \int_{(-\pi, \pi]} e^{ih\cdot} |\varphi|^2 d\mu_Y. \end{aligned}$$

Nach Lemma 5.10 ist $Z(B) := \int_B \varphi dZ_Y$ ein MOW mit

$$\int e^{it\cdot} dZ = \int e^{it\cdot} \varphi dZ_Y = X_t,$$

woraus wegen der Eindeutigkeit des MOW in der Spektraldarstellung die Gleichheit $Z_X = Z$, also (6.3), folgt. (6.4) ergibt sich direkt aus (6.3) mit (5.20). Schließlich impliziert (6.4) die Gleichung (6.5). \square

Für die Anwendungen spielen sogenannte lineare Filter eine übergeordnete Rolle, da sie sich rein im Zeitbereich, d.h. ohne die Verwendung des Spektralsatzes, schreiben lassen, wie der folgende Satz zeigt.

Satz 6.3 (*Filter in Reihendarstellung: Lineare Filter*)

Falls Konstanten $c_j \in \mathbb{C}, j \in \mathbb{Z}$, existieren derart, dass für $\varphi \in L_2^{\mathbb{C}}(\mu_Y)$ die Reihenentwicklung

$$\varphi(\omega) = \sum_{j=-\infty}^{+\infty} c_j e^{-ij\omega}, \quad (6.6)$$

in $L_2^{\mathbb{C}}(\mu_Y)$ gilt, so folgt die Darstellung

$$X_t = \sum_{j=-\infty}^{+\infty} c_j Y_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (6.7)$$

wobei die Konvergenz in $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ zu verstehen ist. Filter in Reihendarstellung werden oft auch als lineare Filter bezeichnet.

Beweis: Unter Beachtung der Rechenregeln (5.19) und (5.20) ergibt sich

$$\begin{aligned} X_t &= \int \varphi e^{it \cdot} dZ_Y = \int \sum_{j=-\infty}^{\infty} c_j e^{-ij \cdot} e^{it \cdot} dZ_Y = \sum_{j=-\infty}^{\infty} c_j \int e^{i(t-j) \cdot} dZ_Y \\ &= \sum_{j=-\infty}^{\infty} c_j Y_{t-j}. \end{aligned}$$

□

Bemerkung 6.4 (Zur Reihendarstellung)

a) Falls eine gegebene Folge $c_j, j \in \mathbb{Z}$, absolut summierbar ist, d.h. wenn $\sum_j |c_j| < \infty$ gilt, so konvergiert die Reihe $\sum_j c_j e^{-ij\omega}$ gleichmäßig in $\omega \in [-\pi, \pi]$ und definiert eine stetige Funktion φ auf $[-\pi, \pi]$. Wegen

$$\int_{(-\pi, \pi]} \left| \sum_{|j| \geq k} c_j e^{-ij \cdot} \right|^2 d\mu_Y \leq \left(\sum_{|j| \geq k} |c_j| \right)^2 \mu_Y(-\pi, \pi]$$

gilt die Konvergenz auch in $L_2^{\mathbb{C}}(\mu_Y)$, sodass (6.6) erfüllt ist und damit die Reihenentwicklung (6.7) gilt.

b) Ist φ eine stetig differenzierbare, reelle Funktion auf $[-\pi, \pi]$ mit $\varphi(-\pi) = \varphi(\pi)$, so lässt sie sich in eine Fourierreihe $\frac{1}{2\pi} \sum_{j=-\infty}^{+\infty} c_j e^{-ij\omega} = \varphi(\omega)$ entwickeln mit absolut summierbaren Fourierkoeffizienten $c_j = \int_{-\pi}^{\pi} \varphi(\omega) e^{ij\omega} d\omega$, vgl. Heuser (1981), Satz 136.5, und Aufgabe 6.1.

c) Ist $Y =: e$ ein weißes Rauschen, gilt also $\mu_e = \frac{\sigma_e^2}{2\pi} \lambda|_{(-\pi, \pi]}$, so konvergiert für

jedes $\varphi \in L_2^{\mathbb{C}}(\mu_e)$ die Fourierreihe aus b) in $L_2^{\mathbb{C}}(\mu_e)$ gegen φ . Der Grund dafür ist, dass in diesem Spezialfall die Funktionen $e^{ij\cdot}$, $j \in \mathbb{Z}$, ein vollständiges Orthonormalsystem in $L_2^{\mathbb{C}}(\mu_e)$ bilden, nach dem φ entwickelt werden kann, vgl. Heuser (1981), Satz 141.1. \square

In den bisherigen Überlegungen durfte die Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ komplexwertig sein. In den allermeisten Anwendungen ist Y allerdings reellwertig. Wir fragen uns deshalb, wann ein Filter φ eine reelle Reihe wieder in eine reelle Reihe transformiert. Nach Bemerkung 3.7 ist das Spektralmaß μ_Y einer reellen Zeitreihe Y auf $(-\pi, \pi)$ symmetrisch um 0, erfüllt also $\mu_Y(B) = \mu_Y(-B)$, $\forall B \in (-\pi, \pi) \cap \mathbb{B}$ mit $-B := \{-b : b \in B\}$. Unter Verwendung dieser Tatsache kann man das folgende Resultat zeigen.

Satz 6.5 (*Filterung reeller Zeitreihen*)

Ist $Y = (Y_t : t \in \mathbb{Z})$ reell mit Spektralmaß μ_Y und $\varphi \in L_2^{\mathbb{C}}$, so ist die gefilterte Reihe $X_t = \int_{(-\pi, \pi]} e^{it\cdot} \varphi dZ_Y$, $t \in \mathbb{Z}$, genau dann reell, wenn μ_Y -f.s. auf $(-\pi, \pi)$ gilt

$$\varphi(\omega) = \overline{\varphi(-\omega)}, \quad (6.8)$$

und im Fall $\mu\{\pi\} > 0$ der Wert $\varphi(\pi)$ reell ist.

Beweis: Wir definieren wie früher die Spiegelungsabbildung $T : (-\pi, \pi] \rightarrow (-\pi, \pi]$ gemäß $T(\omega) = -\omega \forall \omega \in (-\pi, \pi)$ und $T(\pi) = \pi$, wobei dann offenbar $T = T^{-1}$ gilt. Mit dieser Bezeichnung ist die Äquivalenz von $\varphi = \overline{\varphi} \circ T[\mu_Y]$ und $X_t = \overline{X}_t$, $\forall t \in \mathbb{Z}$, zu zeigen. Nach Bemerkung 3.7 wissen wir, dass $\mu_Y = \mu_Y^T$ gilt.

Man kann nun nacheinander zeigen

- (1) Durch $\overline{Z}_Y(B) := \overline{Z_Y(B)}$ wird ein MOW definiert mit $\mu_Y(B) = \|\overline{Z}_Y(B)\|^2$.
- (2) Es gilt $\int \varphi dZ_Y = \int \overline{\varphi} d\overline{Z}_Y$.
- (3) Durch $\widehat{Z}_Y(B) := \overline{Z}_Y(T(B))$ wird ein MOW definiert mit $\mu_Y(B) = \|\widehat{Z}_Y(B)\|^2$.
- (4) Es gilt $\int \varphi \circ T d\overline{Z}_Y = \int \varphi d\widehat{Z}_Y$.
- (5) Es gilt $Z_Y = \widehat{Z}_Y$.
- (6) Es gilt $\overline{X}_t = \int e^{it\cdot} \overline{\varphi} \circ T dZ_Y$, $\forall t \in \mathbb{Z}$.

Aus (6) ergibt sich unmittelbar, dass die gefilterte Zeitreihe $X = (X_t : t \in \mathbb{Z})$ genau dann reell ist, wenn gilt

$$\int e^{it\cdot} \overline{\varphi} \circ T dZ_Y = \int e^{it\cdot} \varphi dZ_Y, \quad \forall t \in \mathbb{Z},$$

was wegen (5.23) äquivalent zu $\overline{\varphi} \circ T = \varphi[\mu_Y]$ ist.

Es bleibt also, die Aussagen (1)–(6) zu zeigen. Dies ist als Aufgabe 6.2 gestellt. \square

Liegt φ in der Reihendarstellung (6.6) mit reellen Koeffizienten $c_j, j \in \mathbb{Z}$, vor, so ist (6.8) offensichtlich erfüllt. In diesem Fall sieht man auch direkt aus der Reihendarstellung (6.7), dass eine reelle Zeitreihe Y in eine reelle Zeitreihe X transformiert wird.

Aus (6.8) folgt, dass die Filterfunktion $|\varphi|^2$, die gemäß (6.4) das Spektralmaß μ_X bestimmt, symmetrisch zum Nullpunkt ist, also $|\varphi(\omega)|^2 = |\varphi(-\omega)|^2$ μ_Y -f.s. auf $(-\pi, \pi)$ gilt.

Zum Abschluß dieses Abschnitts wollen wir die Hintereinanderausführung von Filtern untersuchen

Lemma 6.6 (*Hintereinanderschalten von Filtern*)

Wird die gemäß (6.1) erzeugte Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit einem Filter $\psi \in L_2^{\mathbb{C}}(\mu_X)$ transformiert gemäß

$$Z_t = \int_{(-\pi, \pi]} \psi(\omega) e^{it\omega} dZ_X(\omega), \quad \forall t \in \mathbb{Z},$$

so gilt

$$Z_t = \int_{(-\pi, \pi]} \psi(\omega) \cdot \varphi(\omega) e^{it\omega} dZ_Y(\omega), \quad \forall t \in \mathbb{Z}, \quad (6.9)$$

und

$$\mu_Z(B) = \int_B |\psi|^2 |\varphi|^2 d\mu_Y, \quad \forall B \in (-\pi, \pi] \cap \mathbb{B}. \quad (6.10)$$

Der Beweis ergibt sich direkt aus der Substitutionsregel in Lemma 5.10.

6.2 Spezielle Filter

Dieser Abschnitt stellt eine Reihe von Filtern zusammen, die in den verschiedensten Anwendungen eine wichtige Rolle spielen. Häufig kommen die betrachteten Methodiken gar nicht direkt aus dem Bereich der Filter, sie lassen sich aber als solche darstellen. Die Darstellung als Filter erlaubt dann aber einen guten Einblick in die Wirkungsweise auf Zeitreihendaten. Zu den betrachteten Verfahren gehören die Methode der Differenzenbildung, des gleitenden Mittels und des exponentiellen Glättens, von denen wir die letzten beiden schon in Kapitel 1 kennen gelernt haben.

Beispiel 6.7 (Differenzenoperatoren)

Seien in Satz 6.3 $c_0 = 1, c_1 = -1$ und $c_j = 0$ sonst. Dann ist die gemäß (6.7) transformierte Reihe $X_t =: \Delta Y_t$ gleich

$$\Delta Y_t = Y_t - Y_{t-1}, \quad \forall t \in \mathbb{Z}. \quad (6.11)$$

$(\Delta Y_t : t \in \mathbb{Z})$ bezeichnet man als die erste Differenz der Zeitreihe $(Y_t : t \in \mathbb{Z})$. Die zugehörige Transferfunktion φ berechnet sich zu

$$\begin{aligned} \varphi(\omega) &= 1 - e^{-i\omega} = e^{-i\omega/2} (e^{i\omega/2} - e^{-i\omega/2}) \\ &= 2ie^{-i\omega/2} \sin(\omega/2) = 2e^{i(\pi-\omega)/2} \sin(\omega/2), \end{aligned} \quad (6.12)$$

woraus sich die Filterfunktion $|\varphi|^2$ zu

$$|\varphi(\omega)|^2 = 4 \sin^2(\omega/2) = 2(1 - \cos \omega) \quad (6.13)$$

ergibt.

Wiederholte Differenzenbildung führt auf den Differenzenoperator n -ter Ordnung. Es gilt

$$\begin{aligned} \Delta^n(Y_t) &= \Delta^{n-1}(Y_t) - \Delta^{n-1}(Y_{t-1}) \\ &= \sum_{r=0}^n (-1)^r \binom{n}{r} Y_{t-r}, \quad \forall t \in \mathbb{Z}, \forall n \geq 1, \end{aligned} \quad (6.14)$$

wobei $\Delta^n(Y_t)$ als n -te Differenz der Zeitreihe $(Y_t : t \in \mathbb{Z})$ bezeichnet wird.

Dieser Filter hat wegen Lemma 6.6 die Transferfunktion

$$\varphi^n(\omega) = 2^n e^{in(\pi-\omega)/2} \sin^n(\omega/2)$$

mit der Filterfunktion

$$|\varphi(\omega)|^{2n} = 2^n (1 - \cos \omega)^n. \quad (6.15)$$

In Abbildung 6.1 sind die Filterfunktionen die Differenz erster Ordnung und die d -Differenz ($d = 5$) dargestellt. Differenzenoperatoren Δ^n sind „high pass“-Filter. Sie lassen vornehmlich hohe Frequenzen passieren und unterdrücken niedrige Frequenzen. Der Differenzenoperator Δ wird in der Praxis hauptsächlich angewandt, um Trends aus einer Reihe zu beseitigen, also eigentlich auf eine nichtstationäre Reihe. Jeder Trend kann jedoch als Beginn einer langwelligen Schwingung aufgefasst werden, die dann durch den „high pass“-Filter unterdrückt wird. Entsprechend werden Differenzenoperatoren n -ter Ordnung in der Praxis verwendet, um polynomiale Trends aus einer Reihe zu entfernen. Es gilt nämlich (siehe Aufgabe 6.3)

$$\begin{aligned} \text{Ist } P(t) \text{ ein Polynom } n\text{-ten Grades, so ist } \Delta^n P(t) \\ \text{eine konstante Funktion, also } \Delta^{n+1} P(t) = 0, \end{aligned} \quad (6.16)$$

wobei P sogar stochastische Koeffizienten besitzen darf.

Die Wirkungsweise der Differenzenbildung wollen wir anhand der schon in Ka-

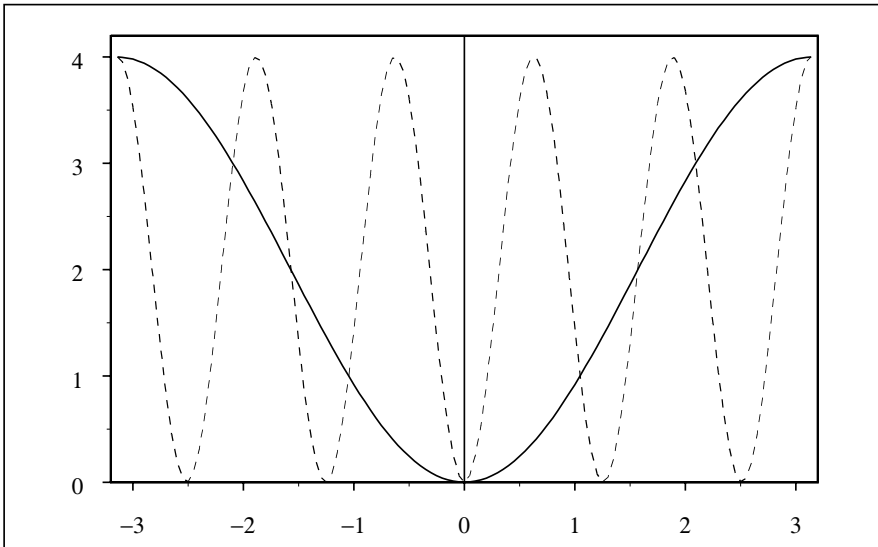


Abb. 6.1 Filterfunktion für die Differenz erster Ordnung (durchgezogen) und für die d -Differenz, $d = 5$ (gestrichelt)

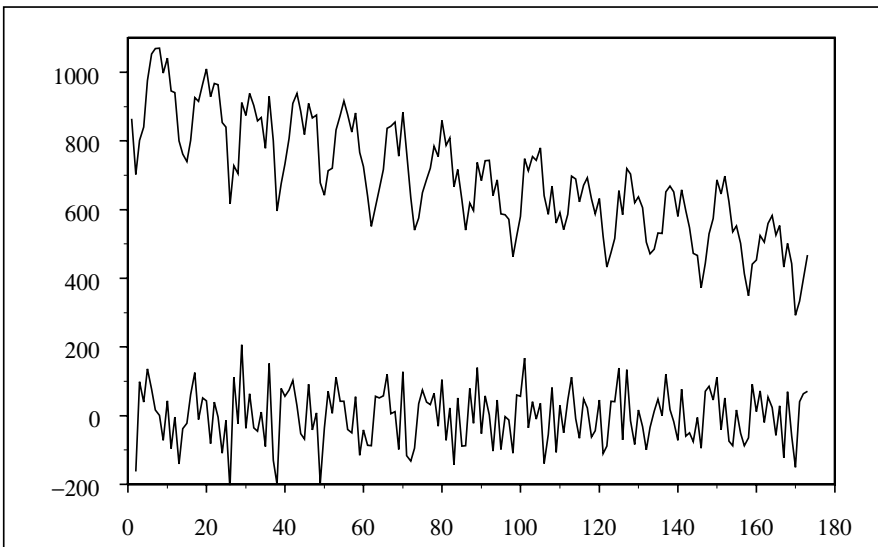


Abb. 6.2 Monatliche Anzahl der durch Verkehrsunfälle getöteten Personen und die durch Differenzenbildung gefilterte Zeitreihe (vgl. mit Abbildung 1.8)

pitel 1 dargestellten Anzahlen der monatlich durch Verkehrsunfälle getöteten Personen demonstrieren (vgl. Abbildung 1.8). In Abbildung 6.2 sind sowohl die sehr deutlich linear abfallenden Unfalltodesfälle wie die hieraus durch Differenzenbildung erhaltene gefilterte Zeitreihe zu sehen. Man erkennt, dass der linear abfallende Trend sehr gut eliminiert wurde. Demgegenüber wurde das saisonale Verhalten der Unfallzahlen über die jeweiligen Kalenderjahre nur unwesentlich verändert.

Anstelle der Differenzenbildung an direkt aufeinanderfolgenden Zeitpunkten, kann man auch für $d = 1, 2, 3, \dots$ die sogenannte d -Differenz

$$X_t = Y_t - Y_{t-d}, \quad \forall t \in \mathbb{Z}, \quad (6.17)$$

bilden mit Transferfunktion $\varphi(\omega) = 1 - e^{-i\omega d}$ und Filterfunktion

$$|\varphi(\omega)|^2 = 2(1 - \cos(\omega d)). \quad (6.18)$$

Diese Transfer- und Filterfunktion verschwinden genau für $\omega = k\omega_d \in (-\pi, \pi]$ mit $\omega_d := (2\pi)/d$, $k \in \mathbb{Z}$. Dies bedeutet, dass aus der Zeitreihe Y alle zyklischen Komponenten mit Kreisfrequenzen $k\omega_d$ durch die d -Differenz entfernt werden. Z.B. wird bei einer Zeitreihe, die eine periodische Komponente mit Periode $d = 12$ (Monate) besitzt, diese bei Anwendung der d -Differenz unterdrückt. Darüberhinaus werden sogar alle zyklischen Komponenten mit Kreisfrequenzen $(k\pi)/6 \in (-\pi, \pi]$, $k \in \mathbb{Z}$, unterdrückt. \square

In Abbildung 6.1 sind die Filterfunktionen für die Differenz erster Ordnung und für die d -Differenz mit $d = 5$ dargestellt.

In den beiden folgenden Beispielen betrachten wir „low pass“-Filter, die vornehmlich kleine Frequenzen passieren lassen und hohe Frequenzen unterdrücken. Letzteres bewirkt eine Glättung der Reihe durch den Filter. Wir beginnen mit dem gleitenden Mittel, das wir schon in Beispiel 1.2 verwendet haben, um bei einer nichtstationären Zeitreihe eine Trendkomponente zu schätzen.

Beispiel 6.8 (Gleitendes Mittel als Filter)

Für $q \geq 1$ betrachten wir das folgende gleitende Mittel von $Y = (Y_t : t \in \mathbb{Z})$

$$X_t = \frac{1}{2q+1} \sum_{j=-q}^{+q} Y_{t-j}, \quad \forall t \in \mathbb{Z}. \quad (6.19)$$

Gemäß Satz 6.3 bedeutet dies, dass mit $c_j = 1/(2q+1)$ für $|j| \leq q$ und $c_j = 0$ für $|j| > q$ durch (6.19) ein Filter definiert wird mit der Transferfunktion

$$\begin{aligned} \varphi(\omega) &= \frac{1}{2q+1} \sum_{j=-q}^q e^{-ij\omega} \\ &= \begin{cases} 1 & \text{für } \omega = 0 \\ \frac{1}{2q+1} \frac{\sin(\omega(2q+1)/2)}{\sin(\omega/2)} & \text{für } \omega \neq 0, \end{cases} \end{aligned} \quad (6.20)$$

siehe (A.6) im Anhang A.1, und zugehöriger Filterfunktion

$$|\varphi(\omega)|^2 = \frac{2\pi}{2q+1} k_{2q+1}(\omega), \quad (6.21)$$

mit dem aus der Fourieranalyse bekannten Fejér-Kern

$$k_N(\omega) = \begin{cases} \frac{\sin^2(\omega N/2)}{2\pi N \sin^2(\omega/2)} & \text{für } \omega \neq 0 \\ \frac{N}{2\pi} & \text{für } \omega = 0, \end{cases} \quad (6.22)$$

$N \geq 1$, für den $\int_{-\pi}^{\pi} k_N(\omega) d\omega = 1$ gilt, wie aus der Orthogonalität der Funktionen $e^{ij\cdot}$ in $L_2^{\mathbb{C}}(-\pi, \pi]$ (Integration bzgl. des Lebesgue-Maßes) folgt.

Aus Abbildung 6.3 erkennt man, dass tatsächlich ein „low pass“-Filter vor-

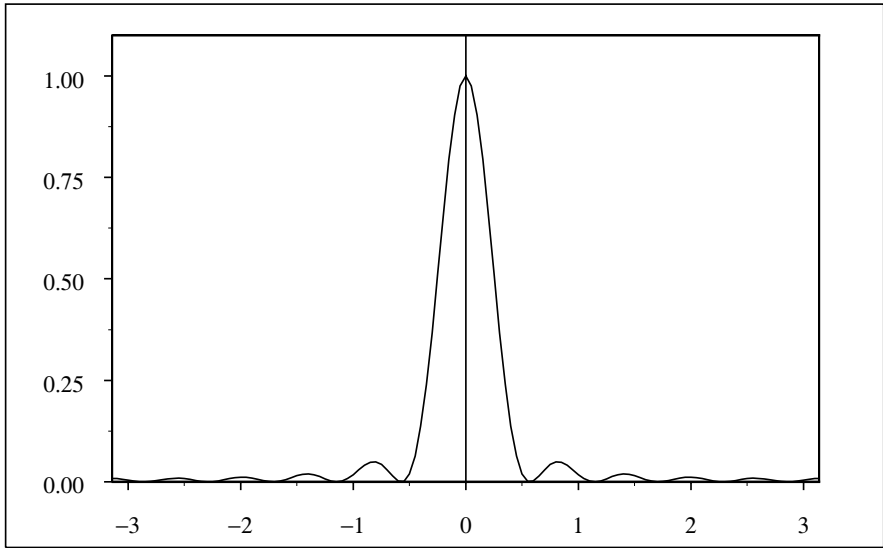


Abb. 6.3 Filterfunktion für das gleitende Mittel mit $q = 5$ (vgl. (6.21))

liegt. Besitzt $Y = (Y_t : t \in \mathbb{Z})$ z.B. eine beschränkte Spektraldichte f_Y auf $(-\pi, \pi]$ so folgt

$$\begin{aligned} \text{Var} X_t = \gamma_X(0) &= \frac{2\pi}{2q+1} \int_{(-\pi, \pi]} k_{2q+1} f_Y d\lambda \\ &\leq \frac{2\pi}{2q+1} \sup_{-\pi < \omega \leq \pi} f_Y(\omega). \end{aligned} \quad (6.23)$$

Durch gleitende Mittelbildung verkleinert sich also die Varianz der gefilterten Reihe umgekehrt proportional zur Zahl der gemittelten Werte. Das gleitende

Mittel haben wir in Kapitel 1 dazu verwendet, eine Trendkomponente einer nichtstationären Zeitreihe zu schätzen (vgl. Beispiel 1.2). \square

Auch die Methode des exponentiellen Glättens aus Beispiel 1.2, die erfolgreich zur Trendschätzung bei nichtstationären Zeitreihen eingesetzt wird, lässt sich als Filter auffassen.

Beispiel 6.9 (Exponentielles Glätten als Filter)

Für $0 < \alpha < 1$ ist das exponentielle Glätten definiert durch

$$X_t = \sum_{j=0}^{\infty} (1 - \alpha) \alpha^j Y_{t-j}, \quad \forall t \in \mathbb{Z}. \quad (6.24)$$

Gemäß Satz 6.3 bedeutet dies, dass mit $c_j = (1 - \alpha) \alpha^j$ für $j \geq 0$ und $c_j = 0$ für $j < 0$ durch (6.24) ein Filter φ definiert wird gemäß

$$\varphi(\omega) = \sum_{j=0}^{\infty} (1 - \alpha) (\alpha e^{-i\omega})^j = \frac{1 - \alpha}{1 - \alpha e^{-i\omega}} \quad (6.25)$$

und Filterfunktion

$$|\varphi(\omega)|^2 = \frac{(1 - \alpha)^2}{|1 - \alpha e^{-i\omega}|^2} = \frac{(1 - \alpha)^2}{1 - 2\alpha \cos \omega + \alpha^2}. \quad (6.26)$$

Abbildung 6.4 zeigt, dass es sich beim exponentiellen Glätten tatsächlich um einen „low pass“-Filter handelt. Die schon in (1.6) angegebene Rekursionsformel, die die Anwendung des exponentiellen Glättens algorithmisch besonders einfach macht, schreibt sich mit den jetzigen Bezeichnungen als

$$X_t = \alpha X_{t-1} + (1 - \alpha) Y_t, \quad \forall t \in \mathbb{Z}. \quad (6.27)$$

Sie folgt direkt aus (6.24). Um den bereits oben erwähnten Glättungseffekt des low pass-Filters (6.24) zu verdeutlichen verweisen wir zum einen auf Abbildung 1.6 und Beispiel 1.2 in Kapitel 1. Zum anderen kommen wir auf die in Kapitel 2 ausführlich besprochenen Sonnenfleckenzahlen zurück (vgl. Beispiel 2.14). In Abbildung 6.5 sind einerseits die um das arithmetische Mittel zentrierten Sonnenfleckenzahlen und andererseits die hieraus mittels exponentiellem Glätten mit $\alpha = 0.8$ gewonnenen gefilterten Werte zu sehen. Es zeigt sich, dass das exponentielle Glätten das auffällige periodische Verhalten der Sonnenfleckenzahlen (mit approximativer Wellenlänge von um die 11 Jahre (vgl. erneut mit Beispiel 2.14) kaum beeinflusst. Lediglich die Amplituden erscheinen nach der Glättung wesentlich kleiner. Betrachtet man aber die zugehörige Filterfunktion in Abbildung 6.4, die ja multiplikativ auf die Spektralverteilung wirkt, in der Nähe der zur Wellenlänge 11 gehörenden Frequenz von $\omega = 2\pi/11 \approx 0.57$ und beachtet man, dass die Spektralverteilung in der Nähe einer bestimmten Frequenz die Größenordnung der Varianz der

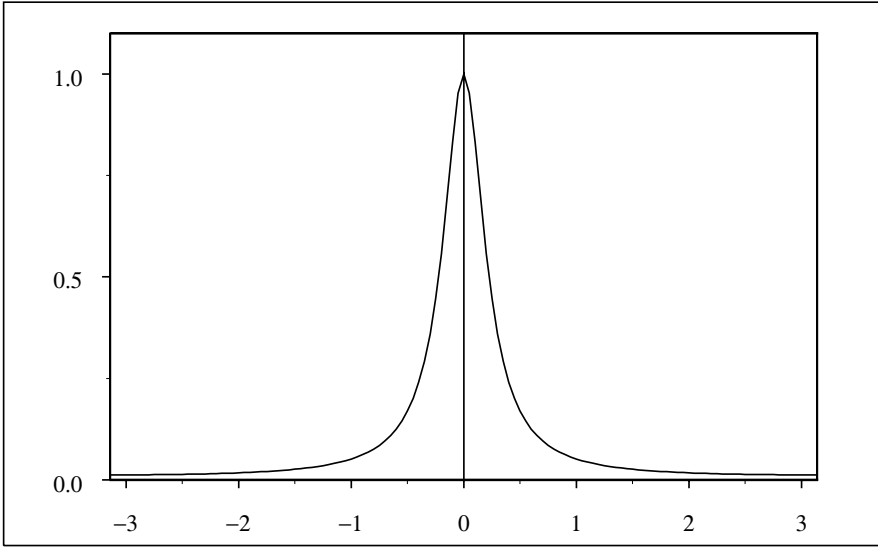


Abb. 6.4 Filterfunktion für das exponentielle Glätten mit $\alpha = 0.8$ (vgl. (6.26))

stochastischen Amplitude genau dieser Frequenz wiedergibt, so ist die beobachtete Amplitudenreduktion in Abbildung 6.5 durchaus im Einklang mit den Resultaten über die Eigenschaften von Filtern.

Die zusätzlich sichtbare leichte zeitliche Verschiebung der Maximalstellen nach der exponentiellen Glättung erklärt sich dadurch, dass das exponentielle Glätten in der Form (6.24) ein rein einseitiger Filter ist. Eine nachträgliche Zeitverschiebung

$$X'_t = X_{t+h}, \quad \forall t \in \mathbb{Z},$$

die dieses Phänomen rückgängig machen würde, ist ebenfalls ein linearer Filter und zwar mit Filterfunktion konstant 1, d.h. ohne Einfluss auf die Spektralverteilung. Eine Hintereinanderausführung des exponentiellen Glättens mit dieser Zeitverschiebung führt unter Berücksichtigung von (6.24) zu folgendem Gesamtfilter

$$X'_t = \sum_{j=0}^{\infty} (1 - \alpha) \alpha^j Y_{t+h-j}, \quad \forall t \in \mathbb{Z}.$$

Abhängig von α könnte man h etwa so wählen, dass die Summen der Gewichte für Y_s mit $s < t$ und für Y_s mit $s > t$ ungefähr gleich groß sind. Im vorliegenden Beispiel würde ein solches Vorgehen zu einem h von 2 oder 3 führen, was in Einklang mit Abbildung 6.5 ist. \square

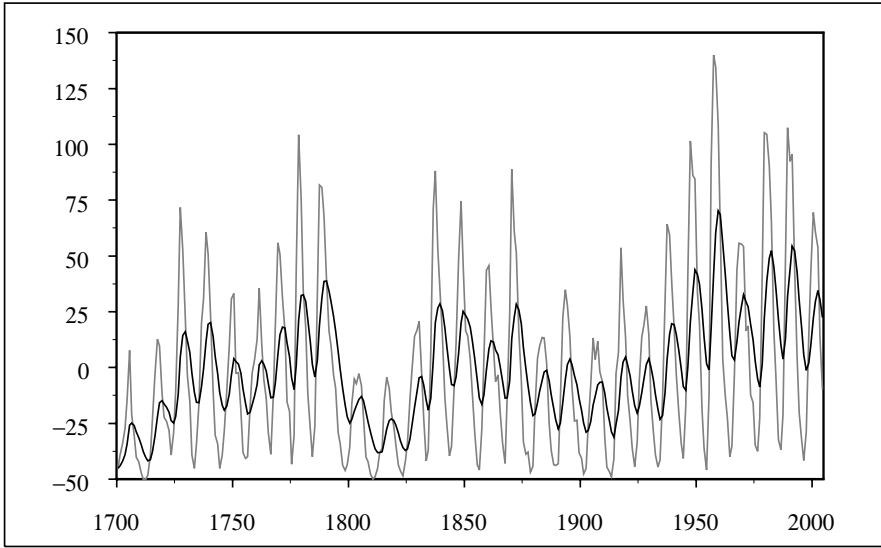


Abb. 6.5 Zentrierte Sonnenfleckenzahlen und gemäß exponentiellem Glätten ($\alpha = 0.8$) gefilterte Zeitreihe

Beispiel 6.10 (Unterdrückung einer festen Frequenz ω_0)

Wir wollen einen Filter konstruieren, der eine vorgegebene Frequenz ω_0 einer vorgegebenen Zeitreihe (Y_t) unterdrückt. Betrachten wir dazu einen Moving Average-Filter der Länge 2 gemäß

$$X_t = Y_t + b_1 Y_{t-1} + b_2 Y_{t-2}, \quad \forall t \in \mathbb{Z}. \quad (6.28)$$

Hierzu gehört die Filterfunktion

$$|\varphi(\omega)|^2 = |1 + b_1 e^{-i\omega} + b_2 e^{-2i\omega}|^2. \quad (6.29)$$

Mit $B(z) = 1 + b_1 z + b_2 z^2$ gilt $\varphi(\omega) = B(e^{-i\omega})$. Gesucht sind nun Koeffizienten b_1 und b_2 derart, dass $B(z)$ gerade an der Stelle $z_0 = e^{-i\omega_0}$ (und damit automatisch auch in $\bar{z}_0 = e^{i\omega_0}$) eine Nullstelle besitzt. Ein solches Polynom zweiten Grades $B(z)$ können wir leicht gemäß

$$\begin{aligned} B(z) &= (1 - ze^{-i\omega_0})(1 - ze^{i\omega_0}) \\ &= 1 - 2\cos(\omega_0)z + z^2 \end{aligned} \quad (6.30)$$

erzeugen. Die resultierende Filterfunktion lautet (vgl. (6.29))

$$|\varphi(\omega)|^2 = 2 + 4\cos^2(\omega_0) - 8\cos(\omega_0)\cos(\omega) + 2\cos(2\omega). \quad (6.31)$$

Für die konkrete Frequenz $\omega_0 = 0.55$ erhalten wir den Filter

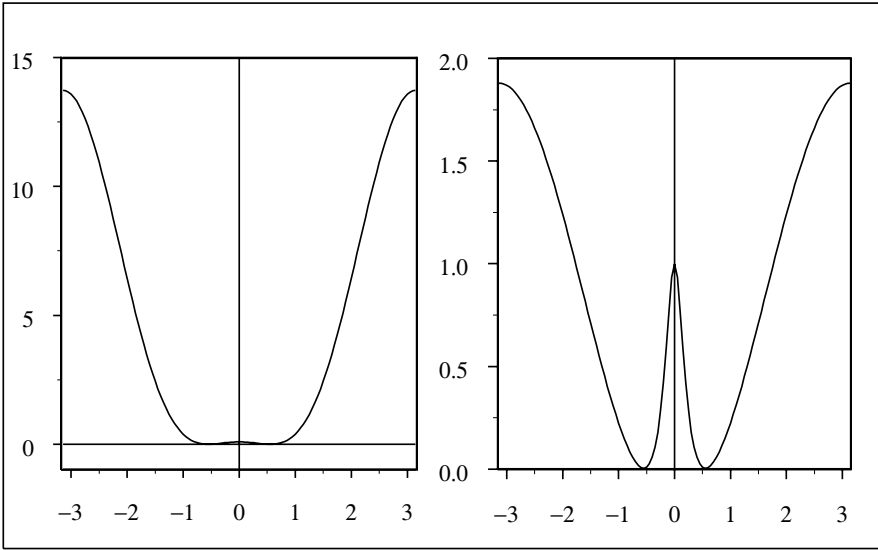


Abb. 6.6 Moving Average–Filterfunktion $|\varphi|^2$ gemäß (6.31) (links) und Filterfunktion der Hintereinanderschaltung des Moving Average–Filters mit einem exponentiellen Glätten ($\alpha = 0.8$) (rechts)

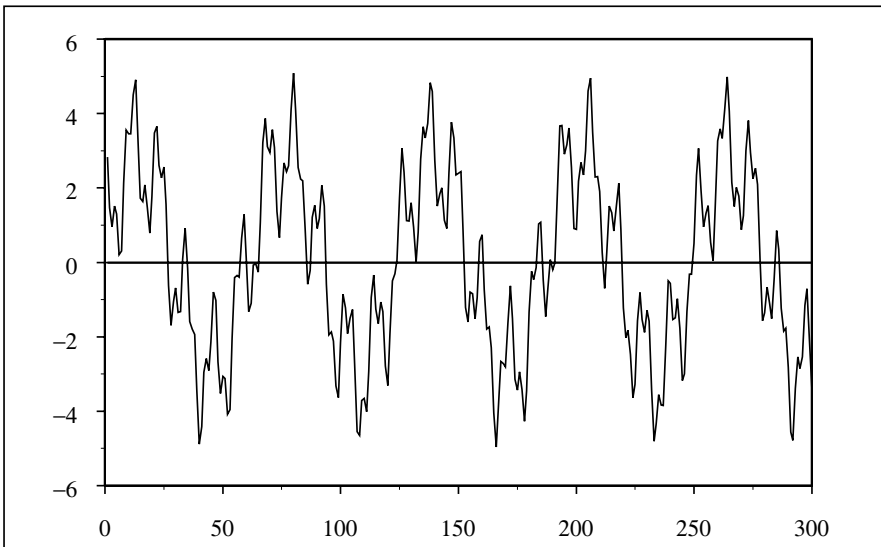


Abb. 6.7 Simulierte zyklische Zeitreihe der Länge $n = 300$ mit Frequenzen $\omega_1 = 0.10, \omega_2 = 0.55, \omega_3 = 1.50$ und Varianzen der Amplituden $\sigma_1^2 = 9, \sigma_2^2 = 4, \sigma_3^2 = 1$ (vgl. Definition 2.9)

$$X_t = Y_t - 1.705Y_{t-1} + Y_{t-2}, \quad \forall t \in \mathbb{Z}. \quad (6.32)$$

In Abbildung 6.6 ist die gemäß (6.30) gegebene Filterfunktion dargestellt. In der Tat hat sie die gewünschte Eigenschaft $\varphi(\omega_0) = 0$. Allerdings ist diese Filterfunktion für große Frequenzen recht groß, so dass wir mit einer deutlichen Verstärkung entsprechender Frequenzanteile in der gefilterten Zeitreihe rechnen müssen. Zudem werden Frequenzen ω , die kleiner als ω_0 sind, nur mit einem relativ kleinen Gewicht versehen. Wenden wir den konkreten Filter (6.32) auf eine simulierte zyklische Zeitreihe an, die sich aus der Überlagerung von drei Schwingungsanteilen mit den Frequenzen $\omega_1 = 0.10$, $\omega_2 = 0.55$ und $\omega_3 = 1.50$ ergibt (vgl. Definition 2.9). Die Varianzen für die zugehörigen Amplituden wurden wie folgt gewählt $\sigma_1^2 = 9$, $\sigma_2^2 = 4$ und $\sigma_3^2 = 1$. In Abbildung 6.7 ist eine Simulation der Länge $n = 300$ einer solchen zyklischen Zeitreihe zu sehen.

Abbildung 6.8 gibt die mit dem Filter (6.32) gefilterte zyklische Zeitreihe an. Die gewünschte Unterdrückung der Frequenz $\omega_0 = 0.55$, aber auch die oben angesprochene Verstärkung hoher und die gleichzeitige Reduzierung kleiner Frequenzen ist recht deutlich zu erkennen. Um die beiden unerwünschten

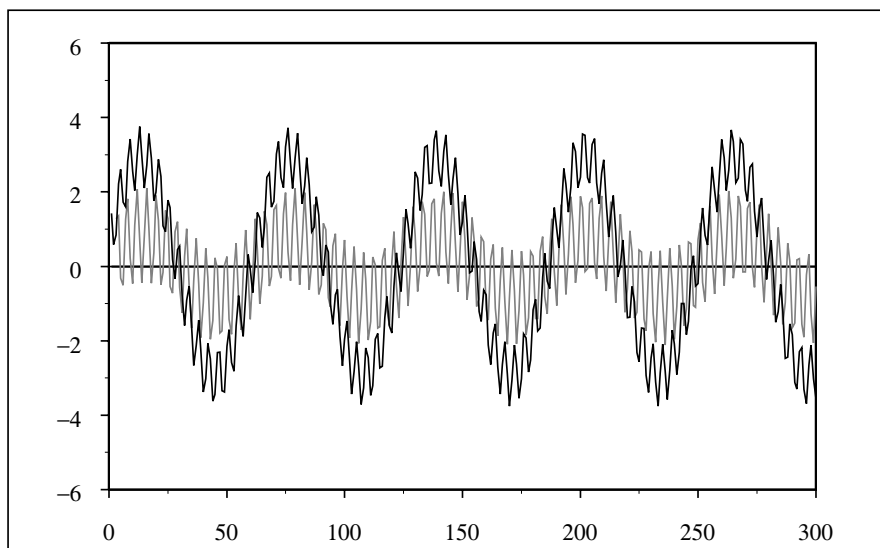


Abb. 6.8 Gemäß (6.32) gefilterte zyklische Zeitreihe aus Abbildung 6.7 (kleiner Amplitude) und Komponente der zyklischen Zeitreihe nach Weglassen des Frequenzanteil ω_2

Effekte zu unterbinden schalten wir hinter den Filter aus (6.32) noch ein exponentielles Glätten mit Parameter $\alpha = 0.8$ (vgl. dazu die Darstellung der

Filterfunktion dieses exponentiellen Glättens in Abbildung 6.4) und eine Multiplikation der gefilterten Zeitreihe mit dem Faktor 3.33 (entsprechend Multiplikation der Filterfunktion mit 3.33^2). Die Multiplikation stellt sicher, dass die sich durch Produktbildung ergebende Gesamtfilterfunktion an der Stelle $\omega = 0$ den Wert 1 annimmt. Eine graphische Darstellung der Gesamtfilterfunktion ist ebenfalls in Abbildung 6.6 angegeben. Diese Gesamtfilterfunktion hat nun genau den gewünschten Effekt der Auslöschung einer bestimmten Frequenz $\omega_0 = 0.55$ und führt nicht zu der unerwünschten Verstärkung der hohen Frequenzen. Abbildung 6.9 gibt die mit dem zweistufigen Filter gefilterte Reihe und zum Vergleich die Komponente der zyklischen Zeitreihe aus Abbildung 6.7 an, die durch Weglassen des Schwungsanteils mit Frequenz $\omega_2 = 0.55$ entsteht. Man erkennt sehr schön die gute Approximationseigenschaft der aus der ursprünglichen zyklischen Zeitreihe durch Filterung entstandenen Reihe. Die Zeitverschiebung resultiert erneut aus der Einseitigkeit der verwendeten Filter und könnte entsprechend der Bemerkung in Beispiel 6.9 kompensiert werden. \square

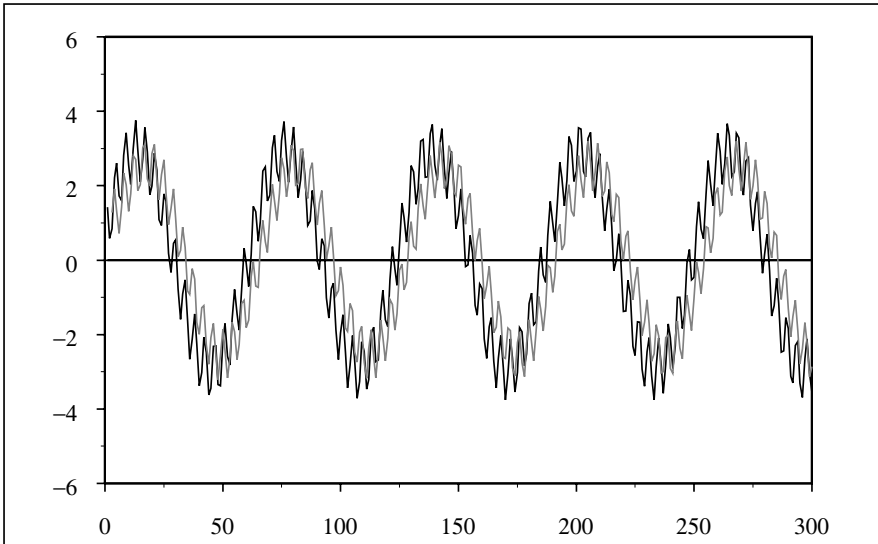


Abb. 6.9 Mit Gesamtfilter gefilterte zyklische Zeitreihe aus Abbildung 6.7 (fett) und zum Vergleich Komponente der zyklischen Zeitreihe nach Weglassen des Frequenzanteil ω_2 .

6.3 Zweiseitige MA-Reihen

In diesem Abschnitt wollen wir die schon in der Einleitung des Kapitels erwähnte Eigenschaft von gefilterten Zeitreihen als Lösungen von Differenzengleichungen vorbereiten. Das folgende Beispiel ist für diese Fragestellung, die in Kapitel 7 ausführlich behandelt wird, grundlegend. Hierin charakterisieren wir die Zeitreihen, die aus der Filterung eines weißen Rauschens entstehen.

Beispiel 6.11 (Zweiseitige MA-Reihen)

Ist $e = (e_t : t \in \mathbb{Z})$ ein weißes Rauschen mit $e_t \sim (0, \sigma_e^2)$, so ist $\mu_e = (\sigma_e^2/(2\pi)) \lambda|_{(-\pi, \pi]}$ das zugehörige Spektralmaß. Dann lässt sich nach Bemerkung 6.4 c) jeder Filter $\varphi \in L_2^{\mathbb{C}}(\mu_e)$ in eine in $L_2^{\mathbb{C}}(\mu_e)$ konvergente Reihe

$$\varphi = \sum_{j=-\infty}^{+\infty} c_j e^{-ij \cdot} \quad (6.33)$$

entwickeln mit Fourier-Koeffizienten $c_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi(\omega) e^{ij\omega} d\omega$, welche quadratsummierbar sind, also $\sum_{j=-\infty}^{+\infty} |c_j|^2 < \infty$ erfüllen. Umgekehrt definiert eine beliebige quadratsummierbare Folge $c_j, j \in \mathbb{Z}$, einen Filter φ gemäß (6.33).

Wir folgern also:

Jede Reihe $X = (X_t : t \in \mathbb{Z})$, die durch Filterung eines weißen Rauschens $e = (e_t : t \in \mathbb{Z})$ entsteht, hat eine in $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ konvergierende Reihendarstellung

$$X_t = \sum_{j=-\infty}^{+\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (6.34)$$

mit quadratsummierbaren, komplexen Konstanten $c_j, j \in \mathbb{Z}$. Umgekehrt gehört zu jeder Reihendarstellung (6.34) mit quadratsummierbaren, komplexen Konstanten $c_j, j \in \mathbb{Z}$, ein Filter $\varphi \in L_2^{\mathbb{C}}(\mu_e)$ gemäß (6.33) derart, dass $X_t = \int_{(-\pi, \pi]} e^{it \cdot} \varphi dZ_e$, $t \in \mathbb{Z}$, gilt. Wir nennen Reihen $X = (X_t : t \in \mathbb{Z})$ mit (6.34) zweiseitige MA-Reihen. Eine zweiseitige MA-Reihe X gemäß (6.34) hat nach (6.5) die Spektraldichte

$$f_X = |\varphi|^2 \frac{\sigma_e^2}{2\pi} \quad \text{mit} \quad \varphi = \sum_{j=-\infty}^{+\infty} c_j e^{-ij \cdot}. \quad (6.35)$$

□

Aufgaben

Aufgabe 6.1 Sei $\varphi : [-\pi, \pi] \rightarrow \mathbb{C}$ eine stetig differenzierbare Funktion mit $\varphi(-\pi) = \varphi(\pi)$. Dann konvergiert bekanntlich die Fourierreihe

$$\frac{1}{2\pi} \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} = \varphi(\omega)$$

im Raum $L_2[-\pi, \pi]$ bzgl. des Lebesguemaßes λ gegen φ , wobei die Koeffizienten gemäß $b_j := \int_{-\pi}^{\pi} \varphi(\omega) e^{ij\omega} \lambda(d\omega)$ definiert sind. Man zeige:

$$\sum_{j \in \mathbb{Z}} |b_j| < \infty.$$

Aufgabe 6.2 Man verifiziere die im Hinweis zum Beweis von Satz 6.5 aufgeführten Beweisschritte (1) – (6).

Aufgabe 6.3 Sei Δ der Differenzenoperator erster Ordnung, also $\Delta Y_t = Y_t - Y_{t-1}$ und Δ^k seine k -fach iterierte Anwendung. Man zeige

$$\Delta^k t^k = k!$$

und folgere daraus die Aussage (6.16).

Aufgabe 6.4 Man zeige, dass beim Übergang von der stationären zentrierten Zeitreihe $Y = (Y_t : t \in \mathbb{Z})$ zur gefilterten Reihe

$$X_t = \frac{1}{3}[Y_t + Y_{t-1} + Y_{t-2}], \quad \forall t \in \mathbb{Z},$$

die Frequenzen von Y in der Nähe von $\frac{2\pi}{3}$ und $-\frac{2\pi}{3}$ unterdrückt werden.

Aufgabe 6.5 Man berechne zu einem reellen weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $e_t \sim (0, \sigma_e^2)$ die Varianz von $Y_t = e_t - e_{t-1}$ (Problem des over-differencing).

Aufgabe 6.6 Sei $X = (X_t : t \in \mathbb{Z})$ ein reeller, zentrierter, stationärer Prozess mit zugehörigem MOW Z und Spektralmaß μ . Weiter sei $T : (-\pi, \pi] \rightarrow (-\pi, \pi]$ mit $T(\omega) = -\omega$, $\forall \omega \in (-\pi, \pi)$ und $T(\pi) = \pi$ die Spiegelungsabbildung, für die nach Bemerkung 3.7 $\mu = \mu^T$ gilt. Das induzierte MOW Z^T gemäß $Z^T(B) = Z(T^{-1}(B))$ hat deshalb nach Aufgabe 5.5 ebenfalls μ als zugehöriges Maß. Nach dem Beweispunkt (5) zu Satz 6.5 gilt $\overline{Z(B)} = Z^T(B)$. Deshalb gilt für $B \in (-\pi, \pi] \cap \mathbb{B}$

$$U(B) := \frac{1}{2}(Z(B) + Z^T(B)) = \operatorname{Re} Z(B)$$

und

$$V(B) := \frac{i}{2}(Z(B) - Z^T(B)) = -\operatorname{Im} Z(B).$$

Man zeige

a) U und V sind reellwertige MOW's auf $((0, \pi), (0, \pi) \cap \mathbb{B})$ mit zugehörigem Maß $\frac{1}{2}\mu|_{(0, \pi)}$, die orthogonal zueinander sind, d.h. mit $U(B_1) \perp V(B_2)$ für

$B_1, B_2 \in (0, \pi) \cap \mathbb{B}$.

b) Mit der Definition $\int g dU := (\int g dZ + \int g dZ^T)/2$ und entsprechend für V zeige man

$$X_t = \int_{(-\pi, \pi]} \cos(t\omega) dU(\omega) + \int_{(-\pi, \pi]} \sin(t\omega) dV(\omega), \quad \forall t \in \mathbb{Z}. \quad (6.36)$$

c) Falls $\mu(\{0\}) = \mu(\{\pi\}) = 0$ gilt, so zeige man

$$X_t = 2 \int_{(0, \pi)} \cos(t\omega) dU(\omega) + 2 \int_{(0, \pi)} \sin(t\omega) dV(\omega), \quad \forall t \in \mathbb{Z}, \quad (6.37)$$

wobei nun die Integrale rechts stochastische Integrale bzgl. der MOW's aus a) sind. Die Darstellung (6.37) ist im Falle eines diskreten Spektralmaßes mit Träger $\{\omega_1, \dots, \omega_M\}$ gerade die reelle zyklische Zeitreihe (2.23).

Aufgabe 6.7 Man betrachte die folgenden vier Filter φ_k , $k = 1, \dots, 4$, in der Reihendarstellung (6.6) mit Koeffizienten $c_{-1} = \dots = c_2 = 1/4$ für $k = 1$, $c_{-2} = \dots = c_1 = 1/4$ für $k = 2$, $c_{-2} = \dots = c_2 = 1/5$ für $k = 3$ und $c_{\pm 2} = -3/4$, $c_{\pm 1} = 3/4$, $c_0 = 1$ für $k = 4$; alle übrigen Koeffizienten für die vier Filter seien 0.

Man zeige, dass der zusammengesetzte Filter $\varphi = \prod_{k=1}^4 \varphi_k$ die Reihendarstellung (6.6) besitzt mit Koeffizienten $c_j = c_{-j}$, $\forall j$, sowie $c_j = 0$, $\forall j > 7$, und

$$[c_0, \dots, c_7] = \frac{1}{320} [74, 67, 46, 21, 3, -5, -6, -3].$$

Dies ist *Spencer's 15-Punkte Moving Average-Formel*.

Aufgabe 6.8 Sei $X = (X_t : t \in \mathbb{Z})$ ein stationärer Prozess mit Autokovarianzfunktion γ . Die *Autokovarianz-erzeugende* Funktion von X ist definiert gemäß

$$G(z) = \sum_{h=-\infty}^{\infty} \gamma(h) z^h, \quad (6.38)$$

falls die Reihe für alle $z \in \mathbb{C}$ mit $\varrho^{-1} < |z| < \varrho$, $\varrho > 1$ geeignet, konvergiert.

a) Man zeige für eine zweiseitige MA-Reihe wie in (6.34) mit φ aus (6.33), wobei zusätzlich $\sum_{j=-\infty}^{\infty} |c_j| |z|^j < \infty$ für alle $z \in \mathbb{C}$ mit $\varrho^{-1} < |z| < \varrho$, $\varrho > 1$ geeignet, gelte, dass für diese z die Gleichung

$$G(z) = \sigma_e^2 \varphi(z) \varphi(z^{-1})$$

erfüllt ist.

b) Für eine MA(2)-Reihe $X_t = e_t + b_1 e_{t-1} + b_2 e_{t-2}$ zeige man

$$G(z) = \sigma_e^2 ((1 + b_1^2 + b_2^2) + (b_1 + b_1 b_2)(z + z^{-1}) + b_2(z^2 + z^{-2}))$$

und leite daraus die Autokovarianzen $\gamma(h)$, $h \in \mathbb{Z}$, ab.

ARMA–Modelle

In diesem Abschnitt untersuchen wir die in Kapitel 2 schon kurz eingeführten „autoregressive moving average“-Zeitreihen, kurz *ARMA*-Reihen, im Detail. Ihre Bedeutung ergibt sich unter anderem daraus, dass man sie so konstruieren kann, dass ihre Spektraldichte jede vorgegebene stetige Spektraldichte beliebig gut approximiert, vgl. Satz 9.12 und 9.13. *ARMA*-Reihen sind spezielle zweiseitige *MA*-Reihen im Sinne von Beispiel 6.11, deren lineare Struktur sie besonders attraktiv macht.

7.1 Definition und Existenz von ARMA–Reihen

Wir wiederholen die Definition 2.8 von *ARMA*-Zeitreihen aus Kapitel 2.

Definition 7.1 (*ARMA-Reihen*)

Eine *ARMA*(p, q)-Reihe $X = (X_t : t \in \mathbb{Z})$ ist eine komplexe, zentrierte, stationäre Zeitreihe, die eine Gleichung

$$X_t + \sum_{j=1}^p a_j X_{t-j} = e_t + \sum_{j=1}^q b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (7.1)$$

erfüllt mit gegebenen Konstanten $a_j, b_j \in \mathbb{C}$; $p, q \in \mathbb{N}$; $a_p \neq 0$, $b_q \neq 0$, und einem geeigneten reellen oder komplexen weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $e_t \sim (0, \sigma_e^2)$.

Mit $\mathbf{a} = (a_1, \dots, a_p)^T$, $\mathbf{b} = (b_1, \dots, b_q)^T$ und der Varianz σ_e^2 des weißen Rauschens schreibt man manchmal in Kurzform $X \sim \text{ARMA}(p, q, \mathbf{a}, \mathbf{b}, \sigma_e^2)$.

Eine Lösung $X = (X_t : t \in \mathbb{Z})$ muss natürlich auf demselben Hintergrundraum (Ω, \mathcal{A}, P) wie das weiße Rauschen $e = (e_t : t \in \mathbb{Z})$ erklärt sein und die Gleichung (7.1) ist wie üblich als Gleichung der Äquivalenzklassen der auftretenden Zufallsvariablen oder gleichbedeutend als fast sichere Gleichheit von Repräsentanten zu verstehen.

Die sogenannte Differenzengleichung (7.1) setzt sich aus dem *autoregressiven* AR-Anteil der linken Seite und dem *moving average* MA-Anteil der rechten Seite zusammen. Ist $q = 0$, d.h. verschwindet der Summenterm auf der rechten Seite in (7.1), so ist X eine $AR(p)$ -Reihe, in Kurzform $X \sim AR(p, \mathbf{a}, \sigma_e^2)$. Ist entsprechend $p = 0$, so ist X eine $MA(q)$ -Reihe, in Kurzform $X \sim MA(q, \mathbf{b}, \sigma_e^2)$.

Wir rekapitulieren die schon früher erwähnten Modellvorstellungen, die zu den ARMA-Prozessen geführt haben: Bei einer $AR(p)$ -Reihe bedeutet (7.1) gerade

$$X_t = -(a_1 X_{t-1} + \cdots + a_p X_{t-p}) + e_t, \quad \forall t \in \mathbb{Z}, \quad (7.2)$$

Dies ist eine Regressionsgleichung wie sie aus der Statistik bekannt ist. Allerdings sind dort die *Regressoren* X_{t-1}, \dots, X_{t-p} meist feste, unabhängige Größen, während sie hier aus den Beobachtungen selber stammen, was zum Namen *auto*-regressiv geführt hat. Die Modellvorstellung, die hinter der Gleichung (7.2) steht, ist offensichtlich: Der Wert X_t der Zeitreihe X zum Zeitpunkt t wird durch die mit $(-a_1), \dots, (-a_p)$ gewichteten, vergangenen Werte X_{t-1}, \dots, X_{t-p} bestimmt, sowie durch eine damit unkorrelierte Störung e_t .

MA-Reihen sind aus der Modellvorstellung entstanden, dass der Wert X_t einer Messgröße durch gewichtete „Schocks“ e_s der Vergangenheit $s < t$ beeinflusst wird. X_t kann z.B. ein ökonomischer Indikator sein, der durch verschiedene zufällige Ereignisse wie Streiks, Katastrophen, usw. beeinflusst wird, die nicht nur momentan wirken, sondern noch Nachwirkungen über einen gewissen Zeitraum haben können.

Das Zusammenfügen von AR- und MA-Gleichungen zu ARMA-Gleichungen geschieht mit dem Ziel, gegebene Zeitreihen mit möglichst wenigen Parametern beschreiben zu können (principle of parsimony).

Bemerkung 7.2 (Die ARMA-Gleichung in Filterdarstellung)

Seien mit $a_0 := b_0 := 1$

$$A(z) = \sum_{j=0}^p a_j z^j, \quad B(z) = \sum_{j=0}^q b_j z^j, \quad \forall z \in \mathbb{C},$$

die zugehörigen sogenannten z -Transformationen, für die $A(0) = B(0) = 1$ gilt, und $\varphi(\omega) = A(e^{-i\omega})$, sowie $\psi(\omega) = B(e^{-i\omega})$, $\omega \in (-\pi, \pi]$, damit gebildete Filter. Dann schreibt sich (7.1) als Gleichung der gefilterten Reihen X und e gemäß

$$\int e^{it\cdot} \varphi dZ_X = \int e^{it\cdot} \psi dZ_e, \quad \forall t \in \mathbb{Z}. \quad (7.3)$$

Dabei bezeichnen Z_X und Z_e die zu X und e gehörenden MOW's. Hieraus folgt mit den Bezeichnungen aus Kapitel 6 und den Spektralmaßen μ_X und μ_e

$$\int_B |\varphi|^2 d\mu_X = \int_B |\psi|^2 d\mu_e, \quad \forall B \in \mathbb{B} \cap (-\pi, \pi]. \quad (7.4)$$

□

Bemerkung 7.3 (Die ARMA-Gleichung in Kompaktform)

Mit dem sogenannten *Lag-Operator* L gemäß $L^i X_t = X_{t-i}$, $i \in \mathbb{Z}$, kann man die Differenzengleichung (7.1) unter Verwendung der formalen Polynome $A(L) = \sum_{j=0}^p a_j L^j$ und $B(L) = \sum_{j=0}^q b_j L^j$ in der kompakten Form

$$A(L) X_t = B(L) e_t, \quad \forall t \in \mathbb{Z}, \quad (7.5)$$

schreiben. Ist dann $C(z) = \sum_{j=0}^r c_j z^j$ ein weiteres Polynom und $\chi(\omega) := C(e^{-i\omega})$ der zugehörige Filter, so kann die Anwendung von χ auf (7.5) auch gemäß

$$C(L) A(L) X_t = C(L) B(L) e_t, \quad \forall t \in \mathbb{Z}, \quad (7.6)$$

geschrieben werden. Wegen $L^i (L^j X_t) = L^{i+j} X_t = L^j (L^i X_t)$ kann man mit den formalen Polynomen wie mit üblichen Polynomen rechnen. Es gilt z.B. $C(L) A(L) X_t = A(L) C(L) X_t$, $t \in \mathbb{Z}$, usw. □

Der folgende Satz gibt eine notwendige und hinreichende Bedingungen für die Existenz von Lösungen der ARMA-Gleichung (7.1) an und beschreibt für den Fall, dass mindestens eine Lösung existiert, die Lösungsmenge. Für seine Formulierung verwenden wir die folgenden Bezeichnungen:

Haben A und B gemeinsame Nullstellen, so gibt es Polynome A_0 und B_0 ohne gemeinsame Nullstellen mit $A_0(0) = B_0(0) = 1$ und ein Polynom H derart, dass $A = A_0 \cdot H$ und $B = B_0 \cdot H$ gilt. (Falls A und B keine gemeinsamen Nullstellen besitzen, werde $H \equiv 1$ gesetzt.)

Die zu A , B , A_0 , B_0 , H gehörigen Filter bezeichnen wir der Reihe nach mit φ , φ_0 , ψ , ψ_0 , η , also z.B. $A(e^{-i\omega}) = \varphi(\omega)$. Außerdem bezeichnen wir wie bisher mit Z_U und μ_U das MOW und Spektralmaß zu einer stationären Zeitreihe $U = (U_t : t \in \mathbb{Z})$. Damit schreibt sich die ARMA-Gleichung in Filterdarstellung gemäß

$$\int e^{it\cdot} \varphi_0 \eta dZ_X = \int e^{it\cdot} \psi_0 \eta dZ_e, \quad \forall t \in \mathbb{Z}. \quad (7.7)$$

Die durch Weglassen von η , d.h. Kürzen gemeinsamer Faktoren von A und B , entstehende Gleichung in Kompaktform

$$A_0(L) X_t = B_0(L) e_t, \quad \forall t \in \mathbb{Z}, \quad (7.8)$$

oder in Filterdarstellung

$$\int e^{it\cdot} \varphi_0 dZ_X = \int e^{it\cdot} \psi_0 dZ_e, \quad \forall t \in \mathbb{Z}, \quad (7.9)$$

nennen wir die *reduzierte ARMA-Gleichung*. Wir werden sehen, dass man sich in allen interessierenden Fällen auf die reduzierte ARMA-Gleichung zurückziehen kann. Man kann dann deshalb ohne Beschränkung der Allgemeinheit von vornherein annehmen, dass die Originalgleichung (7.1) schon in reduzierter Form vorliegt, also dass A und B keine gemeinsamen Nullstellen besitzen.

Satz 7.4 (*Lösungen der ARMA-Gleichung*)

a) Genau dann gibt es zu einem beliebig vorgegebenem weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ mindestens eine ARMA(p, q)-Reihe, die (7.1) erfüllt, wenn die reduzierte z -Transformation A_0 keine Nullstelle z mit $|z| = 1$ besitzt. In diesem Fall ist $X_0 = (X_{0t} : t \in \mathbb{Z})$ mit

$$X_{0t} = \int e^{it \cdot} \frac{\psi_0}{\varphi_0} dZ_e, \quad \forall t \in \mathbb{Z}, \quad (7.10)$$

eine Lösung sowohl der Ursprungsgleichung (7.1) als auch der reduzierten Gleichung (7.9), die nur diese einzige Lösung besitzt.

b) Die allgemeine Lösung X von (7.7) im Fall $A_0(z) \neq 0$ für $|z| = 1$ (falls also eine Lösung existiert) hat die Gestalt

$$X_t = X_{0t} + X_{1t}, \quad \forall t \in \mathbb{Z}, \quad (7.11)$$

mit X_0 aus (7.10) und einer zyklischen Zeitreihe X_1 gemäß

$$X_{1t} = \sum_{j=1}^r e^{i\omega_j t} Z_j, \quad \forall t \in \mathbb{Z}, \quad (7.12)$$

wobei $N := \{\omega_1, \dots, \omega_r\}$ die Menge der Nullstellen ω von $H(e^{-i\omega})$ auf $(-\pi, \pi]$ sei und Z_1, \dots, Z_r unkorrelierte, zentrierte Elemente aus $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ bezeichnen, die untereinander und mit allen X_{0t} unkorreliert sind.

Beweis: a) Wir zeigen zunächst, dass, falls die Originalgleichung (7.7) eine Lösung besitzt, so auch die reduzierte Gleichung und umgekehrt. Sei also X eine Lösung der ARMA-Gleichung (7.7). Dann ist wegen $\eta = \eta 1_{N^c}$ und der Substitutionsregel auch die durch $Y_t := \int e^{it \cdot} 1_{N^c} dZ_X$ definierte Zeitreihe Y eine Lösung von (7.7). Das zugehörige Spektralmaß $\mu_Y(B) = \mu_X(B \cap N^c)$ legt per Konstruktion keine Masse auf die Nullstellen von η . Deshalb gilt $|\eta| > 0$ μ_Y -f.s. und natürlich auch μ_e -f.s.. Eine Anwendung des Filters η^{-1} auf beide Seiten der Gleichung (7.7) für Y liefert nach Kürzen von η auf der rechten Seite die reduzierte Gleichung (7.9) mit Y statt X . Wenn es also eine Lösung X von (7.7) gibt, so auch eine Lösung Y der reduzierten Gleichung. (Wir werden sehen, dass die reduzierte Gleichung höchstens eine einzige Lösung besitzt.) Umgekehrt zeigt eine Anwendung des Filters η auf die reduzierte Gleichung (7.9), dass eine Lösung der reduzierten Gleichung auch eine Lösung der Originalgleichung (7.7) ist. Für den weiteren Existenzbeweis können wir

also die reduzierte Gleichung (7.9) zugrunde legen.

Nehmen wir zunächst an, dass $\varphi_0(\omega) = A_0(e^{-i\omega}) \neq 0 \forall \omega \in [-\pi, \pi]$ gilt. Dann folgt die Ungleichung $\inf \{|\varphi_0(\omega)| : -\pi \leq \omega \leq \pi\} > 0$. Somit ist die Funktion $|\psi_0/\varphi_0|$ beschränkt, insbesondere ein Element von $L_2^{\mathbb{C}}(\mu_e)$. Also ist die Reihe $X_0 = (X_{0t} : t \in \mathbb{Z})$ mit

$$X_{0t} = \int e^{it \cdot} \frac{\psi_0}{\varphi_0} dZ_e, \quad \forall t \in \mathbb{Z}, \quad (7.13)$$

zentriert und stationär mit

$$\int e^{it \cdot} \varphi_0 dZ_{X_0} = \int e^{it \cdot} \varphi_0 \cdot \frac{\psi_0}{\varphi_0} dZ_e = \int e^{it \cdot} \psi_0 dZ_e, \quad \forall t \in \mathbb{Z}.$$

Die Reihe (7.13) ist also Lösung von (7.9). Diese Lösung ist eindeutig bestimmt, denn Anwendung des Filters φ_0^{-1} auf (7.9) liefert gerade obige Formel (7.13).

Umgekehrt gebe es eine stationäre Lösung X von (7.9). Wir zeigen, dass dann A_0 keine Nullstellen auf der Einheitskreislinie hat. Für einen Widerspruchsbeweis sei dazu angenommen, dass $z_0 = e^{-i\omega_0}$, $\omega_0 \in (-\pi, \pi]$, eine Nullstelle von A_0 ist, sodass sich A_0 mit einem geeigneten Polynom A_1 in der Form $A_0(z) = (1 - z/z_0) A_1(z)$ schreiben lässt. Wegen der Gleichung

$$\left| 1 - \frac{e^{-i\omega}}{e^{-i\omega_0}} \right| = \left| e^{\frac{i(\omega - \omega_0)}{2}} - e^{\frac{-i(\omega - \omega_0)}{2}} \right| = 2 |\sin(\omega - \omega_0)|$$

ergibt sich für die gefilterte Reihe $U_t = \int A_1(e^{-i \cdot}) e^{it \cdot} dZ_X$ entsprechend (7.4) die Gleichung

$$\int_E 4 \sin^2((\omega - \omega_0)/2) d\mu_U(\omega) = \int_E |\psi_0|^2 d\mu_e \quad \forall E \in \mathbb{B} \cap (-\pi, \pi]. \quad (7.14)$$

Da φ_0 und ψ_0 keine gemeinsamen Nullstellen besitzen, gilt für ein genügend kleines $\varepsilon > 0$ und eine Konstante $c > 0$ die Ungleichung

$$|\psi_0(\omega)|^2 \geq c \quad \text{für } \omega \in I(\varepsilon) = \{\omega \in (-\pi, \pi] : |\omega - \omega_0| \leq \varepsilon\}.$$

Wegen $|\sin x| \leq |x|$ ergibt sich mit $E = I(\varepsilon)$ aus (7.14) durch Abschätzung der linken Seite nach oben und der rechten Seite nach unten die Ungleichung

$$\varepsilon^2 \mu_U(-\pi, \pi] \geq c \cdot \frac{\sigma_e^2}{2\pi} \cdot \varepsilon,$$

wobei $\mu_U(-\pi, \pi]$ endlich ist. Nach Kürzen von ε ergibt sich für $\varepsilon \downarrow 0$ offenbar ein Widerspruch und damit $A_0(z) \neq 0$ für $|z| = 1$.

b) Ist X Lösung von (7.7), so ist, wie im Beweisteil zu a) schon gezeigt, die Zeitreihe $Y_t = \int e^{i\omega t} 1_{N^c}(\omega) dZ_X(\omega)$, $t \in \mathbb{Z}$, Lösung von (7.9). Da diese aber

nur eine einzige Lösung besitzt, nämlich X_0 aus (7.12), gilt $Y = X_0$. Mit $X_{1t} = \int e^{i\omega t} 1_N(\omega) dZ_X(\omega)$, $t \in \mathbb{Z}$, gilt dann zunächst offenbar (7.11). Wegen $e^{it \cdot} 1_N = \sum_{j=1}^r e^{it\omega_j} 1_{\{\omega_j\}}$ folgt die Gleichheit (7.12) mit $Z_j = Z_X(\{\omega_j\})$, wobei wegen der Grundeigenschaften von MOW's offenbar die dort angegebenen Eigenschaften für die Z_j gelten.

Sind umgekehrt solche Z_j gegeben, so gehört nach Beispiel 5.9 zur Zeitreihe (7.12) das MOW $Z_1(B) := \sum_{j=1}^r Z_j 1_B(\omega_j)$ mit Spektralmaß $\mu_1(B) = \sum_{j=1}^r \|Z_j\|^2 1_B(\omega_j)$. Wegen $\eta(\omega_j) = 0 \forall j$ ergibt die Anwendung des Filters η auf $X_{1t} = \int e^{it \cdot} dZ_1$ den Wert 0, folglich auch $\int \varphi_0 \eta dZ_{X_1} = 0$. Damit ist $X_0 + X_1$ Lösung von (7.7). Insgesamt ist also b) gezeigt. \square

Bemerkung 7.5 (Eindeutigkeit von Lösungen, reelle Lösungen)

a) Aus Satz 7.4 folgt unmittelbar, dass nur dann mehrere Lösungen auftreten können, wenn A und B gemeinsame Nullstellen *auf der Einheitskreislinie* besitzen. Diese Situation ist auch dadurch charakterisiert, dass das Spektralmaß einer Lösung einen diskreten Anteil hat, der von der zyklischen Zeitreihe in (7.12) herrührt.

b) Interessiert man sich nur für Lösungen, die ein stetiges Spektralmaß besitzen (und das wird bei uns stets so sein), so kann es nur eine einzige Lösung geben, nämlich diejenige aus (7.10), selbst wenn A und B gemeinsame Nullstellen z mit $|z| = 1$ haben. Man kann in diesem Fall äquivalent zur reduzierten ARMA-Gleichung übergehen, ob es nun gemeinsame Nullstellen z von A und B mit $|z| = 1$ gibt oder nicht. Damit man nicht mehr reduzieren muss, fordert man deshalb meist, dass die z -Transformation A in der Originalgleichung (7.1) schon in reduzierter Form vorliegt, dass also A und B keine gemeinsamen Nullstellen besitzen. Damit dann überhaupt eine Lösung existiert, darf in diesem Fall A keine Nullstelle auf der Einheitskreislinie besitzen.

Nach Satz 7.4 haben gemeinsame Faktoren von A und B in der Ursprungsgleichung (7.1), die nicht auf der Einheitskreislinie liegen, keinerlei Einfluss auf die Lösungsmenge, man müsste sie also beim Übergang zu einer reduzierten Form, wenn es nur um Existenz und Eindeutigkeit von Lösungen geht, eigentlich nicht wegstreichen. Man tut es jedoch trotzdem, denn derartige gemeinsame Faktoren von A und B würden bewirken, dass die Parametrisierung durch die Koeffizienten a_j und b_j nicht mehr eindeutig ist, siehe dazu auch Abschnitt 8.4.

c) Wir sprechen vom *reellen Fall*, wenn das weiße Rauschen und die Koeffizienten von A_0 und B_0 in der reduzierten Gleichung reell sind. Nach Satz 6.5 folgt dann, dass auch die Lösung der reduzierten Gleichung reell ist, da dann offenbar die Symmetriebedingung (6.8) für ψ_0/φ_0 erfüllt ist. \square

Nachdem wir die Existenz und Eindeutigkeit der ARMA-Zeitreihen eingehend untersucht haben, wollen wir uns nun deren Eigenschaften zuwenden. Eine wichtige Kenngröße einer stationären Zeitreihe ist ihr Spektralmaß bzw. (bei Existenz) ihre Spektraldichte. Im folgenden Satz werden wir sehen, dass

stationäre ARMA-Reihen ohne zyklischen Anteil stets eine stetige Spektraldichte besitzen, die wir darüberhinaus explizit hinschreiben können.

Satz 7.6 (Die Spektraldichte bei ARMA-Reihen)

Ist X Lösung der reduzierten ARMA-Gleichung (7.9), so besitzt sie die stetige Spektraldichte

$$f_X = \frac{\sigma_e^2 |\psi_0|^2}{2\pi |\varphi_0|^2}. \quad (7.15)$$

Im reellen Fall ist f_X symmetrisch um 0. Da f_X strikt positiv ist bis auf höchstens endlich viele Nullstellen von ψ_0 , sind die Maße μ_X und μ_e äquivalent.

Der Beweis zu Satz 7.6 ergibt sich direkt aus (6.5) und (7.10).

Unter denselben Voraussetzungen wie im vorangehenden Satz kann man außerdem zeigen, dass ARMA-Reihen stets eine zweiseitige MA(∞)-Darstellung besitzen.

Satz 7.7 (Lösungen der ARMA-Gleichung in Reihendarstellung)

Sei X Lösung der reduzierten ARMA-Gleichung (7.9). Dann besitzt X die zweiseitige MA-Darstellung

$$X_t = \sum_{j=-\infty}^{+\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad \sum_{j=-\infty}^{+\infty} |c_j| < \infty. \quad (7.16)$$

Die Konstanten $c_j, j \in \mathbb{Z}$, sind eindeutig bestimmt gemäß $c_j = \langle X_t, e_{t-j} \rangle / \sigma_e^2$, $t \in \mathbb{Z}$. Man kann sie als die Koeffizienten der Laurent-Entwicklung von B_0/A_0 erhalten, die für ein geeignetes $\varrho > 1$ im Ringbereich $1/\varrho \leq |z| \leq \varrho$ konvergiert. Im reellen Fall ist die Lösung X nach Bemerkung 7.5 c) reell und damit sind dann die Konstanten c_j ebenfalls reell.

Beweis: Für die Laurent-Reihe

$$\frac{B_0(z)}{A_0(z)} = \sum_{j=-\infty}^{+\infty} c_j z^j, \quad \frac{1}{\varrho} \leq |z| \leq \varrho,$$

gilt stets, siehe z.B. Remmert (1991),

$$\sum_{j=-\infty}^{+\infty} |c_j| < \infty, \quad \text{also auch} \quad \sum_{j=-\infty}^{+\infty} |c_j|^2 < \infty. \quad (7.17)$$

Deshalb konvergiert die Reihe

$$\frac{\psi(\omega)}{\varphi(\omega)} = \sum_{j=-\infty}^{+\infty} c_j e^{-ij\omega}$$

auch in $L_2^{\mathbb{C}}(\mu_e)$. Es folgt

$$X_t = \int e^{it \cdot \frac{\psi}{\varphi}} dZ_e = \sum_{j=-\infty}^{+\infty} c_j \int e^{i(t-j) \cdot} dZ_e = \sum_{j=-\infty}^{+\infty} c_j e_{t-j}.$$

□

Bemerkung 7.8 (Geometrisches Abklingen der Koeffizienten)

Ist X Lösung der reduzierten ARMA-Gleichung, so fallen die Koeffizienten c_j der zweiseitigen MA-Darstellung (7.19) in geometrischer Rate ab, d.h. es gilt

$$|c_j| \leq c \cdot \varrho^{-|j|}, \quad \forall j \in \mathbb{Z}, \quad (7.18)$$

mit geeigneten Konstanten $c > 0$ und $\varrho > 1$. Dies folgt unmittelbar daraus, dass in der Laurent-Entwicklung (7.17) sowohl $\sum_{j=0}^{\infty} c_j z^j$ eine Potenzreihe in z ist, die für $|z| \leq \varrho$ konvergiert als auch $\sum_{j=1}^{\infty} c_{-j} (1/z)^j$ eine Potenzreihe in $1/z$ ist, die entsprechend für $|1/z| \leq \varrho$ konvergiert. Insbesondere gilt $(|c_j| + |c_{-j}|)\varrho^j \rightarrow 0$ für $j \rightarrow \infty$ und damit (7.18). □

7.2 Kausalität und Invertibilität von ARMA-Reihen

In einer zweiseitigen MA-Darstellung (7.16) wird X_t als unendliche Linearkombination von Störungen e_s sowohl der Vergangenheit als auch der Zukunft ausgedrückt. Das ist für in der Realität auftretende Zeitreihen natürlich nicht möglich. Vielmehr sollten in einer solchen Darstellung allein Störungen der Vergangenheit vorkommen. Dies führt zur Begriffsbildung der Kausalität.

Definition 7.9 (Kausalität)

Eine ARMA(p, q)-Reihe $X = (X_t : t \in \mathbb{Z})$, die die Differenzengleichung (7.1) erfüllt, heißt *kausal*, wenn eine einseitige MA-Darstellung existiert gemäß

$$X_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad \sum_{j=0}^{\infty} |c_j| < \infty. \quad (7.19)$$

Die Forderung der Kausalität von X beinhaltet zum einen, dass X überhaupt eine Lösung der ARMA-Gleichung ist und impliziert zum anderen (wegen der Reihendarstellung), dass X ein stetiges Spektralmaß besitzt. Damit ist nach Bemerkung 7.9 klar, dass es sich bei X nur um die eindeutig bestimmte Lösung der reduzierten ARMA-Gleichung handeln kann, die nach Satz 7.7 die eindeutig bestimmte zweiseitige MA(∞)-Darstellung (7.16) besitzt. Kausale Zeitreihen sind demnach genau die Lösungen reduzierter ARMA-Gleichungen, bei denen die Koeffizienten c_j in der Entwicklung (7.16) für negative j verschwinden.

Unter der Voraussetzung, dass X Lösung der reduzierten ARMA-Gleichung ist, gilt wegen $\langle X_t, e_k \rangle = c_{t-k} \sigma_e^2$ die Darstellung (7.19) genau dann, wenn gilt:

$$\langle X_t, e_k \rangle = 0 \quad \forall k, t \in \mathbb{Z} \text{ mit } k > t, \quad (7.20)$$

d.h. wenn zukünftige Störungen e_k unkorreliert mit X_t sind ($k > t$).

Satz 7.10 (*Charakterisierung der Kausalität*)

Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ ist genau dann kausal, wenn die reduzierte z -Transformation A_0 keine Nullstellen z mit $|z| \leq 1$ besitzt.

Beweis: Ist X kausal, so ist wegen $c_j = 0 \ \forall j < 0$ die Laurent-Entwicklung von B_0/A_0 in $1/\varrho \leq |z| \leq \varrho$ sogar eine für $|z| \leq \varrho$ konvergente Potenzreihe

$$P(z) = \sum_{j=0}^{\infty} c_j z^j. \quad (7.21)$$

Wegen $B_0(z) = P(z)A_0(z)$ für $|z| = 1$ gilt nach dem Identitätssatz für holomorphe Funktionen diese Gleichung sogar für $|z| < \varrho$. Da A_0 und B_0 keine gemeinsamen Nullstellen haben, folgt $A_0(z) \neq 0$ für $|z| \leq 1$. Gilt umgekehrt $A_0(z) \neq 0$ für $|z| \leq 1$, so lässt sich bekanntlich B_0/A_0 in eine für $|z| < \varrho$, $\varrho > 1$ geeignet, konvergente Potenzreihe entwickeln, die dann die Laurentreihe von B_0/A_0 ist mit $c_j = 0 \ \forall j < 0$. \square

In Abschnitt 8.2 werden wir eine sehr effektive rekursive Methode zur Berechnung der Koeffizienten c_j der Potenzreihe $P(z)$ besprechen.

Die duale Fragestellung zur Kausalität ist, ob sich die Störungen e_t in ähnlicher Weise aus den Werten X_t, X_{t-1}, \dots zurückberechnen lassen. Dazu geben wir die folgende Definition.

Definition 7.11 (*Invertibilität*)

Eine ARMA(p, q)-Reihe $X = (X_t : t \in \mathbb{Z})$, die die reduzierte Gleichung (7.9) erfüllt, heißt invertibel, wenn Konstanten $d_j \in \mathbb{C}$, $j = 0, 1, 2, \dots$, existieren mit

$$e_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \quad \forall t \in \mathbb{Z}, \quad \sum_{j=0}^{\infty} |d_j| < \infty. \quad (7.22)$$

In Analogie zu Satz 7.10 charakterisiert der folgende Satz die Invertibilität.

Satz 7.12 (*Charakterisierung der Invertibilität*)

Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ ist genau dann invertibel, wenn die reduzierte z -Transformation B_0 nur Nullstellen außerhalb des Einheitskreises besitzt.

Beweis: Ohne Einschränkung der Allgemeinheit können wir annehmen, dass die ARMA-Gleichung schon in reduzierter Form vorliegt, dass also $A = A_0$ und $B = B_0$ gilt. Zunächst sei $B(z) \neq 0$ für $|z| \leq 1$ vorausgesetzt. Dann ist ψ^{-1} gemäß $\psi^{-1}(\omega) = 1/B(e^{-i\omega})$ beschränkt auf $[-\pi, \pi]$, also quadratintegrabel bzgl. jedes Spektralmaßes und somit ein Filter, dessen Anwendung auf (7.3) die Gleichung

$$\int e^{it \cdot} \frac{\varphi}{\psi} dZ_X = e_t, \quad \forall t \in \mathbb{Z}, \quad (7.23)$$

ergibt. Nun ist A/B für $|z| < \varrho$ mit geeignetem $\varrho > 1$ in eine Potenzreihe $A/B(z) = \sum_{j=0}^{\infty} d_j z^j$ entwickelbar mit $\sum_{j=0}^{\infty} |d_j| < \infty$. Mit

$$\frac{\varphi}{\psi}(\omega) = \sum_{j=0}^{\infty} d_j e^{-ij\omega}, \quad \forall \omega \in (-\pi, \pi],$$

folgt die Reihenentwicklung (7.22) direkt aus Bemerkung 6.4 a). Gelte nun umgekehrt (7.22), also mit

$$D(z) := \sum_{j=0}^{\infty} d_j z^j \quad \text{und} \quad \chi(\omega) := D(e^{-i\omega})$$

die Gleichung

$$e_t = \int e^{it \cdot} \chi dZ_X, \quad \forall t \in \mathbb{Z}.$$

Anwendung des Filters $\varphi \in L_2^{\mathbb{C}}(\mu_e)$ auf beide Seiten liefert mit $t = 0$ nach Lemma 6.6 die Gleichung

$$\int \varphi dZ_e = \int \chi \varphi dZ_X,$$

zusammen mit (7.3) und der Substitutionsregel (5.39) also die Beziehung

$$\int \varphi dZ_e = \int \chi \psi dZ_e,$$

woraus nach der Rechenregel (5.23) die Gleichheit $\varphi = \psi \cdot \chi$ in $L_2^{\mathbb{C}}(\mu_e)$ folgt. Wegen der Stetigkeit von φ, ψ, χ gilt diese Gleichheit auf dem gesamten Intervall $[-\pi, \pi]$. Nun ist $B(z)D(z) = \sum_{j=0}^{\infty} \alpha_j z^j$ eine Potenzreihe, die für $|z| \leq 1$ konvergiert und für $|z| = 1$ mit $A(z)$ übereinstimmt. Damit gilt $\varphi(\omega) = \sum_{j=0}^{\infty} \alpha_j e^{-i\omega} \forall \omega \in [-\pi, \pi]$. Heranmultiplizieren auf beiden Seiten dieser Gleichung mit den Funktionen $e^{-ik\omega}$ und Ausintegrieren mit dem Lebesgue-Maß liefert wegen der Orthogonalität dieser Funktionen (vgl. Bemerkung 6.4 c)) die Gleichheit $a_k = \alpha_k \forall k \in \mathbb{Z}$, also die Gleichheit $A(z) = B(z)D(z) \forall |z| \leq 1$. Dann kann B keine Nullstelle z mit $|z| \leq 1$ besitzen, denn diese wäre auch eine solche von A , was nach Voraussetzung ausgeschlossen ist. \square

Der folgende Satz stellt sicher, dass zu einer Zeitreihe, die Lösung einer ARMA-Gleichung in reduzierter Form ist, eine neue ARMA-Gleichung und ein neues weißes Rauschen gebildet werden kann derart, dass die Zeitreihe in der neuen Situation kausal und invertibel ist.

Satz 7.13 (*Übergang zu einer kausalen und invertiblen Reihe*)

Sei $X = (X_t : t \in \mathbb{Z})$ eine $\text{ARMA}(p, q)$ -Reihe, wobei die ARMA -Gleichung in reduzierter Form vorliege. Seien $\alpha_1, \dots, \alpha_p \in \mathbb{C}$ die Nullstellen von A , $\beta_1, \dots, \beta_q \in \mathbb{C}$ die Nullstellen von B , sowie $J := \{j : |\alpha_j| < 1\}$ und $K := \{k : |\beta_k| < 1\}$. Ferner sei

$$\tilde{A}(z) := \prod_{j \in J} (1 - \bar{\alpha}_j \cdot z) \prod_{j \notin J} (1 - \alpha_j^{-1} \cdot z),$$

ähnlich $\tilde{B}(z)$. Dann ist $\epsilon := (\epsilon_t : t \in \mathbb{Z})$ gemäß

$$\epsilon_t := \frac{B(L) \tilde{A}(L)}{A(L) \tilde{B}(L)} e_t, \quad \forall t \in \mathbb{Z}, \quad (7.24)$$

ein weißes Rauschen mit $\sigma_\epsilon^2 = \lambda \cdot \sigma_e^2$

$$\lambda := \prod_{j \in J} |\alpha_j|^2 \cdot \prod_{k \in K} |\beta_k|^{-2}, \quad (7.25)$$

und es gilt

$$\tilde{A}(L) X_t = \tilde{B}(L) \epsilon_t, \quad \forall t \in \mathbb{Z}. \quad (7.26)$$

Eventuell gemeinsame Faktoren von \tilde{A} und \tilde{B} können gekürzt werden, vgl. Bemerkung 7.5 b). Nach Satz 7.10 und Satz 7.12 ist X dann kausal und invertibel bzgl. des neuen weißen Rauschens $\epsilon = (\epsilon_t : t \in \mathbb{Z})$.

Der einfache Beweis, bei dem die Spektraldichte von $\epsilon = (\epsilon_t : t \in \mathbb{Z})$ zu berechnen ist, wird als Aufgabe 7.1 gestellt.

7.3 Lineare L_1 -Filter

Neben der L_2 -basierten Filtertheorie in Kapitel 6, ist es auch möglich, für lineare Filter mit absolut summierbaren Koeffizienten entsprechende Aussagen zu formulieren, wobei die Filter auf beliebige reelle oder komplexe stochastische Prozesse $X = (X_t : t \in \mathbb{Z})$ mit der einzigen Bedingung

$$\sup_{t \in \mathbb{Z}} E|X_t| < \infty$$

angewendet werden können.

Seien also a_0, a_1, \dots Koeffizienten mit $\sum_{j=0}^{\infty} |a_j| < \infty$. Dann wird durch

$$Y_t = \sum_{j=0}^{\infty} a_j X_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (7.27)$$

ein Prozess $Y = (Y_t : t \in \mathbb{Z})$ mit $\sup_{t \in \mathbb{Z}} E|Y_t| < \infty$ definiert. Dabei gilt die Konvergenz in (7.27) sowohl f.s. als auch im Raum $L_1^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ bzw. $L_1^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ (s. Aufgabe 7.8). Unter Verwendung der z -Transformation $A(z) = \sum_{j=0}^{\infty} a_j z^j$, die eine für $|z| \leq 1$ konvergente Potenzreihe darstellt, schreibt sich (7.27) mit dem Lag-Operator L in Kompaktform gemäß

$$Y_t = A(L)X_t, \quad \forall t \in \mathbb{Z}. \quad (7.28)$$

Wir nennen die Koeffizienten (a_j) oder auch A einen *linearen L_1 -Filter*. Genauer müssten wir nach unserer früheren Sprechweise eigentlich von einem *kausalen* linearen Filter sprechen. Da wir es aber bei späteren Anwendungen in Kapitel 14 einzig mit solchen Filtern zu tun haben werden, lassen wir diesen Zusatz weg. Mit linearen L_1 -Filtern kann man wie mit üblichen Potenzreihen rechnen, ganz analog zu Kapitel 6. Ist also zum Beispiel $B(z) = \sum_{j=0}^{\infty} b_j z^j$ ein weiterer L_1 -Filter, so ist auch die Potenzreihe $C(z) = \sum_{k=0}^{\infty} c_k z^k = A(z)B(z) = B(z)A(z)$ mit

$$c_k = \sum_{j=0}^k a_j b_{k-j} = \sum_{j=0}^k b_j a_{k-j}$$

ein L_1 -Filter, und es gilt (s. Aufgabe 7.9)

$$C(L)X_t = A(L)[B(L)X_t] = B(L)[A(L)X_t], \quad \forall t \in \mathbb{Z}, \quad (7.29)$$

wobei die Konvergenzen wiederum sowohl f.s. als auch in $L_1^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ gelten.

Aufgaben

Aufgabe 7.1 Man beweise Satz 7.13. Hinweis: $|1 - \bar{\alpha}e^{-i\omega}| = |\alpha||1 - \frac{1}{\alpha}e^{-i\omega}|$.

Aufgabe 7.2 Sei $e = (e_t : t \in \mathbb{Z})$ ein weißes Rauschen, $e_t \sim (0, \sigma_e^2)$ und $X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}$, $t \in \mathbb{Z}$, mit einer beliebigen absolut summierbaren Folge b_j , $j \in \mathbb{Z}$. Man zeige für die Autokovarianzfunktion γ_X von $X = (X_t : t \in \mathbb{Z})$ die Gleichheit

$$\sum_{h=-\infty}^{\infty} \gamma_X(h) = \sigma_e^2 \left| \sum_{j=-\infty}^{\infty} b_j \right|^2.$$

Aufgabe 7.3 Sei eine $ARMA(1, 1)$ -Reihe X gegeben durch

$$X_t + aX_{t-1} = e_t + be_{t-1}, \quad e_t \sim (0, \sigma_e^2), \quad |a| < 1.$$

Man berechne die Spektraldichte f von X und zeichne ein Schaubild von f für den Spezialfall $a = 1/2$, $b = 1/4$, $\sigma_e^2 = 1$.

Aufgabe 7.4 Sei $X = (X_t : t \in \mathbb{Z})$ ein reeller $AR(2)$ -Prozess mit z -Transformation $A(z) = 1 + a_1z + a_2z^2$. Man zeige, dass die Kausalitätsbedingung $A(z) \neq 0 \quad \forall z \text{ mit } |z| \leq 1$ äquivalent zu

$$a_2 < 1, \quad |a_1| - a_2 < 1,$$

ist.

Aufgabe 7.5 Für eine reelle $MA(1)$ -Reihe $X = (X_t : t \in \mathbb{Z})$ mit $X_t = e_t + \rho e_{t-1}$, $|\rho| < 1$, und weißem Rauschen $e = (e_t : t \in \mathbb{Z})$ gebe man ein weißes Rauschen $W = (W_t : t \in \mathbb{Z})$ an derart, dass

$$X_t = W_t + \frac{1}{\rho} W_{t-1}, \quad \forall t \in \mathbb{Z},$$

gilt. Dies zeigt, dass ohne die Invertibilitätsannahme die Parametrisierung eines MA -Prozesses nicht eindeutig ist.

Aufgabe 7.6 Man zeige mit einer zum Beweis von Satz 7.12 analogen Beweismethode, dass bei einer $ARMA(p, q)$ -Reihe $X = (X_t : t \in \mathbb{Z})$, bei der A und B keine gemeinsamen Nullstellen besitzen, aus einer Darstellung $X_t = \sum_{j=0}^{\infty} c_j e_{t-j}$, $\sum_{j=0}^{\infty} |c_j| < \infty$, die Aussage $A(z) \neq 0$ für $|z| \leq 1$ folgt.

Aufgabe 7.7 Man zeige, dass bei einer reellen kausalen $AR(p)$ -Reihe gilt

$$\pi(k) = 0 \text{ für } k > p, \text{ sowie } \pi(p) = -a_p.$$

Aufgabe 7.8 Sei $X = (X_t : t \in \mathbb{Z})$ ein reeller oder komplexer stochastischer Prozess mit $\sup_t E|X_t| < \infty$. Falls a_0, a_1, \dots absolut summierbare Koeffizienten sind, so zeige man, dass für alle $t \in \mathbb{Z}$ die Reihe

$$\sum_{j=0}^{\infty} b_j X_{t-j}$$

mit Wahrscheinlichkeit 1 absolut und im Raum $L_1^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ konvergiert und unter der zusätzlichen Voraussetzung $\sup_t EX_t^2 < \infty$ auch im quadratischen Mittel gegen denselben Grenzwert konvergiert.

Aufgabe 7.9 Man zeige die Gleichung (7.29).

Die Autokovarianz und Autokorrelation von ARMA-Reihen im reellen Fall

In diesem Kapitel wollen wir uns mit der Berechnung und allgemeinen Eigenschaften von Autokovarianzen und Autokorrelationen bei ARMA-Reihen $X = (X_t : t \in \mathbb{Z})$ beschäftigen. Dabei setzen wir voraus, dass die ARMA-Gleichung (7.1) in reduzierter Form vorliegt, also A und B keine gemeinsamen Nullstellen besitzen. Weiter liege der reelle Fall vor, d.h. in der definierenden Gleichung (7.1) sind alle auftretenden Koeffizienten a_j und b_j und das weiße Rauschen e und damit X selbst reell.

Weiter werde für das ganze Kapitel angenommen, dass die z -Transformation $A(z) = \sum_{j=0}^p a_j z^j$ keine Nullstellen z mit $|z| \leq 1$ besitzt, dass also X kausal ist. Dann hat X , wie im vorigen Kapitel diskutiert, die Darstellung

$$X_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (8.1)$$

wobei die Koeffizienten $c_j \in \mathbb{R}$ nach Bemerkung 7.8 in geometrischer Rate abklingen und durch die Potenzreihenentwicklung

$$\frac{B(z)}{A(z)} = \sum_{j=0}^{\infty} c_j z^j$$

festgelegt sind.

8.1 Die Berechnung der Autokovarianzen von ARMA-Reihen aus der MA-Darstellung

Aus (8.1) erhalten wir die folgende Darstellung für die Autokovarianz $\gamma(h) = \langle X_{t+h}, X_t \rangle = EX_{t+h}X_t$ einer ARMA-Zeitreihe (man beachte die Zentriertheit einer jeden ARMA-Zeitreihe)

$$\gamma(h) = \sigma_e^2 \sum_{j=0}^{\infty} c_j c_{j+|h|}, \quad \forall h \in \mathbb{Z}. \quad (8.2)$$

Aus (8.2) und (7.18) folgt nun unmittelbar

Bemerkung 8.1

Die Autokovarianzen $\gamma(h)$ einer kausalen ARMA-Reihe fallen in geometrischer Rate ab, d.h. es gilt

$$|\gamma(h)| \leq c \cdot \varrho^{-|h|}, \quad \forall h \in \mathbb{Z}, \quad (8.3)$$

mit geeigneten Konstanten $c \in (0, \infty)$ und $\varrho > 1$. \square

Wir wollen nun die explizite Struktur der Autokovarianzfunktion für spezielle ARMA-Zeitreihen anhand zweier Beispiele diskutieren.

Beispiel 8.2 (ARMA(1, 1)-Reihe)

Sei für $a, b \in \mathbb{R}, a \neq b, |a| < 1$

$$X_t + aX_{t-1} = e_t + be_{t-1}, \quad \forall t \in \mathbb{Z}, \quad (8.4)$$

die Gleichung einer kausalen ARMA(1, 1)-Reihe. Mit $A(z) = 1 + az, B(z) = 1 + bz$ ergibt sich für $|z| < 1/|a|$

$$\frac{B(z)}{A(z)} = \sum_{j=0}^{\infty} (-az)^j (1 + bz) = 1 + \sum_{j=1}^{\infty} (b-a)(-a)^{j-1} z^j$$

also $c_0 = 1, c_j = (b-a)(-a)^{j-1}, j \geq 1$. Aus (8.2) erhält man

$$\gamma(0) = \sigma_e^2 \sum_{j=0}^{\infty} c_j^2 = \sigma_e^2 \left(1 + \frac{(b-a)^2}{1-a^2} \right), \quad (8.5)$$

sowie

$$\gamma(h) = \sigma_e^2 \frac{(b-a)(1-ab)}{1-a^2} (-a)^{|h|-1}, \quad |h| = 1, 2, \dots \quad (8.6)$$

Setzt man $b = 0$, so erhält man die Kovarianzfunktion γ einer AR(1)-Reihe mit $X_t + aX_{t-1} = e_t, |a| < 1$, vgl. (2.17):

$$\gamma(h) = \sigma_e^2 \frac{(-a)^{|h|}}{1-a^2}, \quad \forall h \in \mathbb{Z}. \quad (8.7)$$

\square

Beispiel 8.3 (ARMA(2, 1)-Reihe)

Für $a \in \mathbb{R}$ mit $|a| > 1$ und $b \in \mathbb{R}$ mit $ab \neq -1$ sei $X = (X_t : t \in \mathbb{Z})$ Lösung von

$$X_t - \frac{2}{a}X_{t-1} + \frac{1}{a^2}X_{t-2} = e_t + be_{t-1}, \quad \forall t \in \mathbb{Z}. \quad (8.8)$$

Wegen

$$A(z) = 1 - \frac{2}{a}z + \frac{1}{a^2}z^2 = \left(1 - \frac{z}{a}\right)^2$$

besitzt also die z -Transformation A die zweifache reelle Nullstelle a . Zur Berechnung der Potenzreihenentwicklung von $B(z)/A(z)$ mit $B(z) = 1 + bz$ und der Autokovarianz $\gamma(h)$ gemäß (8.2) werden wir die Gleichung

$$(1-x)^{-(k+1)} = \sum_{j=0}^{\infty} \binom{j+k}{k} x^j, \quad |x| < 1, \quad k = 0, 1, 2, \dots, \quad (8.9)$$

wiederholt anwenden. Mit (8.9) für $k = 1$ ergibt sich

$$\begin{aligned} \frac{B(z)}{A(z)} &= \sum_{j=0}^{\infty} (j+1) \left(\frac{z}{a}\right)^j + b \sum_{j=1}^{\infty} \frac{j}{a^{j-1}} z^j \\ &= \sum_{j=0}^{\infty} \frac{j(ab+1)+1}{a^j} \cdot z^j, \end{aligned}$$

also

$$X_t = \sum_{j=0}^{\infty} c_j e_{t-j} \quad \text{mit} \quad c_j = \frac{j(ab+1)+1}{a^j}, \quad j = 0, 1, 2, \dots$$

Für $\gamma(h)$ ergibt sich aus (8.2) für $h \geq 0$ mit der Abkürzung $d := ab + 1$

$$\begin{aligned} \gamma(h) &= \sigma_e^2 \cdot \frac{1}{a^h} \sum_{j=0}^{\infty} \frac{(jd+1)((j+h)d+1)}{a^{2j}} \\ &= \sigma_e^2 \cdot \frac{1}{a^h} \left\{ \sum_{j=0}^{\infty} \frac{(jd+1)^2}{a^{2j}} + h \cdot d \sum_{j=0}^{\infty} \frac{jd+1}{a^{2j}} \right\}. \end{aligned}$$

Eine Umrechnung unter Verwendung von (8.9) für $k = 0, 1, 2$ liefert schließlich die Formel

$$\gamma(h) = \sigma_e^2 \cdot \frac{1}{a^h} (\kappa_1 + \kappa_2 h), \quad h = 0, 1, 2, \dots, \quad (8.10)$$

mit

$$\begin{cases} \kappa_1 = a^4 \frac{(a+b)^2 + (1+ab)^2}{(a^2-1)^3} \\ \kappa_2 = a^3 \frac{(ab+1)(a+b)}{(a^2-1)^2} \end{cases} \quad (8.11)$$

Man bemerkt eine große Ähnlichkeit in der Struktur der Folgen c_j , $j \geq 0$, und $\gamma(j)$, $j \geq 0$. Beide sind Linearkombinationen der Folgen a^{-j} , $j \geq 0$, und $j a^{-j}$, $j \geq 0$. Dies ist kein Zufall, sondern folgt daraus, dass beide dieselbe Differenzengleichung erfüllen. Im folgenden Abschnitt werden wir noch allgemeine Überlegungen hierzu durchführen. \square

8.2 Die Differenzengleichung für die Koeffizienten der MA-Darstellung

In diesem Abschnitt geht es um die grundsätzliche Struktur der Koeffizienten der $MA(\infty)$ -Darstellung einer $ARMA$ -Reihe. Bisher haben wir diese c_j als Koeffizienten der Potenzreihenentwicklung von B/A kennen gelernt. Hieraus können wir wie folgt eine rekursive Methode zur Berechnung der c_j ableiten, denn es gilt

$$\begin{aligned} \sum_{j=0}^q b_j z^j &= \sum_{k=0}^p a_k z^k \cdot \sum_{r=0}^{\infty} c_r z^r \\ &= \sum_{r=0}^{\infty} \sum_{k=0}^p c_r a_k z^{r+k} \\ &= \sum_{j=0}^{\infty} (c_0 a_j + c_1 a_{j-1} + \cdots + c_j a_0) z^j, \end{aligned} \quad (8.12)$$

wobei $a_j := 0$ für $j > p$ gesetzt wird. Setzt man entsprechend $b_j := 0$ für $j > q$, so erhält man durch Koeffizientenvergleich die Gleichung

$$b_j = \sum_{k=0}^j c_{j-k} a_k, \quad j = 0, 1, 2, \dots, \quad (8.13)$$

oder als Rekursionsformel geschrieben

$$c_j = b_j - \sum_{k=1}^j c_{j-k} a_k, \quad j \geq 1, \quad c_0 = 1, \quad (8.14)$$

also

$$c_1 = b_1 - a_1, \quad c_2 = b_2 - (c_1 a_1 + a_2) = b_2 - b_1 a_1 + a_1^2 - a_2, \dots$$

Über die reine Berechnung der c_j hinaus kann man aus den Gleichungen (8.13) auch die prinzipielle Struktur der Lösungen $\{c_j\}$ gewinnen. Dazu spalten wir (8.13) in zwei Gruppen von Gleichungen auf.

Nebenbedingungen:

$$\sum_{k=0}^j c_{j-k} a_k = b_j \quad \text{für } j = 0, \dots, \max(p-1, q). \quad (8.15)$$

Homogene Gleichungen:

$$\sum_{k=0}^p c_{j-k} a_k = 0 \quad \text{für } j = \max(p, q+1), \dots \quad (8.16)$$

Mit $d_j := c_{j+\max(p, q+1)-p}$ wird (8.16) zu

$$\sum_{k=0}^p d_{j-k} a_k = 0 \quad \text{für } j \geq p. \quad (8.17)$$

Die Gesamtheit V aller komplexen Lösungsfolgen $\{d_j\} = \{d_j : j = 0, 1, \dots\}$ für (8.17) bei festen Werten $a_k \in \mathbb{R}, a_0 = 1, a_p \neq 0$, bildet mit den üblichen, komponentenweise definierten Operationen einen \mathbb{C} -Vektorraum der Dimension p , da sich zu beliebig vorgegebenen Werten d_0, \dots, d_{p-1} die restlichen Werte d_p, d_{p+1}, \dots rekursiv durch (8.17) bestimmen lassen. Ist z eine $(n+1)$ -fache Nullstelle von $A(z) = \sum_{k=0}^p a_k z^k$, so gilt $z \neq 0$ und die $n+1$ Folgen

$$\{j^r z^{-j} : j = 0, 1, 2, \dots\}, \quad r = 0, \dots, n, \quad (8.18)$$

sind Lösungen von (8.17), die sogenannten *Basislösungen*, von denen es offenbar genau p gibt, wenn man alle Nullstellen von A mit ihren Vielfachheiten durchläuft. Die p Basislösungen bilden eine Basis von V , sodass sich die allgemeine Lösung als eindeutig bestimmte Linearkombination von Basislösungen schreiben lässt, siehe Anhang A.4.

Im Hinblick auf unsere Anwendungen interessieren uns nur reelle Lösungsfolgen. In diesem Fall lässt sich zeigen, dass die allgemeine reelle Lösung $\{d_j\}$ sich als Summe von Lösungen der beiden folgenden Typen schreiben lässt.

1. Typ: ($A(z) = 0, z$ reell)

$$d_j = \kappa j^r z^{-j}, \quad j = 0, 1, 2, \dots, \quad (8.19)$$

mit $\kappa \in \mathbb{R}, r = 0, \dots, n$, wobei $n+1$ die Vielfachheit der Nullstelle z ist.

2. Typ: ($A(z) = 0, z$ nicht reell)

$$d_j = \varrho j^r |z|^{-j} \cos(j\varphi + \vartheta), \quad j = 0, 1, 2, \dots, \quad (8.20)$$

mit $\varrho > 0, \varphi, \vartheta \in (-\pi, \pi], \varphi \notin \{0, \pi\}, r = 0, \dots, n$, wobei $n+1$ die Vielfachheit der Nullstelle $z = |z| e^{i\varphi}$ ist.

Die Form (8.20) ergibt sich durch Zusammenfassung der Basislösungen für z und \bar{z} , siehe Anhang A.4.

Hat A nur Nullstellen außerhalb des Einheitskreises, so gilt in beiden obigen Fällen $|z| > 1$. Dann ist (8.19) eine mit geometrischer Rate abfallende Folge und (8.20) eine mit geometrischer Rate gedämpfte Kosinusschwingung. Die Gesamtlösung $\{d_j\}$ setzt sich wie gesagt aus Summanden der beiden Typen zusammen, wobei die Werte von φ gemäß $z = |z|e^{i\varphi}$ durch die Nullstellen z bestimmt werden und die Werte $\kappa, \varrho, \vartheta$ durch d_0, \dots, d_{p-1} festgelegt sind.

8.3 Die Differenzengleichung für die Autokovarianzen und die Yule–Walker–Gleichungen bei AR-Reihen

Ausgangspunkt unserer Beschäftigung mit der Differenzengleichung (8.17) waren die Gleichungen (8.15) und (8.16) zur Berechnung der Koeffizienten der Potenzreihenentwicklung von B/A , über die dann mit Formel (8.2) die Autokovarianzen $\gamma(\cdot)$ berechnet werden können. Nun werden wir gleich sehen, dass die Folge $\gamma(j), j \geq 0$, der Autokovarianzen ebenfalls die homogenen Gleichungen (8.16) mit c_j ersetzt durch $\gamma(j)$ erfüllt. Allerdings gibt es ein anderes System von Nebenbedingungen als (8.15). Dazu beachten wir, dass für alle $j \in \mathbb{Z}$ aus der ARMA-Gleichung (7.1) und ihrer MA-Darstellung (8.1) folgt

$$\begin{aligned} \left\langle \sum_{i=0}^p a_i X_{t-i}, X_{t-j} \right\rangle &= \left\langle \sum_{i=0}^q b_i e_{t-i}, X_{t-j} \right\rangle \\ &= \sum_{i=0}^q b_i \left\langle e_{t-i}, \sum_{k=0}^{\infty} c_k e_{t-j-k} \right\rangle \\ &= \sum_{i=0}^q b_i c_{i-j} \sigma_e^2, \end{aligned}$$

mit $c_k := 0$ für $k < 0$; hierbei wird die Kausalitätsannahme benutzt. Wegen $\langle X_{t-i}, X_{t-j} \rangle = \gamma(j-i)$ ergeben sich analog zu (8.15) und (8.16) für $\gamma(j), j = 0, 1, \dots$, die folgenden Beziehungen

$$\sum_{k=0}^p \gamma(j-k) a_k = \sum_{k=0}^q c_{k-j} b_k \sigma_e^2, \quad j = 0, 1, 2, \dots, \quad (8.21)$$

wobei die rechte Seite für $j \geq q+1$ verschwindet.

Im Fall der Zeitreihe aus (8.8) ergibt sich aus (8.21)

Beispiel 8.4 (Fortsetzung von Beispiel 8.3, ARMA(2,1)-Reihe)

$$\begin{aligned} \gamma(0) - \gamma(1) \frac{2}{a} + \gamma(2) \frac{1}{a^2} &= \sigma_e^2 (1 + b(b + 2/a)) & (j = 0) \\ \gamma(1) - \gamma(0) \frac{2}{a} + \gamma(1) \frac{1}{a^2} &= \sigma_e^2 b & (j = 1) \\ \gamma(j) - \gamma(j-1) \frac{2}{a} + \gamma(j-2) \frac{1}{a^2} &= 0 & (j \geq 2) \end{aligned} \quad (8.22)$$

Die allgemeine Form von $\gamma(j)$ ist, da a zweifache Nullstelle von $A(z)$ ist,

$$\gamma(j) = \sigma_e^2 (\kappa_1 + \kappa_2 j) a^{-j}, \quad j = 0, 1, 2, \dots, \quad (8.23)$$

mit geeigneten reellen Konstanten κ_1 und κ_2 .

Einsetzen von (8.23) in (8.22) für $j = 0$ und $j = 1$ liefert die Gleichungen

$$\begin{aligned} \kappa_1 - 2\kappa_2(a^2 - 1)^{-1} &= (a^4/(a^2 - 1)^2)(1 + b(b + 2/a)) \\ -\kappa_1 + \kappa_2(a^2 + 1)/(a^2 - 1) &= ba^3/(a^2 - 1). \end{aligned}$$

Man rechnet sofort nach, dass die Lösung (κ_1, κ_2) dieses Gleichungssystems erneut (8.11) ergibt. \square

Die Beziehungen (8.21) können auch dazu verwendet werden, bei *gegebenen* Werten der $\gamma(j)$, z.B. insbesondere bei vorliegenden Schätzwerten, die zugehörigen Parameter a_1, \dots, a_p und b_1, \dots, b_q sowie σ_e^2 zu berechnen. Diese Vorgehensweise ist besonders für $AR(p)$ -Reihen geeignet, also für den Fall $q = 0$, den wir nun voraussetzen wollen. Mit den Bezeichnungen

$$\Gamma_p := (\gamma(i - j))_{i,j=1,\dots,p}, \quad \gamma_p := (\gamma(1), \dots, \gamma(p))^T \quad (8.24)$$

sowie $\mathbf{a} = (a_1, \dots, a_p)^T$ ergibt (8.21) für $j = 0, 1, \dots, p$ die *Yule-Walker-Gleichungen*

$$\gamma(0) + \mathbf{a}^T \gamma_p = \sigma_e^2 \quad (j = 0) \quad (8.25)$$

und für $j = 1, \dots, p$ in vektorieller Form geschrieben

$$\Gamma_p \mathbf{a} + \gamma_p = 0. \quad (8.26)$$

Wie schon bemerkt, können (8.25) und (8.26) dazu verwendet werden, um aus Schätzwerten $\hat{\gamma}(0), \dots, \hat{\gamma}(p)$ der Autokovarianzen die Varianz σ_e^2 und die Parameterwerte a_1, \dots, a_p der $AR(p)$ -Reihe zu schätzen.

Ein naheliegender Schätzer $\hat{\gamma}(h)$ für $\gamma(h)$ bei Vorliegen von Beobachtungen X_1, \dots, X_n ist die *Stichprobenautokovarianz*, vgl. (2.29).

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \bar{X}_n) (X_{t+|h|} - \bar{X}_n), \quad |h| \leq n-1, \quad (8.27)$$

mit dem *Stichprobenmittel* $\bar{X}_n = (X_1 + \dots + X_n)/n$. Setzt man noch $\hat{\gamma}(h) = 0$ für $|h| \geq n$, so lässt sich leicht zeigen (Aufgabe 8.3), dass für jede Folge von Realisierungen der X_1, \dots, X_n die Funktion $\hat{\gamma}: \mathbb{Z} \rightarrow \mathbb{R}$ positiv semidefinit ist. Wenn nicht alle Realisierungen gleich sind, so gilt offenbar $\hat{\gamma}(0) > 0$. Das folgende Resultat liefert dann eindeutig bestimmte Lösungen $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_p)$ und $\hat{\sigma}^2$ für (8.25) und (8.26) (mit γ ersetzt durch $\hat{\gamma}$). Man spricht dann von *Yule-Walker-Schätzern* $\hat{a}_1, \dots, \hat{a}_p$ und $\hat{\sigma}^2$.

Satz 8.5

Sei $\gamma: \mathbb{Z} \rightarrow \mathbb{R}$ eine reelle Kovarianzfunktion mit $\gamma(h) \rightarrow 0$ für $h \rightarrow \infty$, sowie $\gamma(0) > 0$. Dann ist $\Gamma_n := (\gamma(i-j))_{i,j=1,\dots,n}$ nicht ausgeartet für jedes $n \geq 1$.

Beweis: Sei $X = (X_t: t \in \mathbb{Z})$ ein reeller, stationärer, zentrierter Prozess mit γ als Kovarianzfunktion. Ist Γ_n für ein n ausgeartet, so gibt es ein maximales $r \geq 1$ derart, dass Γ_r nicht ausgeartet ist. Es existiert dann ein Vektor $\mathbf{c} = (c_1, \dots, c_{r+1})^T \in \mathbb{R}^{r+1}$, $c_{r+1} = 1$, mit

$$\left\| \sum_{j=1}^{r+1} c_j X_{t+j} \right\|^2 = \mathbf{c}^T \Gamma_{r+1} \mathbf{c} = 0, \quad \forall t \in \mathbb{Z}.$$

Insbesondere gilt $X_t \in \text{span}\{X_{t-1}, \dots, X_{t-r}\}$, $\forall t \in \mathbb{Z}$, also iterativ auch $X_t \in \text{span}\{X_1, \dots, X_r\}$, $\forall t \geq 1$. Für einen Vektor $\mathbf{a}_t \in \mathbb{R}^r$ gilt demnach mit $\mathbf{X} = (X_1, \dots, X_r)^T$ die Gleichung $X_t = \mathbf{a}_t^T \mathbf{X}$ $\forall t \geq 1$. Multiplikation mit X_j , $j = 1, \dots, r$, und Erwartungswertbildung liefert für $\gamma_t = (\gamma(t-1), \dots, \gamma(t-r))^T$ die Gleichung $\gamma_t^T = \mathbf{a}_t^T \Gamma_r$, also $X_t = \mathbf{a}_t^T \mathbf{X} = \gamma_t^T \Gamma_r^{-1} \mathbf{X}$ und damit schließlich

$$\gamma(0) = \|X_t\|^2 = \gamma_t^T \Gamma_r^{-1} \Gamma_r \Gamma_r^{-1} \gamma_t = \gamma_t^T \Gamma_r^{-1} \gamma_t, \quad \forall t \geq 1.$$

Wegen $\gamma_t \rightarrow 0$ für $t \rightarrow \infty$ und $\gamma(0) > 0$ ergibt sich ein Widerspruch. □

Beispiel 8.6 (Sonnenfleckendaten)

Aus den Daten der Sonnenflecktätigkeit, vergleiche Abbildung 2.11, ergaben sich in Kapitel 2 für die geschätzten Autokorrelationen $\hat{\varrho}_h := \hat{\gamma}(h)/\hat{\gamma}(0)$ die Werte

$$\hat{\varrho}_1 = 0.82043, \quad \hat{\varrho}_2 = 0.45021.$$

Wir wollen zu diesen Werten die Parameter \hat{a}_1, \hat{a}_2 , einer $AR(2)$ -Reihe berechnen. Die Division von (8.21) durch $\hat{\gamma}(0)$ liefert zunächst entsprechende Gleichungen für die Autokorrelation $\varrho_h = \gamma(h)/\gamma(0)$. Wir ersetzen dabei gleich ϱ_h durch die Schätzwerte $\hat{\varrho}_h$ und erhalten

$$\begin{aligned} 1 + \hat{a}_1 \hat{\varrho}_1 + \hat{a}_2 \hat{\varrho}_2 &= \hat{\sigma}_e^2 / \hat{\gamma}(0) & (j=0) \\ \hat{\varrho}_1 + \hat{a}_1 + \hat{a}_2 \hat{\varrho}_1 &= 0 & (j=1) \\ \hat{\varrho}_2 + \hat{a}_1 \hat{\varrho}_1 + \hat{a}_2 &= 0 & (j=2). \end{aligned}$$

Aus den Gleichungen für $j=1$ und $j=2$ ergibt sich

$$\begin{aligned} \hat{a}_2 &= (\hat{\varrho}_1^2 - \hat{\varrho}_2) / (1 - \hat{\varrho}_1^2) = 0.682 \\ \hat{a}_1 &= -\hat{\varrho}_1 (1 + \hat{a}_2) = -1.380. \end{aligned}$$

Unter Verwendung von $\hat{\gamma}(0) = 1634.08$ erhält man einen Schätzwert für σ_e^2 gemäß

$$\hat{\sigma}_e^2 = \hat{\gamma}(0) (1 + \hat{a}_1 \hat{\varrho}_1 + \hat{a}_2 \hat{\varrho}_2) = 285.72 = 16.90^2.$$

Die zugehörige z -Transformation des geschätzten $AR(2)$ -Modells lautet

$$A(z) = 1 + \hat{a}_1 z + \hat{a}_2 z^2 = 1 - 1.380 \cdot z + 0.682 \cdot z^2$$

und besitzt die komplexen Nullstellen $z_{1/2} = 1.012 \pm 0.665 \, i$ bzw. in Polarkoordinaten-Darstellung $z_{1/2} = 1.211 e^{\pm 0.581 \, i}$. Die Tatsache, dass die z -Transformation lediglich zwei komplexe Nullstellen hat, führt uns sowohl bei der Berechnung der Koeffizienten der $MA(\infty)$ -Darstellung wie der Autokorrelationsfunktion zum 2. Typ (vgl. (8.20)) von Lösungen der zugehörigen Differenzengleichung. Die graphische Darstellung der Autokorrelationsfunktion des geschätzten $AR(2)$ -Modells mit Parametern \hat{a}_1 und \hat{a}_2 in Abbildung 8.1 zeigt demgemäß das Verhalten einer geometrisch gedämpften Kosinusschwingung.

Aus der Polarkoordinatendarstellung erkennt man sofort, dass die Nullstellen

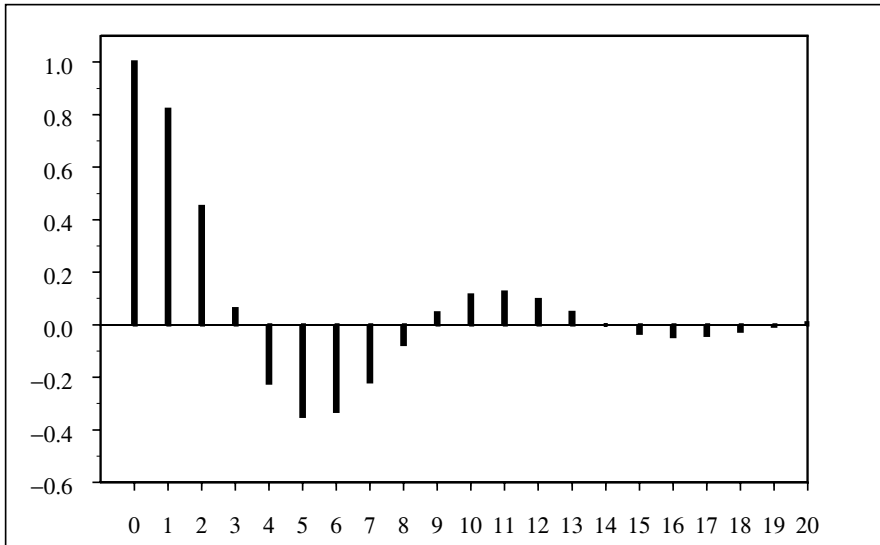


Abb. 8.1 Autokorrelationsfunktion des aus den Sonnenfleckenzahlen geschätzten $AR(2)$ -Modells $X_t - 1.380X_{t-1} + 0.682X_{t-2} = e_t$

$z_{1/2}$ der z -Transformation A betragsmäßig größer als 1 sind. Die geschätzte $AR(2)$ -Reihe mit den Koeffizienten \hat{a}_1, \hat{a}_2 ist also kausal. Dass dies kein glücklicher Zufall ist, sondern bei Verwendung des Yule-Walker-Schätzers stets gilt, besagt der folgende Satz. \square

Satz 8.7

Sei $\gamma(h)$, $h \in \mathbb{N}_0$, eine reelle Kovarianzfunktion derart, dass die Kovarianzmatrix $\Gamma_{p+1} = (\gamma(|i-j|))_{i,j=1,\dots,p+1}$ positiv definit ist für ein festes $p \in \mathbb{N}$. Mit $\gamma_p := (\gamma(1), \dots, \gamma(p))^T$ sei $\mathbf{a} := (a_1, \dots, a_p)^T$ Lösung der Yule-Walker-Gleichung

$$\Gamma_p \mathbf{a} + \gamma_p = 0.$$

Dann hat die z -Transformation $A(z) = 1 + \sum_{j=1}^p a_j z^j$ nur Nullstellen außerhalb des Einheitskreises.

Beweis: Sei $Y = (Y_t : t \in \mathbb{Z})$ eine reelle stationäre Zeitreihe mit Autokovarianzfunktion $\gamma(\cdot)$. Da wir es im Beweisverlauf auch mit komplexwertigen Zufallsvariablen zu tun haben werden, fassen wir dabei alle vorkommenden Zufallsvariablen als Elemente des komplexen $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ auf mit dem komplexen inneren Produkt $\langle \cdot, \cdot \rangle$ und zugehöriger Norm $\| \cdot \|$. Wir werden jetzt zeigen, dass

$$\hat{Y}_{p+1} := -(a_1 Y_p + \dots + a_p Y_1)$$

die Projektion von Y_{p+1} auf $\text{span}\{Y_1, \dots, Y_p\}$ ist. Dazu ist zu zeigen, dass $Y_{p+1} - \hat{Y}_{p+1}$ senkrecht auf allen Y_j , $j = 1, \dots, p$, steht. Nun gilt aber

$$\langle Y_{p+1} + \sum_{k=1}^p a_k Y_{p+1-k}, Y_j \rangle = \gamma(p+1-j) + \sum_{k=1}^p a_k \gamma(p+1-k-j),$$

für $j = 1, \dots, p$, was gerade der komponentenweise geschriebene Ausdruck $\Gamma_p \mathbf{a} + \gamma_p$ ist, der nach Voraussetzung verschwindet.

Sei $1/a \in \mathbb{C}$ eine Nullstelle von A . Dann hat $A(z)$ die Darstellung

$$A(z) = (1 - az) A_0(z)$$

mit einem geeignetem Polynom $A_0(z) = 1 + \sum_{k=1}^{p-1} \alpha_k z^k$. Wir zeigen nun die Gleichheit

$$a = \varrho := \frac{\langle Z_{p+1}, Z_p \rangle}{\|Z_{p+1}\| \|Z_p\|} \quad (8.28)$$

mit $Z_j := Y_j + \sum_{k=1}^{p-1} \alpha_k Y_{j-k}$. Dazu definieren wir das Polynom

$$\tilde{A}(z) = 1 + \sum_{k=1}^p \tilde{a}_k z^k := (1 - \varrho z) A_0(z)$$

und erhalten mit dem Lag-Operator L gemäß $LY_t = Y_{t-1}$

$$\begin{aligned}
E \left| Y_{p+1} + \sum_{k=1}^p \tilde{a}_k Y_{p+1-k} \right|^2 &= E |\tilde{A}(L) Y_{p+1}|^2 \\
&= E |(1 - \varrho L) A_0(L) Y_{p+1}|^2 = E |(1 - \varrho L) Z_{p+1}|^2 \\
&= E |Z_{p+1} - \varrho Z_p|^2.
\end{aligned} \tag{8.29}$$

Eine entsprechende Rechnung für $A(z)$ liefert

$$E \left| Y_{p+1} + \sum_{k=1}^p a_k Y_{p+1-k} \right|^2 = E |Z_{p+1} - a Z_p|^2. \tag{8.30}$$

Die Projektionseigenschaft von \hat{Y}_{p+1} besagt nun, dass die linke Seite von (8.30) kleiner oder gleich der linken Seite von (8.29) ist. Da wegen der Stationarität von Y auch $\|Z_{p+1}\| = \|Z_p\|$ gilt, ist $\varrho Z_p = (Z_p / \|Z_p\|) \cdot \langle Z_{p+1}, Z_p / \|Z_p\| \rangle$ die Projektion von Z_{p+1} auf $\text{span}\{Z_p\}$, woraus nun folgt, dass die rechte Seite von (8.30) größer oder gleich der rechten Seite von (8.29) ist. Beide Ausdrücke (8.29) und (8.30) sind also gleich, woraus wegen der Eindeutigkeit der Projektion $\varrho = a$ folgt.

Aus der Cauchy-Schwarz-Ungleichung folgt unmittelbar, dass $|\varrho| \leq 1$ gilt. Wäre $|\varrho| = 1$, so erhielte man nach einfacher Rechnung $\|Z_{p+1} - \varrho Z_p\|^2 = 0$, also $Z_{p+1} = \varrho Z_p$. Einsetzen in die Definitionsgleichung für Z_p bzw. Z_{p+1} führt zu einer Darstellung $Y_{p+1} = \sum_{k=1}^p b_k Y_k$ für geeignete Koeffizienten $b_1, \dots, b_p \in \mathbb{C}$. Hieraus folgt für $\mathbf{b} := (1, -b_1, \dots, -b_p)^T$ die Gleichung $\mathbf{b}^T \Gamma_{p+1} \mathbf{b} = 0$ im Widerspruch zur positiven Definitheit von Γ_{p+1} , vergleiche Bemerkung 3.3. Insgesamt gilt für die Nullstelle $1/a$ von A also $|1/a| > 1$. \square

Bemerkung 8.8

Für die Anwendung wird der Satz 8.7 besonders wichtig, wenn wir dort für γ eine geschätzte Autokovarianzfunktion verwenden und die daraus resultierenden Yule-Walker-Parameterschätzer im Auge haben. Wenn wir sicherstellen, dass unser Schätzer der Autokovarianzfunktion selbst die Eigenschaften einer Autokovarianzfunktion gemäß Satz 3.1 trägt (was wir in Aufgabe 8.3 für die Stichprobenautokovarianz nachgewiesen haben!) und die Matrix Γ_{p+1} positiv definit ist (was im konkreten Einzelfall schnell überprüft werden kann), dann ist garantiert, dass das geschätzte autoregressive Modell der Ordnung p stets stationär und auch kausal ist. Diese Tatsache ist für die reale Anwendung von großem Nutzen und ist sicher mit ein Grund für die große Bedeutung autoregressiver Zeitreihenmodelle. \square

8.4 Identifizierbarkeit der Parameter von ARMA-Zeitreihen

Im vorangegangenen Abschnitt haben wir bereits den Yule-Walker-Schätzer für autoregressive Zeitreihenmodelle verwendet (siehe Beispiel 8.6). In Kapitel

11 werden wir eingehend Schätzverfahren für die Parameter in $ARMA(p, q)$ -Modellen untersuchen. Bevor man sich sinnvoll mit Schätzverfahren auseinander setzen kann, ist die grundlegende Frage zu klären, inwieweit die Parameter von $AR(p)$ - und genauso von $ARMA(p, q)$ -Modellen überhaupt eindeutig bestimmt sind. Wir wollen diesbezügliche Überlegungen jetzt anstellen.

Betrachten wir dazu zunächst $AR(p)$ -Zeitreihen $X = (X_t : t \in \mathbb{Z})$. Die Frage ist, ob X möglicherweise für zwei verschiedene Parametersätze $p, a_1, \dots, a_p, \sigma_e^2$ und $p^*, a_1^*, \dots, a_{p^*}^*, \sigma_{e^*}^2$ jeweils einem autoregressiven Schema genügen kann. Der folgende Satz besagt, dass dies für kausale und stationäre Autoregressionen nicht möglich ist.

Satz 8.9 (*Identifizierbarkeit autoregressiver Modelle*)

Die Autokorrektionsfunktion einer kausalen AR-Zeitreihe $X = (X_t : t \in \mathbb{Z})$ bestimmt sowohl die Ordnung p als auch die Koeffizienten a_1, \dots, a_p eindeutig. Damit ist offensichtlich auch das weiße Rauschen und insbesondere dessen Varianz σ_e^2 festgelegt.

Beweis: Zunächst gilt für $ARMA$ -Zeitreihen stets $\gamma(0) > 0$, denn $\gamma(0) = 0$ führt wegen $EX_t = 0$ zu $X_t = 0$ f.s. und gemäß (7.1) zu dem Widerspruch $0 = \text{Var}(e_t + \sum_{j=1}^q b_j e_{t-j}) = \sigma_e^2 \cdot (1 + \sum_{j=1}^q b_j^2)$. Mit ϱ bezeichnen wir die zugrunde liegende Autokorrelationsfunktion. Da $\varrho(h) \rightarrow 0$ für $h \rightarrow \infty$ (vergleiche Bemerkung 8.1) folgt wegen Satz 8.5, dass die Autokorrelationsmatrix $(\varrho(r-s) : r, s = 1, \dots, p)$ für alle $p \in \mathbb{N}$ nicht ausgeartet ist.

Nehmen wir an, dass $X = (X_t : t \in \mathbb{Z})$ sowohl eine $AR(p)$ -Darstellung mit Koeffizienten a_1, \dots, a_p als auch eine $AR(p^*)$ -Darstellung mit Koeffizienten $a_1^*, \dots, a_{p^*}^*$ erfüllt. Hierbei gilt notwendigerweise $a_p, a_{p^*}^* \neq 0$.

Ohne Einschränkung sei $p \leq p^*$. Dann gelten die beiden folgenden Yule-Walker-Gleichungen (vergleiche (8.26))

$$\varrho(h) + \sum_{j=1}^p a_j \varrho(h-j) = 0, \quad h = 1, 2, \dots \quad (8.31)$$

und

$$\varrho(h) + \sum_{j=1}^{p^*} a_j^* \varrho(h-j) = 0, \quad h = 1, 2, \dots \quad (8.32)$$

Das lineare Gleichungssystem

$$(\varrho(1), \dots, \varrho(p^*))^T = \begin{pmatrix} \varrho(r-s) \\ r, s = 1, \dots, p^* \end{pmatrix} \mathbf{c}$$

besitzt also mindestens die beiden Lösungen

$$\mathbf{c}_1 = -(a_1^*, \dots, a_{p^*}^*)^T \quad \text{und} \quad \mathbf{c}_2 = -(a_1, \dots, a_p, 0, \dots, 0)^T,$$

was im Fall $p^* > p$ wegen der Invertierbarkeit der Autokorrelationsmatrix und der daraus resultierenden Eindeutigkeit der Lösungen des obigen Gleichungssystems zum Widerspruch $a_{p^*}^* \neq 0$ und im Fall $p = p^*$ zu $a_j = a_j^*$ für $j = 1, \dots, p (= p^*)$ führt. \square

Wir wollen dieses Resultat nun auf ARMA-Zeitreihen erweitern. Es gilt

Satz 8.10 (*Identifizierbarkeit von ARMA-Modellen*)

$X = (X_t : t \in \mathbb{Z})$ sei eine ARMA-Zeitreihe. Die aus den Parametern gebildeten Polynome $A(z) = 1 + \sum_{j=1}^p a_j z^j$ und $B(z) = 1 + \sum_{k=1}^q b_k z^k$ mögen keine gemeinsamen Nullstellen z und nur solche mit $|z| > 1$ besitzen. Letzteres ist gerade die Forderung, dass X kausal und invertibel ist. In dieser Modellklasse sind sowohl die Ordnungen p und q sowie die Koeffizienten a_1, \dots, a_p und b_1, \dots, b_q (und mit Hilfe der Invertibilität auch das weiße Rauschen) eindeutig bestimmt.

Beweis: Gemäß (7.15) lautet die Spektraldichte von $X = (X_t : t \in \mathbb{Z})$

$$f_X(\omega) = \frac{\sigma_e^2 |B(e^{-i\omega})|^2}{2\pi |A(e^{-i\omega})|^2}, \quad \omega \in [-\pi, \pi]. \quad (8.33)$$

X genüge nun einer weiteren ARMA(p^*, q^*)-Gleichung mit einem Parametersatz $a_1^*, \dots, a_{p^*}^*, b_1, \dots, b_{q^*}$ und $\sigma_{e^*}^2$, wobei die entsprechenden Polynome $A^*(z)$ und $B^*(z)$ ebenfalls keine gemeinsamen Nullstellen z und nur solche mit $|z| > 1$ besitzen. Genauso wie eben erhalten wir die weitere Darstellung der Spektraldichte

$$f_X(\omega) = \frac{\sigma_{e^*}^2 |B^*(e^{-i\omega})|^2}{2\pi |A^*(e^{-i\omega})|^2}, \quad \omega \in [-\pi, \pi]. \quad (8.34)$$

Nun betrachten wir die folgende kausale AR-Gleichung

$$A(L)B^*(L)Y_t = \eta_t, \quad \forall t \in \mathbb{Z}, \quad (8.35)$$

für ein weißes Rauschen $\eta = (\eta_t : t \in \mathbb{Z})$ mit $\sigma_\eta^2 = 1$. Erneut nach Satz 7.6 lautet die Spektraldichte der eindeutig bestimmten Lösung $Y = (Y_t : t \in \mathbb{Z})$

$$f_Y(\omega) = \frac{1}{2\pi} |A^*(e^{-i\omega})B^*(e^{-i\omega})|^{-2}, \quad \omega \in [-\pi, \pi]. \quad (8.36)$$

Nun wenden wir den Filter $A^*(L)B(L)$ auf Y an und erhalten

$$\eta_t^* := A^*(L)B(L)Y_t, \quad \forall t \in \mathbb{Z}, \quad (8.37)$$

wobei $\eta^* = (\eta_t^* : t \in \mathbb{Z})$ aufgrund von (6.5) die Spektraldichte

$$\frac{1}{2\pi} \frac{|A^*(e^{-i\omega})B(e^{-i\omega})|^2}{|A(e^{-i\omega})B^*(e^{-i\omega})|^2}, \quad \omega \in [-\pi, \pi],$$

besitzt. Wegen (8.33) und (8.34) ist dieser Ausdruck gleich $(2\pi)^{-1}\sigma_{e^*}^2/\sigma_e^2$, so dass η^* ebenfalls ein weißes Rauschen darstellt.

Y erfüllt also sowohl die autoregressive Gleichung (8.35) als auch (8.37). Da beide Darstellungen kausal sind, folgt mit der Eindeutigkeit der Parameter aus Satz 8.9

$$A(z)B^*(z) = A^*(z)B(z) \quad \forall z \in \mathbb{C} \text{ und } \sigma_\eta^2 = \sigma_{\eta^*}^2.$$

Da A und B sowie A^* und B^* jeweils keine gemeinsamen Nullstellen besitzen folgt, dass sowohl die Nullstellen der Polynome A und A^* als auch die Nullstellen der Polynome B und B^* komplett übereinstimmen müssen. Hieraus folgt, dass sich die Polynome A und A^* (und genauso B und B^*) jeweils nur um einen konstanten Faktor unterscheiden können. Wegen $A(0) = A^*(0) = B(0) = B^*(0) = 1$ folgt jedoch, dass dieser Faktor gleich eins sein muss. Damit ergibt sich schließlich $A^* = A$ und $B^* = B$, woraus wie behauptet $p = p^*, q = q^*, a_j = a_j^* (j = 1, \dots, p)$ und $b_k = b_k^* (k = 1, \dots, q)$ folgt. \square

Aufgaben

Aufgabe 8.1 Gegeben sei ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $e_t \sim (0, 1)$. Man bestimme die Autokovarianzfunktion der AR(3)-Zeitreihe

$$X_t - 0.6X_{t-1} + 0.11X_{t-2} - 0.006X_{t-3} = e_t.$$

Hinweis: Für das Aufstellen der allgemeinen Lösung der Differenzengleichung für die Autokovarianzen ist die Zerlegung $(1-0.1L)(1-0.2L)(1-0.3L)X_t = e_t$ hilfreich.

Aufgabe 8.2 Man löse die folgenden homogenen linearen Differenzengleichungen unter Beachtung der Anfangsbedingungen:

a)

$$d_j - \frac{3}{2}d_{j-1} + \frac{9}{16}d_{j-2} - \frac{1}{16}d_{j-3} = 0, \quad j = 3, 4, \dots$$

$$d_0 = 19, d_1 = 9, d_2 = 5.$$

b)

$$d_j - \frac{8}{25}d_{j-1} + \frac{1}{25}d_{j-2} = 0, \quad j = 2, 3, \dots$$

$$d_0 = 5, d_1 = 1.$$

Hinweis: Aufgabenteil b) führt zunächst auf komplexe Lösungen. Man stelle die zugehörigen reellen Lösungen gemäß Gleichung (A.50) auf und bestimme

die dortigen Parameter ρ und θ auf numerischem Wege gemäß der Anfangsbedingungen. Man überprüfe die (expliziten) Lösungen in a) und b) durch direkte (rekursive) Berechnung der ersten Folgenglieder d_j .

Aufgabe 8.3 Seien x_1, \dots, x_n reelle Zahlen und

$$g(h) := \sum_{t=1}^{n-h} x_t x_{t+h}, \quad h = 0, 1, 2, n-1,$$

und $g(h) = 0$ für $h \geq n$, sowie $g(-h) := g(h)$. Für $G(p) := (g(|r-s|) : r, s = 1, \dots, p)$, $p \in \mathbb{N}$, zeige man die Darstellung

$$G(p) = A^T A$$

für eine geeignete $(n+p-1) \times p$ -Matrix A . Hieraus folgere man die positive Semidefinitheit von $G(p)$ und zeige darüber hinaus, dass $G(p)$ nicht ausgeartet ist, sobald nicht alle x_t verschwinden. Insbesondere folgt, dass der Autokovarianzschätzer $\hat{\gamma}_X$ positiv semidefinit ist.

Aufgabe 8.4 Gegeben sei eine reelle, stationäre, zentrierte Zeitreihe $X = (X(t) : t \in \mathbb{Z})$ mit Kovarianzfunktion $\gamma(t) \rightarrow 0$ für $t \rightarrow \infty$. Man definiere die partielle Autokorrelationsfunktion $\pi(k)$ wie in Abschnitt 2.5. Für festes t definiere man

$$X^*(t+n+1) := \sum_{i=1}^n a_{ni} X(t+n+1-i); \quad X^*(t) := \sum_{j=1}^n b_{nj} X(t+j),$$

wobei die Koeffizienten a_{ni}, b_{nj} zu bestimmen sind durch Minimierung der Zielfunktion $Z_1 = E(X(t+n+1) - X^*(t+n+1))^2$, $Z_2 = E(X(t) - X^*(t))^2$.

a) Man zeige durch Differentiation der Zielfunktionen Z_1, Z_2 nach a_{ni} , bzw. b_{nj} , dass $\mathbf{a}_n = (a_{n1}, \dots, a_{nn})^T$ Lösung des Gleichungssystems

$$\Gamma_n \cdot \mathbf{a}_n = (\gamma(1), \dots, \gamma(n))^T$$

ist und $a_{ni} = b_{ni} \forall i$ gilt. Dabei ist Γ_n die $n \times n$ -Kovarianzmatrix.

b) Mit a) leite man direkt aus der Definition der partiellen Autokorrelation die folgende Gleichung her:

$$\pi(n+1) = \frac{\gamma(n+1) - \sum_{i=1}^n a_{ni} \gamma(n+1-i)}{\gamma(0) - \sum_{i=1}^n a_{ni} \gamma(i)}, \quad n = 1, 2, \dots$$

c) Man folgere aus (2.39) mit Hilfe der Cramerschen Regel für die partielle Autokorrelationsfunktion $\pi(n)$:

$$\pi(n) = \frac{\det \hat{\Gamma}_n}{\det \Gamma_n},$$

wobei $\hat{\Gamma}_n$ die Matrix bezeichnet, die aus Γ_n durch Ersetzen der letzten Spalte durch $(\gamma(1), \dots, \gamma(n))^T$ entsteht.

Aufgabe 8.5 Gegeben sei eine reelle $MA(1)$ -Reihe $X_t = e_t + be_{t-1}$, $\forall t \in \mathbb{Z}$, $b \in (-1, 1)$ mit weißem Rauschen $e_t \sim (0, \sigma_e^2)$. Man zeige für die partielle Autokorrelation

$$\pi(n) = \frac{-(-b)^n(1-b^2)}{1-b^{2(n+1)}}.$$

Hinweis: Man verwende Teil c) der vorigen Aufgabe. Zur Berechnung von $\det \Gamma_n =: c_n$ entwickle man mit dem Laplaceschen Entwicklungssatz nach der ersten Spalte von Γ_n und verwende diesen danach noch weiter, um schließlich auf die Differenzengleichung

$$c_n - \sigma_e^2(1+b^2)c_{n-1} + \sigma_e^4 b^2 c_{n-2} = 0$$

zu kommen. Diese löse man wie in Abschnitt 8.2 beschrieben. Man kommt auf den Fall mit zwei reellen Nullstellen z_1, z_2 , wobei in der allgemeinen Darstellung der Lösung $c_n = \kappa_1 z_1^{-n} + \kappa_2 z_2^{-n}$ die Werte κ_1, κ_2 durch die bekannten (zu bestimmenden) Werte von c_1 und c_2 festgelegt werden. Den Wert von $\det \hat{\Gamma}_n$ bestimmt man direkt durch Entwicklung nach der letzten Spalte.

Aufgabe 8.6 Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, stationäre, zentrierte Zeitreihe mit Kovarianzfunktion $\gamma(h)$ und Spektralmaß μ .

a) Für reelle Konstanten c_1, \dots, c_r zeige man mit $C(z) := \sum_{j=1}^r c_j z^j$

$$\sum_{j=1}^r \sum_{k=1}^r c_j c_k \gamma(j-k) = \int_{(-\pi, \pi]} |C(e^{i\omega})|^2 \mu(d\omega).$$

b) Man zeige: Die Kovarianzmatrix $\Gamma_r := (\gamma(j-k))_{j,k=1,\dots,r}$ ist ausgeartet für mindestens ein r , genau dann wenn μ ein Maß mit endlichem Träger ist.

Aufgabe 8.7 Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre, zentrierte, reelle Zeitreihe und

$$\pi(n) := \frac{\langle X_{n+1} - P_{\text{span}\{X_2, \dots, X_n\}} X_{n+1}, X_1 - P_{\text{span}\{X_2, \dots, X_n\}} X_1 \rangle}{\|X_{n+1} - P_{\text{span}\{X_2, \dots, X_n\}} X_{n+1}\| \cdot \|X_1 - P_{\text{span}\{X_2, \dots, X_n\}} X_1\|}$$

die partielle Autokorrelation, $n \geq 2$.

a) Man berechne $\pi(2)$ direkt aus der Definition.

b) Man berechne $\pi(2)$ mit Hilfe von Aufgabe 8.4 c).

c) Ist X eine $AR(2)$ -Reihe, so drücke man $\pi(2)$ mit Hilfe von a_1, a_2 aus.

Deterministische und rein nicht-deterministische Zeitreihen

Wir wollen in diesem Abschnitt bei einer stationären, komplexen, zentrierten Zeitreihe $X = (X_t : t \in \mathbb{Z})$ die Frage untersuchen, wie gut die zukünftige Zufallsvariable $X_{t+\tau}$ durch eine Linearkombination von $X_s, s \leq t$, also der gesamten Vergangenheit bis zum Zeitpunkt t , im komplexen Raum $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ approximiert (*linear vorhergesagt*) werden kann.

In der Realität ist natürlich nicht die gesamte unendliche Vergangenheit $X_s, s \leq t$, zum Zeitpunkt t bekannt. Die Überlegungen dieses Abschnitts können demgemäß nur approximativ gelten, wenn eine genügend lange Vorgeschichte bekannt ist. Den realistischeren Fall von Vorhersagen bei endlicher Vergangenheit hatten wir schon in Kapitel 4 behandelt. Die Verwendung der unendlichen Vergangenheit erlaubt es jedoch, einige grundsätzliche Überlegungen zur Natur schwach stationärer Zeitreihen anzustellen.

Wir bezeichnen mit $H(X)$ den linearen Teilraum $\overline{\text{span}}\{X_t : t \in \mathbb{Z}\}$ von $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ und entsprechend $H_t(X) = \overline{\text{span}}\{X_s : s \leq t\}$. Offenbar gilt $H_s(X) \subset H_t(X)$ für $s \leq t$. $H_{-\infty}(X) := \bigcap_{t \in \mathbb{Z}} H_t(X)$ ist ein Teilhilbertraum von $H(X)$. Er besteht anschaulich gesprochen aus all den Elementen von $H(X)$, die zu jedem Zeitpunkt $t \in \mathbb{Z}$ bekannt sind. Dies führt zu nachfolgender Definition. Zuvor sei bemerkt, dass wir im Folgenden konsequent im komplexen Raum $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ arbeiten.

Definition 9.1 (*Deterministische und rein nicht-deterministische Reihen*)

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre, zentrierte, komplexe Zeitreihe.

- a) X heißt *deterministisch*, falls $H_{-\infty}(X) = H(X)$ gilt, d.h. falls für alle $t \in \mathbb{Z}$ die Gleichheit $H_t(X) = H(X)$ erfüllt ist.
- b) X heißt *rein nicht-deterministisch*, falls $H_{-\infty}(X) = \{0\}$ gilt.

Es wird sich im nachfolgenden Abschnitt 9.1 zeigen, dass stationäre Zeitreihen eindeutig in deterministische und rein nicht-deterministische Zeitreihen zerlegt werden können (Wold-Zerlegung). Die im Rahmen einer stochastischen Behandlung von Zeitreihen besonders interessanten rein nicht-

deterministischen Zeitreihen besitzen stets eine $MA(\infty)$ -Darstellung, allerdings nur mit einem unkorrelierten weißen Rauschen (vgl. Satz 9.3). Über den Vorhersagefehler bei stationären Zeitreihen gibt Satz 9.4 Auskunft.

Für die Beurteilung der Bedeutung der Modellklasse der autoregressiven und moving average Zeitreihen betrachten wir in Abschnitt 9.2 eine entsprechende Approximation der Spektraldichten von allgemeinen stationären Zeitreihen durch solche von AR - bzw. MA -Zeitreihen.

Die auf den ersten Blick aufwändiger erscheinende grundsätzliche Formulierung vieler unser Resultate in der Welt der komplexen Zeitreihen hat hauptsächlich den schon früher erwähnten Grund, dass der wichtige Spektralsatz, der immer wieder verwendet wird, viel besser in der komplexen Welt formuliert werden kann. Würden wir von reellen Zeitreihen ausgehen, so müsste stets die Frage untersucht werden, ob durch die Anwendung des Spektralsatzes nicht ein Übergang in die komplexe Welt erzwungen würde. Folglich müssten wir untersuchen, ob resultierende Zeitreihen dann auch wieder reell sind. Wir haben uns dazu entschieden, grundsätzlich von einer komplexwertigen Darstellung der Zeitreihen auszugehen und werden, wo immer dies möglich ist (z.B. in Lemma 9.9), zeigen, dass die resultierenden Zeitreihen bei reellwertigen Ausgangszeitreihen selbst wieder reell sind. So hoffen wir der Tatsache Rechnung zu tragen, dass reellwertige Zeitreihen natürlich in der realen Anwendung die absolut dominierende Rolle spielen.

9.1 Die Wold-Zerlegung

Der folgende Satz von Wold gibt einen grundlegenden Aufschluss über die Struktur stationärer Zeitreihen.

Satz 9.2 (Zerlegungssatz von Wold)

Eine stationäre, zentrierte, komplexe Zeitreihe $X = (X_t : t \in \mathbb{Z})$ kann auf genau eine Weise zerlegt werden gemäß

$$X_t = X_{1t} + X_{2t}, \quad \forall t \in \mathbb{Z}, \quad (9.1)$$

wobei $X_1 = (X_{1t} : t \in \mathbb{Z})$ rein nicht-deterministisch und $X_2 = (X_{2t} : t \in \mathbb{Z})$ deterministisch ist, sowie $X_{1s} \perp X_{2t} \forall t, s \in \mathbb{Z}$ (X_1 und X_2 heißen dann orthogonal) und $X_{1t}, X_{2t} \in H_t(X) \forall t \in \mathbb{Z}$ gelten.

Beweis: Zum Nachweis der Existenz sei $X_{2t} := P_{-\infty} X_t$ und $X_{1t} := X_t - X_{2t}, t \in \mathbb{Z}$. Dabei bezeichnet $P_{-\infty}$ die Projektion auf den linearen Teilraum $H_{-\infty}(X)$. Dann folgt $X_{2t} \in H_{-\infty}(X) \subset H_t(X)$, also auch $X_{1t} \in H_t(X)$.

Wegen $X_{2t} \in H_{-\infty}(X)$ und $X_{1t} \perp H_{-\infty}(X)$ folgt $X_{1s} \perp X_{2t} \forall s, t \in \mathbb{Z}$.

Aus $X_{1t} \in H_t(X) \forall t \in \mathbb{Z}$ erhalten wir $H_{-\infty}(X_1) \subset H_{-\infty}(X)$. Andererseits gilt $X_{1t} \perp H_{-\infty}(X)$ nach Konstruktion, also $H_{-\infty}(X_1) = \{0\}$, d.h. X_1 ist rein nicht-deterministisch. In Aufgabe 9.6 wird die Stationarität von X_1

nachgewiesen, aus der dann auch diejenige von X_2 folgt.

Wir zeigen nun, dass X_2 deterministisch ist.

Wegen $X_{2t} \in H_{-\infty}(X) \quad \forall t \in \mathbb{Z}$ folgt $H(X_2) \subset H_{-\infty}(X)$. Andererseits folgt aus (9.1) $H_t(X) = H_t(X_1) \oplus H_t(X_2)$. Hieraus ergibt sich $H_{-\infty}(X) \subset H_t(X_1) \oplus H_t(X_2) \quad \forall t \in \mathbb{Z}$. Da aber $H_t(X_1) \perp H_{-\infty}(X)$ gilt, erhält man sogar $H_{-\infty}(X) \subset H_t(X_2) \subset H_{-\infty}(X) \quad \forall t \in \mathbb{Z}$, woraus die Behauptung $H_t(X_2) = H(X_2) \quad \forall t \in \mathbb{Z}$ folgt.

Zum abschließenden Nachweis der Eindeutigkeit der Zerlegung sei $X_t = X_{1t} + X_{2t}$ eine beliebige Zerlegung mit den Eigenschaften des Satzes. Dann gilt zunächst wie oben $H_t(X) = H_t(X_1) \oplus H_t(X_2) = H_t(X_1) \oplus H(X_2)$, $\forall t \in \mathbb{Z}$, da X_2 deterministisch ist. Hieraus folgt auch die Gleichheit $H_{-\infty}(X) = H_{-\infty}(X_1) \oplus H(X_2)$, wobei die Inklusion „ \supset “ offensichtlich ist und „ \subset “ aus der Tatsache folgt, dass ein Element $Z \in H_{-\infty}(X) \subset H_t(X_1) \oplus H(X_2) \quad \forall t \in \mathbb{Z}$ eine orthogonale Zerlegung $Z = Z_1 + Z_2$ besitzt mit $Z_2 \in H(X_2)$ und $Z_1 \in H_t(X_1) \quad \forall t$, also mit $Z_1 \in H_{-\infty}(X_1)$.

Wegen $H_{-\infty}(X_1) = \{0\}$ folgt also $H_{-\infty}(X) = H(X_2)$. Damit gilt $X_{2t} \in H_{-\infty}(X)$ und $X_{1t} \perp H(X_2) = H_{-\infty}(X)$, also $X_{2t} = P_{-\infty}X_t \quad \forall t \in \mathbb{Z}$. \square

Wir werden nun zeigen, dass jede rein nicht-deterministische Zeitreihe eine $MA(\infty)$ -Darstellung besitzt.

Satz 9.3 (*$MA(\infty)$ -Darstellung rein nicht-deterministischer Zeitreihen*)

a) Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre, zentrierte, komplexe, rein nicht-deterministische Zeitreihe, die nicht identisch verschwindet. Dann gibt es ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ und Konstanten $c_j \in \mathbb{C}, j = 0, 1, 2, \dots$, mit $\sum_{j=0}^{\infty} |c_j|^2 < \infty$ derart, dass

$$X_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (9.2)$$

gilt. Das weiße Rauschen e kann so gewählt werden, dass $H_t(X) = H_t(e)$ für alle $t \in \mathbb{Z}$ gilt.

b) Umgekehrt folgt aus einer Darstellung (9.2) mit einem beliebigen weißen Rauschen $e = (e_t : t \in \mathbb{Z})$, dass X rein nicht-deterministisch ist.

Bevor wir uns dem Beweis des obigen Satzes zuwenden, definieren wir für $t \in \mathbb{Z}$ den Shift-Operator $T_t : H(X) \rightarrow H(X)$ durch die Festlegung $T_t X_s = X_{s+t} \quad \forall s \in \mathbb{Z}$. Wegen der Stationarität von X gilt $\langle X_{s_1}, X_{s_2} \rangle = \langle X_{s_1+t}, X_{s_2+t} \rangle = \langle T_t X_{s_1}, T_t X_{s_2} \rangle \quad \forall s_1, s_2 \in \mathbb{Z}$. Nach dem Isometriesatz 5.3 ist T_t als Isometrie durch die obige Festlegung bestimmt. Folglich gilt

$$T_t H_s(X) = H_{t+s}(X), \quad \forall s, t \in \mathbb{Z}. \quad (9.3)$$

Weiter gilt für die Projektionen P_t auf den Raum $H_t(X)$

$$T_r P_s X_t = P_{r+s} X_{r+t}, \quad \forall r, s, t \in \mathbb{Z}. \quad (9.4)$$

Letzteres sieht man folgendermaßen: Wegen $P_s X_t \in H_s(X)$ und $X_t - P_s X_t \perp H_s(X)$ folgt $T_r P_s X_t \in H_{r+s}(X)$ und $X_{t+r} - T_r P_s X_t = T_r(X_t - P_s X_t) \perp T_r H_s(X) = H_{r+s}(X)$. Damit ist $T_r P_s X_t$ die Projektion von X_{t+r} auf $H_{r+s}(X)$, d.h. es gilt (9.4).

Die $MA(\infty)$ -Darstellung (9.2) für eine rein nicht-deterministische Zeitreihe ist bemerkenswert und demonstriert, wie umfassend die Klasse der $MA(\infty)$ -Zeitreihen in der Tat ist. Ausdrücklich sei aber darauf hingewiesen, dass in einer solchen $MA(\infty)$ -Darstellung einer beliebigen rein nicht-deterministischen Zeitreihe die Koeffizienten c_j lediglich quadratsummierbar sind und das weiße Rauschen aus unkorrelierten Zufallsvariablen besteht. Im Gegensatz zu dieser sehr großen Klasse repräsentieren die $MA(\infty)$ -Zeitreihen mit absolut summierbaren Koeffizienten und stochastisch unabhängigen Innovationen e_t (die sogenannten *linearen Prozesse*) deutlich weniger stationäre Zeitreihen.

Beweis: (von Satz 9.3)

a) Falls für ein $s \in \mathbb{Z}$ die Gleichung $X_s = P_{s-1} X_s$, also $X_s \in H_{s-1}(X)$, gilt, so folgt nach (9.4) für ein beliebiges $t \in \mathbb{Z}$ auch $X_t = T_{t-s} X_s = T_{t-s} P_{s-1} X_s = P_{t-1} X_t$. Folglich gilt $H_t(X) = H(X) \quad \forall t \in \mathbb{Z}$, also $H(X) = H_{-\infty}(X) = \{0\}$. Wegen $X \neq 0$ ergibt sich ein Widerspruch. Es gilt also $X_s - P_{s-1} X_s \neq 0 \quad \forall s \in \mathbb{Z}$. Die Einheitsvektoren $e_s := (X_s - P_{s-1} X_s) / \|X_s - P_{s-1} X_s\|$, $s \in \mathbb{Z}$, stehen wegen $e_s \perp H_{s-1}(X)$ paarweise aufeinander senkrecht und es gilt $\langle X_t, e_{t-j} \rangle = \langle T_{-t} X_t, T_{-t} e_{t-j} \rangle = \langle X_0, e_{-j} \rangle =: c_j$, unabhängig von t mit $c_j = 0$ für $j < 0$. Wegen der Besselschen Ungleichung, siehe Anhang A.3, folgt weiter

$$\sum_{j=0}^{\infty} |c_j|^2 = \sum_{j=0}^{\infty} |\langle X_0, e_{-j} \rangle|^2 \leq \|X_0\|^2 < \infty.$$

Mit $D_s := \text{span}\{e_s\}$ gilt für $i > k$

$$H_{t-k}(X) = H_{t-i}(X) \oplus D_{t-i+1} \oplus \cdots \oplus D_{t-k} \quad (9.5)$$

also auch

$$P_{t-k} X_t = P_{t-i} X_t + \sum_{j=k}^{i-1} c_j e_{t-j}. \quad (9.6)$$

Da die Reihe in (9.6) bei festem k für $i \rightarrow \infty$ konvergiert, existiert auch $\lim_{i \rightarrow \infty} P_{t-i} X_t$ und gehört offenbar zu $H_{-\infty}(X) = \{0\}$. Aus (9.6) folgt demnach

$$P_{t-k} X_t = \sum_{j=k}^{\infty} c_j e_{t-j} \quad \forall t, k \in \mathbb{Z}. \quad (9.7)$$

Für $k = 0$ ergibt sich (9.2). Aus (9.2) folgt einerseits $H_t(X) \subset H_t(e)$ und andererseits gilt $e_t \in H_t(X)$ nach Konstruktion, d.h. $H_t(e) \subset H_t(X)$, insgesamt

also Gleichheit.

b) Wegen $H_{t-1}(e) \perp e_t \forall t$ gilt $H_{-\infty}(e) \perp H(e)$, also $H_{-\infty}(e) = \{0\}$. Andererseits gilt wegen (9.2) $H_{-\infty}(X) \subset H_{-\infty}(e)$, also insgesamt $H_{-\infty}(X) = \{0\}$, d.h. X ist rein nicht-deterministisch. \square

In der Regel lassen sich, wie wir in Kapitel 7, Satz 7.10, gesehen haben, $ARMA(p, q)$ -Reihen als $MA(\infty)$ -Reihen darstellen, sind also rein nicht-deterministisch.

Wegen (9.4) ist der sogenannte lineare Vorhersagefehler $\delta_X(\tau) = \delta(\tau)$ einer stationären, zentrierten, komplexen Zeitreihe $X = (X_t : t \in \mathbb{Z})$ gemäß

$$\delta(\tau) = \|X_{t+\tau} - P_t X_{t+\tau}\|^2 = \|X_\tau - P_0 X_\tau\|^2, \quad \forall \tau \in \mathbb{Z}, \quad (9.8)$$

unabhängig von t . Wegen $H_s(X) \subset H_t(X)$ für $s \leq t$ ist $\delta(\tau)$ monoton nichtfallend in τ , und es gilt offenbar $\delta(\tau) = 0$ für $\tau \leq 0$. Für eine rein nicht-deterministische Zeitreihe $X \neq 0$ wie in Satz 9.3 a) gilt gemäß (9.7) für $\tau \rightarrow \infty$

$$\delta(\tau) = \sum_{j=0}^{\tau-1} |c_j|^2 \rightarrow \sum_{j=0}^{\infty} |c_j|^2 = \|X_t\|^2 = \gamma_X(0). \quad (9.9)$$

Allgemein können wir das folgende Resultat beweisen.

Satz 9.4 (*Vorhersagefehler*)

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre, zentrierte, komplexe Zeitreihe. Dann gilt mit der Wold-Zerlegung $X = X_1 + X_2$ (9.1) die Gleichheit $\delta_X(\tau) = \delta_{X_1}(\tau)$ und

$$\delta_{X_1}(\tau) \rightarrow \gamma_{X_1}(0) \quad \text{für} \quad \tau \rightarrow \infty. \quad (9.10)$$

Beweis: Ist $X_1 \neq 0$, so folgt (9.10) wegen (9.9), während sie bei $X_1 \equiv 0$ trivial ist. Wegen $X_{2s} \in H_{-\infty}(X)$ gilt $P_t X_s = P_t X_{1s} + X_{2s}$ und wegen $H_t(X) = H_t(X_1) \oplus H_t(X_2)$ zunächst $P_t X_{1s} \in H_t(X_1)$. Wegen $X_{1s} - P_t X_{1s} \perp H_t(X) \supset H_t(X_1)$ ergibt sich also die Gleichheit $P_t X_{1s} = P_t^{(1)} X_{1s}$, wobei $P_t^{(1)}$ die Projektion auf $H_t(X_1)$ im Raum $H(X_1)$ sei.

Es folgt $X_{t+\tau} - P_t X_{t+\tau} = X_{1,t+\tau} - P_t^{(1)} X_{1,t+\tau}$ und damit $\delta_X(\tau) = \delta_{X_1}(\tau)$. \square

Korollar 9.5

Sei $X = (X_t : t \in \mathbb{Z})$ eine stationäre, zentrierte, komplexe Zeitreihe. Dann gilt

- a) X ist deterministisch genau dann, wenn $\delta_X(\tau) = 0 \forall \tau \in \mathbb{Z}$ gilt.
- b) X ist genau dann rein nicht-deterministisch, wenn $\delta_X(\tau) \rightarrow \gamma_X(0)$ für $\tau \rightarrow \infty$ gilt.

Beweis: a) Aus $\delta_X(\tau) = 0 \forall \tau$ folgt nach (9.10) $\|X_{1t}\|^2 = \gamma_{X_1}(0) = 0 \forall t$, d.h. $X_1 = 0$. Also ist $X = X_2$ deterministisch. Die Umkehrung ist klar.

b) Es gilt: X ist rein nicht-deterministisch $\Leftrightarrow X_2 = 0 \Leftrightarrow \gamma_{X_2}(0) = 0 \Leftrightarrow \gamma_X(0) = \gamma_{X_1}(0)$ (letzteres, da $\gamma_X(0) = \gamma_{X_1}(0) + \gamma_{X_2}(0)$). (9.10) liefert nun die Behauptung. \square

Während sich also deterministische Reihen ohne Vorhersagefehler linear vorhersagen lassen, ist bei rein nicht-deterministischen Reihen für $\tau \rightarrow \infty$ keine lineare Vorhersagefunktion besser als die Vorhersagefunktion identisch 0. Das heißt, dass sich rein nicht-deterministische Zeitreihen bei großem Vorhersagehorizont nicht besser als durch den Erwartungswert vorhersagen lassen. Hierin kommt zum Ausdruck, dass die vorhandene Abhängigkeitsstruktur der einzelnen Zeitreihenwerte untereinander, die ja erst eine verbesserte Prognose möglich macht, mit wachsendem Vorhersagehorizont abklingt.

Bemerkung 9.6

Nach Satz 9.3 in Verbindung mit Beispiel 6.11 besitzt jede rein nicht-deterministische Zeitreihe, die nicht identisch verschwindet, eine Spektraldichte. Genauer kann man zeigen, dass bei der Wold-Zerlegung $X = X_1 + X_2$ gemäß Satz 9.2 für die Spektralmaße $\mu_X, \mu_{X_1}, \mu_{X_2}$ gilt

$$\mu_X = \mu_{X_1} + \mu_{X_2} \quad (\text{Aufgabe 9.7}). \quad (9.11)$$

Ist $X_1 \not\equiv 0$, so ist (9.11) die Lebesgue-Zerlegung von μ_X , d.h. $\mu_{X_1} \ll \lambda_1$ und $\mu_{X_2} \perp \lambda_1$ (d.h. es gibt eine Borelmenge $N \subset (-\pi, \pi]$ mit $\lambda_1(N) = 0$ derart, dass $\mu_{X_2}((-\pi, \pi] \setminus N) = 0$ ist.) Zum Beweis der letzten Aussage siehe Ash/Gardner (1975), Theorem 2.5.3. \square

Beispiel 9.7

Sei $X_1 = (X_{1t} : t \in \mathbb{Z})$ eine rein nicht-deterministische $L_2^{\mathbb{C}}$ -Reihe, $X_1 \not\equiv 0$. X_1 besitzt wegen Bemerkung 9.6 stets eine Spektraldichte. Ist $Y \in L_2^{\mathbb{C}}$ eine zentrierte Zufallsvariable, so ist der konstante Prozess $X_{2t} := Y, t \in \mathbb{Z}$, deterministisch, da offenbar $H_t(X_2) = \overline{\text{span}}\{Y\} \forall t$ gilt. Ist Y mit allen X_{1t} unkorreliert, so ist $X_1 + X_2$ gerade die Wold-Zerlegung von

$$X_t := X_{1t} + Y, \quad \forall t \in \mathbb{Z}.$$

Um dies zu sehen, reicht es nach Satz 9.2 zu zeigen, dass $X_{1t}, X_{2t} \in H_t(X)$ für $t \in \mathbb{Z}$ gilt. Hinreichend hierfür ist offenbar, dass Y zu $H_t(X)$ gehört. Nach dem später bewiesenen Satz 10.1 gilt aber

$$\frac{1}{N} \sum_{j=0}^{N-1} X_{t-j} = \frac{1}{N} \sum_{j=0}^{N-1} X_{1,t-j} + Y \rightarrow Z_{X_1}(\{0\}) + Y = Y,$$

da $Z_{X_1}(\{0\}) = 0$ ist (X_1 besitzt eine Spektraldichte). Es folgt $Y \in H_t(X)$. Tatsächlich werden wir in Satz 10.1 die obige Aussage nur für reellwertige

Zeitreihen formulieren, da dort nur solche betrachtet werden. Der Beweis gilt aber genauso für komplexe Zeitreihen. \square

Beispiel 9.8 (Vorhersage bei AR -Zeitreihen mit unendlicher Vergangenheit)

Sei $X = (X_t : t \in \mathbb{Z})$ eine kausale $AR(p)$ -Reihe, vgl. Definition 7.9 in Verbindung mit Bemerkung 7.8 im Spezialfall $q = 0$.

Anwendung der Projektion P_t auf die definierende Gleichung

$$X_{t+1} + \sum_{j=1}^p a_j X_{t+1-j} = e_{t+1}, \quad \forall t \in \mathbb{Z},$$

liefert wegen $e_{t+1} \perp H_t(X)$

$$P_t X_{t+1} = - \sum_{j=1}^p a_j X_{t+1-j}. \quad (9.12)$$

Der „one-step“-Vorhersagefehler (d.h. $\tau = 1$) ist

$$\delta(1) = \|X_{t+1} - P_t X_{t+1}\|^2 = \|e_{t+1}\|^2 = \sigma_e^2. \quad (9.13)$$

In Verallgemeinerung von (9.12) kann man leicht zeigen, dass für $\tau > 1$ die folgende Rekursionsformel gilt

$$P_t X_{t+\tau} = - \sum_{j=1}^{\tau-1} a_j P_t X_{t+\tau-j} - \sum_{j=\tau}^p a_j X_{t+\tau-j} \quad (9.14)$$

falls $1 \leq \tau \leq p$ und

$$P_t X_{t+\tau} = - \sum_{j=1}^p a_j P_t X_{t+\tau-j} \quad \text{falls } \tau > p. \quad (9.15)$$

Da jede AR -Reihe trivialerweise invertibel ist, gilt $H_t(X) = H_t(e) \forall t$, und es ergibt sich mit der Darstellung (9.2)

$$P_t X_{t+\tau} = \sum_{j=\tau}^{\infty} c_j e_{t+\tau-j}, \quad \forall \tau \geq 1, \quad (9.16)$$

und damit als linearer Vorhersagefehler

$$\delta(\tau) = \|X_{t+\tau} - P_t X_{t+\tau}\|^2 = \sum_{j=0}^{\tau-1} |c_j|^2 \sigma_e^2, \quad \forall \tau \geq 1. \quad (9.17)$$

Für eine $AR(1)$ -Reihe $X_t + a_1 X_{t-1} = e_t$ gilt z.B.

$$P_t X_{t+\tau} = (-a_1)^\tau X_t, \quad \forall t \in \mathbb{Z}, \tau \geq 1. \quad (9.18)$$

\square

Wie bereits in der Einleitung zu diesem Kapitel erwähnt, haben wir bisher konsequent alle auftretenden Zeitreihen im komplexen Raum $L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ betrachtet, selbst wenn sie reell waren. Im folgenden Lemma gehen wir nun auf den Fall reeller Zeitreihen ein.

Zur genaueren Kennzeichnung versehen wir die verschiedenen Symbole mit einem oberen Index \mathbb{C} , wenn wir im komplexen Raum $L_2^{\mathbb{C}} = L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$ arbeiten, also $H^{\mathbb{C}}(X), H_t^{\mathbb{C}}(X), H_{-\infty}^{\mathbb{C}}(X), P_t^{\mathbb{C}}, P_{-\infty}^{\mathbb{C}}$ und mit einem oberen Index \mathbb{R} , wenn X reell ist und wir im reellen Raum $L_2^{\mathbb{R}} = L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$ arbeiten, also $H^{\mathbb{R}}(X)$ usw.

Lemma 9.9

Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, stationäre, zentrierte Zeitreihe.

a) Dann gelten für alle $t \in \mathbb{Z}$

$$H_t^{\mathbb{C}}(X) = \{Z_1 + iZ_2 : Z_1, Z_2 \in H_t^{\mathbb{R}}(X)\}, \quad (9.19)$$

$$H_{-\infty}^{\mathbb{C}}(X) = \{Z_1 + iZ_2 : Z_1, Z_2 \in H_{-\infty}^{\mathbb{R}}(X)\}, \quad (9.20)$$

$$H^{\mathbb{C}}(X) = \{Z_1 + iZ_2 : Z_1, Z_2 \in H^{\mathbb{R}}(X)\}, \quad (9.21)$$

$$Z \perp H_t^{\mathbb{R}}(X) \Rightarrow Z \perp H_t^{\mathbb{C}}(X), \quad \forall Z \in H^{\mathbb{C}}(X), \quad (9.22)$$

$$P_t^{\mathbb{R}}Z = P_t^{\mathbb{C}}Z, \quad \forall Z \in H^{\mathbb{R}}(X), \quad (9.23)$$

$$P_{-\infty}^{\mathbb{R}}Z = P_{-\infty}^{\mathbb{C}}Z, \quad \forall Z \in H^{\mathbb{R}}(X). \quad (9.24)$$

b) Die Gleichung $H_t^{\mathbb{C}}(X) = H^{\mathbb{C}}(X) \quad \forall t \in \mathbb{Z}$ gilt genau dann, wenn die Gleichung $H_t^{\mathbb{R}}(X) = H^{\mathbb{R}}(X) \quad \forall t \in \mathbb{Z}$ gilt. Die reelle Zeitreihe X ist also genau dann komplex deterministisch, wenn sie reell deterministisch ist.

c) Genau dann gilt $H_{-\infty}^{\mathbb{C}}(X) = \{0\}$, wenn $H_{-\infty}^{\mathbb{R}}(X) = \{0\}$ gilt. Eine reelle Zeitreihe X ist also genau dann komplex rein nicht-deterministisch, wenn sie reell rein nicht-deterministisch ist.

Der Beweis ist als Aufgabe 9.10 gestellt. Insbesondere folgt aus Lemma 9.9, dass für eine reelle Zeitreihe X in der Wold-Zerlegung $X = X_1 + X_2$, X_1 und X_2 reell sind. Außerdem ist dann das weiße Rauschen $e = (e_t : t \in \mathbb{Z})$ in (9.2) reell, ebenso wie die Konstanten c_j .

9.2 Approximation durch AR- und MA-Reihen

Wir wollen nun zeigen, dass die Spektraldichten sowohl von MA- als auch von AR-Reihen fast jede stetige Spektraldichte beliebig gut approximieren können. Als Vorbereitung dient das folgende Resultat, welches in einem gewissen Sinne als Umkehrung der Aussage verstanden werden kann, dass die Autokovarianzfunktion einer MA(q)-Zeitreihe ab dem Lag $q+1$ identisch Null ist (vgl. Beispiel 2.6).

Lemma 9.10

Sei $X = (X_t : t \in \mathbb{Z})$ eine komplexe, stationäre, zentrierte Zeitreihe, deren Autokovarianzfunktion γ für $|h| > q$ verschwindet und $\gamma(q) \neq 0$ erfüllt. Dann ist X eine $MA(q)$ -Reihe, d.h. es existieren ein weißes Rauschen $e = (e_t : e_t \in \mathbb{Z})$ sowie Konstanten $b_0, \dots, b_q \in \mathbb{C}, b_0 = 1, b_q \neq 0$, mit

$$X_t = \sum_{j=0}^q b_j e_{t-j}, \quad \forall t \in \mathbb{Z}. \quad (9.25)$$

Ist X reell, so können sowohl das weiße Rauschen e als auch alle Koeffizienten b_j reell gewählt werden.

Beweis: (Aufgabe 9.1) Hinweis: Man zeige, dass $X_s \perp X_t$ für $|s - t| > q$ gilt und schließe daraus, dass der Vorhersagefehler $\delta(\tau) = \|X_{t+\tau} - P_t X_{t+\tau}\|^2$ für $|\tau| > q$ die Gleichung $\delta(\tau) = \gamma(0)$ erfüllt, woraus nach Korollar 9.5 folgt, dass X rein nicht-deterministisch ist. Man wende sodann Satz 9.3 an und zeige in der Entwicklung (9.2), dass $0 = b_{q+1} = b_{q+2} = \dots$ gilt, sowie $b_q \neq 0$. \square

Wir formulieren hier noch die zu Lemma 9.10 duale Aussage für AR-Reihen, obwohl wir sie nicht für die nachfolgenden Approximationssätze benötigen.

Lemma 9.11

Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, stationäre, zentrierte Zeitreihe, deren partielle Autokorrelation π für $h > p$ verschwindet und $\pi(p) \neq 0$ erfüllt. Dann ist X eine $AR(p)$ -Reihe, d.h. es existieren ein weißes Rauschen $e = (e_t : e_t \in \mathbb{Z})$ sowie Konstanten $a_1, \dots, a_p \in \mathbb{R}, a_p \neq 0$, mit

$$X_t + \sum_{j=1}^p a_j X_{t-j} = e_t, \quad \forall t \in \mathbb{Z}. \quad (9.26)$$

Beweis: (Aufgabe 9.2) Hinweis: Seien P_t bzw. $P_{s,t}$ die Projektionen in $L_2^{\mathbb{R}}$ auf $\overline{\text{span}}\{X_k : k \leq t\}$ bzw. $H_{s,t} = \text{span}\{X_k : s < k \leq t\}$. Aus $\pi(h) = 0$ für $h \geq p + 1$ folgere man die Beziehung

$$\langle X_{t+1} - P_{s-p,t} X_{t+1}, X_{s-p} \rangle = 0, \quad \forall s \leq t, \quad (9.27)$$

und schließe daraus

$$e_{t+1} := X_{t+1} - P_{t-p,t} X_{t+1} \perp X_s, \quad \forall s \leq t, \quad (9.28)$$

d.h. $P_t X_{t+1} = P_{t-p,t} X_{t+1} \forall t$. Hieraus schließe man, dass $(e_t : t \in \mathbb{Z})$ ein weißes Rauschen ist. Sind a_1, \dots, a_p die (von t unabhängigen?) eindeutig bestimmten Koeffizienten mit

$$P_{t-p,t} X_{t+1} = - \sum_{j=1}^p a_j X_{t+1-j}, \quad \forall t \in \mathbb{Z}, \quad (9.29)$$

so folgt (9.26). \square

Satz 9.12 (Approximation durch eine MA-Reihe)

Sei $f : [-\pi, \pi] \rightarrow [0, \infty)$ eine stetige Spektraldichte mit $f(-\pi) = f(\pi)$. Dann gibt es zu jedem $\varepsilon > 0$ eine MA(q)-Reihe $X = (X_t : t \in \mathbb{Z})$ mit Spektraldichte f_ε , für die $|f(\omega) - f_\varepsilon(\omega)| \leq \varepsilon \forall \omega \in [-\pi, \pi]$ gilt. Ist f symmetrisch um Null, so ist X reell wählbar.

Beweis: Nach Rudin (1970), Theorem 4.25, gibt es ein trigonometrisches Polynom $\sum_{j=-q}^q c_j e^{ij\omega} =: f_\varepsilon(\omega)$ auf $[-\pi, \pi]$, das die Funktion $k(\omega) := f(\omega) + 3\varepsilon/4$ bis auf $\varepsilon/4$ approximiert, d.h. $|k(\omega) - f_\varepsilon(\omega)| \leq \varepsilon/4 \forall \omega \in [-\pi, \pi]$. Weil $k(\cdot)$ reell ist, kann dabei $c_j = \bar{c}_{-j}$ gewählt werden; man ersetze sonst f_ε durch $\frac{1}{2}(f_\varepsilon + \bar{f}_\varepsilon)$. Dann ist $f_\varepsilon(\omega)$ reell mit $f_\varepsilon(\omega) \geq \frac{\varepsilon}{2}$ und $|f(\omega) - f_\varepsilon(\omega)| \leq 3\varepsilon/4 + \varepsilon/4 = \varepsilon$. Zu f_ε gehöre die Autokovarianzfunktion γ und $X = (X_t : t \in \mathbb{Z})$ sei eine Zeitreihe die diese Autokovarianzfunktion besitzt, vgl. Satz 3.1. Wegen der Orthogonalität der Funktionen $e^{in\cdot}$, vgl. Bemerkung 6.4, Teil c), gilt dann gemäß (3.7) $\gamma(h) = 2\pi c_{-h}$ für $|h| \leq q$ und $\gamma(h) = 0$ für $|h| > q$, wobei o.B.d.A. $c_q \neq 0$ gelte. Nach Lemma 9.10 folgt nun die Behauptung.

Ist f symmetrisch um 0, so kann auch f_ε als symmetrisch um 0 vorausgesetzt werden, sodass die zugehörige Kovarianzfunktion reell ist. Dann aber ist die Zeitreihe X reell wählbar, siehe Satz 3.1. \square

Satz 9.13 (Approximation durch eine AR-Reihe)

Sei $f : [-\pi, \pi] \rightarrow [0, \infty)$ eine stetige Spektraldichte mit $f(-\pi) = f(\pi)$. Dann gibt es zu jedem $\varepsilon > 0$ eine kausale AR(p)-Reihe $X = (X_t : t \in \mathbb{Z})$ mit Spektraldichte f_ε , für die $|f(\omega) - f_\varepsilon(\omega)| \leq \varepsilon \forall \omega \in [-\pi, \pi]$ gilt.

Ist f symmetrisch um 0, so ist X reell wählbar.

Beweis: Die Funktion $\tilde{f}(\omega) := f(\omega) + \varepsilon/2$ ist eine stetige Spektraldichte auf $[-\pi, \pi]$ mit $\tilde{f}(-\pi) = \tilde{f}(\pi)$ und $|f(\omega) - \tilde{f}(\omega)| \leq \varepsilon/2 \forall \omega$. Damit ist auch $g(\omega) := 1/\tilde{f}(\omega)$ eine positive, stetige Funktion auf $[-\pi, \pi]$ mit $g(-\pi) = g(\pi)$. Nach Satz 9.12 gibt es zu jedem $\varepsilon' > 0$ eine MA(q)-Reihe mit Spektraldichte $g_{\varepsilon'}$ derart, dass $|g(\omega) - g_{\varepsilon'}(\omega)| \leq \varepsilon' \forall \omega$ gilt. Mit $M := \sup \{f(\omega) : \omega \in [-\pi, \pi]\}$ und einem $\varepsilon' = \varepsilon'(\varepsilon, M) > 0$, welches $0 \leq \varepsilon'(M + \varepsilon/2)^2 / (1 - \varepsilon'(M + \varepsilon/2)) < \varepsilon/2$ erfüllt, ergibt sich für $f_\varepsilon := 1/g_{\varepsilon'}$

$$\begin{aligned} \left| \tilde{f}(\omega) - f_\varepsilon(\omega) \right| &= \left| \frac{1}{g(\omega)} - \frac{1}{g_{\varepsilon'}(\omega)} \right| = \frac{|g_{\varepsilon'}(\omega) - g(\omega)| \cdot \tilde{f}(\omega)}{g_{\varepsilon'}(\omega)} \\ &\leq \varepsilon' \frac{\tilde{f}^2(\omega)}{1 - \varepsilon' \tilde{f}(\omega)} < \frac{\varepsilon}{2}, \end{aligned} \quad (9.30)$$

also insgesamt $|f(\omega) - f_\varepsilon(\omega)| \leq \varepsilon \forall \omega$.

Da zu $g_{\varepsilon'}$ eine MA(q)-Reihe gehört, gibt es ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ und ein Polynom $B(z) = \sum_{j=0}^q b_j z^j$, $b_0 = 1, b_q \neq 0$ derart, dass $g_{\varepsilon'}(\omega) = |B(e^{-i\omega})|^2 \frac{\sigma_e^2}{2\pi}$ gilt. Dann ist $X_t := \int e^{it\cdot} (1/B(e^{-i\cdot})) (2\pi/\sigma_e^2) dZ_e$

eine stationäre Zeitreihe mit Spektraldichte f_ε , s. Satz 6.2. Weiter gilt nach Lemma 6.6

$$\sum_{j=0}^q b_j X_{t-j} = \int B(e^{-i\cdot}) dZ_X = \frac{2\pi}{\sigma_\varepsilon^2} e_t =: \varepsilon_t, \quad \forall t \in \mathbb{Z},$$

wobei $\varepsilon = (\varepsilon_t : t \in \mathbb{Z})$ ein weißes Rauschen ist. Damit ist $X = (X_t : t \in \mathbb{Z})$ eine $AR(p)$ -Reihe (mit $p = q$). Nach Satz 7.13 kann o.B.d.A. angenommen werden, dass X kausal ist. Ist f symmetrisch, so ist auch g_ε symmetrisch wählbar und damit sind die Konstanten b_j und das weiße Rauschen ε reell wählbar. Nach Satz 6.5 ist dann schließlich auch X reell. \square

Aufgaben

Aufgabe 9.1 Man beweise Lemma 9.10.

Aufgabe 9.2 Man beweise Lemma 9.11.

Aufgabe 9.3 Sei $X_t = X_{1t} + X_{2t}$ die Wold-Zerlegung der zentrierten, stationären Zeitreihe $X = (X_t : t \in \mathbb{Z})$. μ_X bezeichne das Spektralmaß von X . Man zeige:

a) Für den Shift-Operator T_t auf $H(X)$ gilt:

$$T_t X_j(s) = X_j(s+t), \quad j = 1, 2.$$

b) Für $\varphi \in L_2 := L_2((-\pi, \pi], \mathbb{B} \cap (-\pi, \pi], \mu_X)$ gilt:

$$T_t \int \varphi(\omega) Z_X(d\omega) = \int \exp(it\omega) \varphi(\omega) Z_X(d\omega).$$

c) Es existieren $\varphi_j \in L_2, j = 1, 2$, mit

$$X_j(t) = \int \exp(it\omega) \varphi_j(\omega) Z_X(d\omega).$$

Aufgabe 9.4 Für eine kausale $AR(1)$ -Reihe $X_t + aX_{t-1} = e_t$, mit reellem weißen Rauschen $e_t \sim (0, \sigma_e^2)$ und $a \in \mathbb{R}, |a| < 1$, weise man für die Projektion P_t auf $\overline{\text{span}}\{X_s : s \leq t\}$ die Beziehung

$$P_t X_{t+\tau} = (-a)^\tau X_t, \quad \forall t \in \mathbb{Z}, \tau \geq 1,$$

auf zwei Arten nach:

a) Durch Verwendung der $MA(\infty)$ -Darstellung $X_t = \sum_{j=0}^{\infty} c_j e_{t-j}$ und Formel (9.7).

b) Durch Verwendung der Formel (9.12) und durch Verwendung der Iterationseigenschaft von Projektionen.

Aufgabe 9.5 Man drücke den Vorhersagefehler $\delta(\tau)$ in Aufgabe 9.4 mit Hilfe von a und σ_e^2 aus.

Aufgabe 9.6 Sei $X = (X_t : t \in \mathbb{Z})$ eine zentrierte, komplexe, stationäre Zeitreihe und T_h der Shift-Operator $T_h : H(X) \rightarrow H(X)$ gemäß $T_h X_t = X_{t+h}$.

Man zeige:

a) $T_h \circ T_{-h}$ ist die Identität auf $H(X)$, $\forall h \in \mathbb{Z}$, und es gilt

$$\langle T_h Y_1, Y_2 \rangle = \langle Y_1, T_{-h} Y_2 \rangle, \quad \forall Y_1, Y_2 \in H(X) \quad \forall h \in \mathbb{Z}.$$

b) $T_h(H_{-\infty}(X)) = H_{-\infty}(X)$.

c) $T_h P_{-\infty} X_t = P_{-\infty} X_{t+h}$.

d) $(P_{-\infty} X_t : t \in \mathbb{Z})$ ist stationär.

Aufgabe 9.7 Man beweise Bemerkung 9.6. Die Stationarität von X_2 ergibt sich aus Aufgabe 9.6

Aufgabe 9.8 Sei $X_t = e_t + b e_{t-1}$, $t \in \mathbb{Z}$, eine reelle MA(1)-Reihe mit $b \in (-1, 1)$, $e_t \sim (0, \sigma_e^2)$. Sei $\hat{X}_{t+1} := P_{1,t} X_{t+1}$ die Projektion von X_{t+1} auf $\text{span}\{X_1, \dots, X_t\}$, $t \geq 1$, und $v_t = \|\hat{X}_{t+1} - X_{t+1}\|^2$ der Vorhersagefehler. Man zeige direkt:

$$\hat{X}_{t+1} = \frac{\sigma_e^2 \cdot b}{v_{t-1}} (X_t - \hat{X}_t) \quad (9.31)$$

$$\|\hat{X}_{t+1}\|^2 = \gamma(0) - v_t \quad (9.32)$$

$$v_t = \sigma_e^2 \left[1 + b^2 - \frac{\sigma_e^2 b^2}{v_{t-1}} \right], \quad \forall t \geq 2. \quad (9.33)$$

Hinweis: Ist P_t die Projektion auf $\overline{\text{span}}\{X_s : s \leq t\}$, so gilt

- (i) $\hat{X}_{t+1} = P_{1,t} \circ P_t X_{t+1}$,
- (ii) $P_t X_{t+1} = P_t(e_{t+1} + b e_t) = b e_t$,
- (iii) $e_t \perp X_1, \dots, X_{t-1}$. Es folgt $P_{1,t} X_{t+1} = P_{1,t}(b e_t) = b P_{\text{span}\{X_t - \hat{X}_t\}} e_t$.

Aufgabe 9.9 Seien H_n , $n \in \mathbb{Z}$, Teilhilberträume des Hilbertraumes H mit $H_n \subset H_{n+1} \quad \forall n$. Seien weiter P_n die Projektionen auf H_n und $P_{-\infty}$ die Projektion auf $H_{-\infty} = \bigcap_{n \in \mathbb{Z}} H_n$. Man zeige, dass für $n \rightarrow \infty$ gilt :

$$P_{-n} x \rightarrow P_{-\infty} x, \quad \forall x \in H.$$

Aufgabe 9.10 Man beweise Lemma 9.9.

Aufgabe 9.11 Sei $X = (X_t : t \in \mathbb{Z})$ eine komplexe, zentrierte, stationäre Zeitreihe mit zugehörigem MOW Z_X und Spektralmaß μ_X . Man zeige

- a) Es gilt $Z_X(\{\omega\}) \in H_{-\infty}(X)$, $\forall \omega \in (\pi, \pi]$.
- b) Ist μ_X diskret, so ist X deterministisch.
- c) Man zeige, dass die Zerlegung (7.11) der allgemeinen Lösung X der ARMA-Gleichung in Satz 7.4 die Wold Zerlegung ist, wobei X_0 rein nicht-deterministisch und X_1 deterministisch ist.

Hinweis zu a): Man verwende eine ähnliche Beweismethode wie in Beispiel 9.7.

Asymptotische Eigenschaften von Schätzverfahren in linearen Zeitreihenmodellen

In diesem Kapitel sollen die grundlegenden statistischen Schätzer der Zeitreihenanalyse, wie das empirische Mittel und die empirische Autokovarianz- sowie Autokorrelationsfunktion, in linearen Zeitreihenmodellen auf ihre Eigenschaften hin untersucht werden. In Abschnitt 10.1 werden dazu zunächst einfache asymptotische Eigenschaften wie Erwartungstreue und Konsistenz abgeleitet. Bezüglich der Konvergenzgeschwindigkeit fällt schon in diesem Abschnitt auf, dass für $MA(\infty)$ -Zeitreihen mit stochastisch unabhängigem und identisch verteiltem weißen Rauschen die gleichen \sqrt{n} -Konvergenzraten wie im Fall unabhängiger Beobachtungen sowohl für den Erwartungswert als auch für die Autokovarianzfunktion auftreten. Um zentrale Grenzwertaussagen für diese Größen erhalten zu können, stellen wir im Abschnitt 10.2 das grundlegende Konzept der schwachen Abhängigkeit vor. Auf dieser Basis beweisen wir dann anschließend in Abschnitt 10.3 einen allgemeinen zentralen Grenzwertsatz, den wir schließlich in Abschnitt 10.4 im Rahmen sogenannter linearer Prozesse (vgl. Definition 10.3) auf das arithmetische Mittel und die empirische Autokovarianz- wie Autokorrelationsfunktion anwenden können, um jeweils eine asymptotische Normalität zu beweisen. Hieraus können wir dann abschließend Konfidenzintervalle für den Erwartungswert und die Autokovarianz wie Autokorrelation einer linearen Zeitreihe herleiten.

Es sei darauf hingewiesen, dass in diesem und in den folgenden Kapiteln zum Teil technisch aufwändigere Beweise an der einen oder anderen Stelle nicht zu vermeiden sind. Um den Schwierigkeitsgrad in vernünftigen Grenzen zu halten, wird zum einen in einigen Beweisen bewusst nur ein Spezialfall behandelt. Andererseits werden manche Rechenschritte nicht bis ins letzte Detail ausgeführt. Es ist aber in hoffentlich ausreichendem Maße darauf geachtet worden, dass die grundsätzlichen Ideen und Ansätze deutlich werden.

Alle in den folgenden Kapiteln 10 – 13 betrachteten Zeitreihen sind reell, wenn nichts anderes gesagt wird.

10.1 Einfache asymptotische Eigenschaften des Stichprobenmittels und der Stichprobenautokovarianz

In Kapitel 2 haben wir bereits empirische Größen für die Autokovarianz– sowie Autokorrelationsfunktion einer stationären Zeitreihe erklärt. Für die folgenden Überlegungen seien stets Zufallsvariable X_1, \dots, X_n aus einer reellen stationären Zeitreihe gegeben.

Das Stichprobenmittel $\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$ ist offenbar ein erwartungstreuer Schätzer für den Erwartungswert $EX_t = \mu$. Der folgende Satz untersucht sein Konsistenzverhalten.

Satz 10.1 (*Ergodensatz für stationäre Zeitreihen*)

$X = (X_t : t \in \mathbb{Z})$ sei eine reelle stationäre Zeitreihe mit Erwartungswert μ . Das MOW der zugehörigen zentrierten Zeitreihe $(X_t - \mu : t \in \mathbb{Z})$ sei mit Z_X bezeichnet und das zugehörige Spektralmaß mit μ_X . Dann gilt für $n \rightarrow \infty$

$$\frac{1}{n} \sum_{t=1}^n X_t \rightarrow \mu + Z_X(\{0\}) \text{ in } L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P). \quad (10.1)$$

$$\frac{1}{n} \sum_{t=1}^n \gamma_X(t) \rightarrow \mu_X(\{0\}). \quad (10.2)$$

Beweis: O.B.d.A. sei $\mu = 0$ angenommen. Es gilt mit dem Spektralsatz

$$\frac{1}{n} \sum_{t=1}^n X_t = \int \frac{1}{n} \sum_{t=1}^n e^{it\omega} Z_X(d\omega). \quad (10.3)$$

Da $\left| \frac{1}{n} \sum_{t=1}^n e^{it\omega} \right| \leq 1$ und für $\omega \neq 0$

$$\left| \frac{1}{n} \sum_{t=1}^n e^{it\omega} \right| = \frac{1}{n} \left| \frac{1 - e^{i(n+1)\omega}}{1 - e^{i\omega}} \right| \xrightarrow{n \rightarrow \infty} 0, \quad (10.4)$$

folgt nach dem Satz von Lebesgue $g_n(\omega) := \frac{1}{n} \sum_{t=1}^n e^{it\omega} \rightarrow 1_{\{0\}}(\omega)$ in $L_2^{\mathbb{C}}(\mu_X)$, also nach Rechenregel (5.20)

$$\frac{1}{n} \sum_{t=1}^n X_t = \int g_n dZ_X \rightarrow Z_X(\{0\}) \quad (10.5)$$

ebenfalls im quadratischen Mittel, dieses Mal aber in $L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$. Damit ist (10.1) bewiesen. Entsprechend gilt (3.7)

$$\frac{1}{n} \sum_{t=1}^n \gamma_X(t) = \int g_n d\mu_X \rightarrow \mu_X(\{0\}) \quad (10.6)$$

und damit auch (10.2). □

Bemerkung 10.2

i) Da L_2 -Konvergenz die Konvergenz nach Wahrscheinlichkeit impliziert, gilt die Konvergenz in (10.1) des vorherigen Satzes auch nach Wahrscheinlichkeit.

ii) Wegen $\|Z_X(\{0\})\|^2 = \mu_X(\{0\})$ gilt $Z_X(\{0\}) = 0$ genau dann, wenn das Spektralmaß μ_X in 0 keine Masse trägt. Nach (10.2) ist dies genau dann der Fall, wenn $\frac{1}{n} \sum_{t=1}^n \gamma_X(t) \rightarrow 0$ gilt, also z.B. wenn $\gamma_X(h) \rightarrow 0$ für $h \rightarrow \infty$ gilt.

Bei $\mu_X(\{0\}) = 0$ gilt demnach $\overline{X}_n \rightarrow \mu$ sowohl in $L_2^{\mathbb{R}}$ als auch nach Wahrscheinlichkeit.

iii) Ist γ_X sogar absolut summierbar, so besitzt μ_X nach der Inversionsformel (vgl. Satz 3.8) eine Spektraldichte f_X , und es gilt (siehe den Beweis zur Inversionsformel und verwende (A.11) aus Anhang A.1)

$$\begin{aligned} n \operatorname{Var} \overline{X}_n &= \frac{1}{n} \sum_{s=1}^n \sum_{t=1}^n \gamma_X(s-t) \\ &= \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma_X(h) \rightarrow \sum_{h=-\infty}^{\infty} \gamma_X(h) = 2\pi f_X(0). \end{aligned} \quad (10.7)$$

Insbesondere gilt

$$n \operatorname{Var} \overline{X}_n = \mathcal{O}(1). \quad (10.8)$$

Dies ist dieselbe Größenordnung wie bei stochastischer Unabhängigkeit der X_t .

iv) Falls $X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}$ gilt, mit absolut summierbaren Koeffizienten b_j und weißem Rauschen $(e_t : t \in \mathbb{Z})$ mit $\operatorname{Var}(e_t) = \sigma_e^2$, so sind auch die Autokovarianzen $\gamma_X(h)$ absolut summierbar, sodass dann iii) anwendbar ist. Wegen $\gamma_X(h) = \sigma_e^2 \sum_{j=-\infty}^{\infty} b_j b_{j+h}$ für $h \in \mathbb{Z}$ gilt nämlich,

$$\begin{aligned} \sum_{h=-\infty}^{\infty} |\gamma_X(h)| &\leq \sigma_e^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |b_j| \cdot |b_{j+h}| = \sigma_e^2 \sum_{j=-\infty}^{\infty} |b_j| \sum_{h=-\infty}^{\infty} |b_{j+h}| \\ &= \sigma_e^2 \left(\sum_{j=-\infty}^{\infty} |b_j| \right)^2 < \infty. \end{aligned}$$

□

Nun wollen wir uns dem Schätzen der Autokovarianzfunktion einer stationären Zeitreihe zuwenden. Wir hatten in (2.29) die Stichprobenautokovarianz $\hat{\gamma}(h) = \hat{\gamma}_n(h)$ als Schätzer für $\gamma(h) = \gamma_X(h)$ eingeführt gemäß

$$\hat{\gamma}_n(h) := \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \overline{X}_n) (X_{t+h} - \overline{X}_n), \quad h = 0, 1, 2, \dots \quad (10.9)$$

Als Vorbereitung zur Bestimmung der asymptotischen Verteilung von $\widehat{\gamma}(h)$ führen wir im folgenden Satz asymptotische Berechnungen für Erwartungswert und Varianz bzw. Kovarianz dieses Autokovarianzschätzers, bzw. der Vergleichsgröße

$$\widetilde{\gamma}_n(h) := \frac{1}{n} \sum_{t=1}^n (X_t - \mu)(X_{t+h} - \mu), \quad h = 0, 1, 2, \dots, \quad (10.10)$$

durch.

Bei den nachfolgenden Resultaten in diesem Kapitel wollen wir als zugrunde liegende Zeitreihe in der Regel ein $MA(\infty)$ -Modell mit stochastisch unabhängigem und identisch verteiltem weißen Rauschen annehmen. Derartige Zeitreihen wollen wir gemäß der nachfolgenden Definition als lineare Prozesse bezeichnen.

Definition 10.3 (*Linearer Prozess*)

Ein reeller stochastischer Prozess $X = (X_t : t \in \mathbb{Z})$, der für alle $t \in \mathbb{Z}$ eine $MA(\infty)$ -Darstellung

$$X_t = \mu + \sum_{j=-\infty}^{\infty} b_j e_{t-j} \quad (10.11)$$

besitzt mit absolut summierbaren, reellen Koeffizienten b_j und stochastisch unabhängigen, identisch verteilten (i.i.d.) Zufallsvariablen $(e_t : t \in \mathbb{Z})$ mit $Ee_t = 0, Ee_t^2 = \sigma_e^2 \in (0, \infty)$, $\mu \in \mathbb{R}$, heißt linearer Prozess oder lineare Zeitreihe. X ist stets streng stationär.

Falls $b_j = 0$ für alle $j < 0$, so heißt X kausale lineare Zeitreihe.

Satz 10.4

$X = (X_t : t \in \mathbb{Z})$ sei ein linearer Prozess wie in Definition 10.3 mit der zusätzlichen Voraussetzung $Ee_t^4 =: \eta\sigma_e^4 \in (0, \infty)$. Dann gilt für $h, k \in \mathbb{N}_0$ und $n \rightarrow \infty$

$$E |\widehat{\gamma}_n(h) - \widetilde{\gamma}_n(h)| = \mathcal{O}\left(\frac{1}{n}\right). \quad (10.12)$$

Hieraus folgt die asymptotische Erwartungstreue von $\widehat{\gamma}_n(h)$, genauer

$$E \widehat{\gamma}_n(h) = \gamma(h) + \mathcal{O}\left(\frac{1}{n}\right). \quad (10.13)$$

Außerdem folgt aus (10.12)

$$\sqrt{n}(\widehat{\gamma}_n(h) - \widetilde{\gamma}_n(h)) \rightarrow 0 \quad n.W.. \quad (10.14)$$

Weiter gilt

$$n \operatorname{Cov}(\tilde{\gamma}_n(h), \tilde{\gamma}_n(k)) \rightarrow c(h, k) \quad (10.15)$$

mit

$$\begin{aligned} c(h, k) = & \gamma(h) \gamma(k) (\eta - 3) \\ & + \sum_{r=-\infty}^{\infty} (\gamma(r) \gamma(r+k-h) + \gamma(r+k) \gamma(r-h)). \end{aligned} \quad (10.16)$$

Damit sind sowohl $\hat{\gamma}_n(h)$ als auch $\hat{\varrho}_n(h) = \hat{\gamma}_n(h)/\hat{\gamma}_n(0)$ \sqrt{n} -konsistente Schätzer für $\gamma(h)$ und $\varrho(h) = \gamma(h)/\gamma(0)$.

Beweis: O.B.d.A. sei $\mu = 0$. Dann gilt für $n > h$ mit $R_n(h) := \hat{\gamma}_n(h) - \tilde{\gamma}_n(h)$

$$\begin{aligned} R_n(h) &= -\frac{1}{n} \bar{X}_n \sum_{t=1}^{n-h} (X_t + X_{t+h}) + \frac{n-h}{n} \bar{X}_n^2 - \frac{1}{n} \sum_{t=n-h+1}^n X_t X_{t+h} \\ &= -\left(1 + \frac{h}{n}\right) \bar{X}_n^2 + \frac{1}{n} \bar{X}_n \left(\sum_{t=1}^h X_t + \sum_{t=n-h+1}^n X_t \right) - \frac{1}{n} \sum_{t=n-h+1}^n X_t X_{t+h}, \end{aligned}$$

also mit der Cauchy-Schwarz-Ungleichung

$$E |\hat{\gamma}_n(h) - \tilde{\gamma}_n(h)| \leq 2E \bar{X}_n^2 + \frac{2h}{n} \sqrt{E \bar{X}_n^2} \sqrt{\gamma(0)} + \frac{h}{n} \gamma(0). \quad (10.17)$$

Da nach Bemerkung 10.2, iv) γ_X absolut summierbar ist, gilt (10.8) und damit (10.12). Unmittelbar folgen daraus (10.13) und (10.14). Bevor wir (10.15) beweisen, benötigen wir einige vorbereitende Rechnungen. Wegen

$$E e_i e_j e_k e_l = \begin{cases} E e_1^4 & , \text{ falls } i = j = k = l \\ \sigma_e^4 & , \text{ falls zwei Paare von Indizes auftreten} \\ 0 & , \text{ sonst} \end{cases} \quad (10.18)$$

ergibt sich zunächst

$$\begin{aligned} & E(X_t X_{t+h} X_s X_{s+k}) \\ &= \sum_{\nu_1, \nu_2=-\infty}^{\infty} \sum_{\mu_1, \mu_2=-\infty}^{\infty} b_{\nu_1} b_{\nu_2} b_{\mu_1} b_{\mu_2} E(e_{t-\nu_1} e_{t+h-\nu_2} e_{s-\mu_1} e_{s+k-\mu_2}) \\ &= \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} b_{\nu+s-t} b_{\nu+s-t+k} (E e_1^4 - 3\sigma_e^4) \\ &\quad + \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+s-t} \sum_{\mu=-\infty}^{\infty} b_{\mu} b_{\mu+s-t+k-h} \cdot \sigma_e^4 \\ &\quad + \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+s-t+k} \sum_{\mu=-\infty}^{\infty} b_{\mu} b_{\mu+s-t-h} \cdot \sigma_e^4 \\ &\quad + \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} \sum_{\mu=-\infty}^{\infty} b_{\mu} b_{\mu+k} \cdot \sigma_e^4. \end{aligned} \quad (10.19)$$

Hieraus folgt mit

$$\gamma(h) = E \sum_{\nu, \mu=-\infty}^{\infty} b_{\nu} b_{\mu} e_{t-\nu} e_{t+h-\mu} = \sigma_e^2 \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} \quad (10.20)$$

$$\begin{aligned} E(X_t X_{t+h} X_s X_{s+k}) &= (\eta - 3) \sigma_e^4 \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} b_{\nu+s-t} b_{\nu+s-t+k} \quad (10.21) \\ &\quad + \gamma(s-t) \gamma(s-t+k-h) \\ &\quad + \gamma(s-t+k) \gamma(s-t-h) \\ &\quad + \gamma(h) \gamma(k). \end{aligned}$$

Damit gilt für $h, k \geq 0$

$$\begin{aligned} n \operatorname{Cov}(\tilde{\gamma}_n(h), \tilde{\gamma}_n(k)) &= n \left(E \tilde{\gamma}_n(h) \tilde{\gamma}_n(k) - E \tilde{\gamma}(h) \cdot E \tilde{\gamma}(k) \right) \\ &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E(X_t X_{t+h} X_s X_{s+k}) - n \gamma(h) \gamma(k), \end{aligned}$$

also

$$\begin{aligned} n \operatorname{Cov}(\tilde{\gamma}_n(h), \tilde{\gamma}_n(k)) &= (\eta - 3) \sigma_e^4 \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} b_{\nu+s-t} b_{\nu+s-t+k} \quad (10.22) \\ &\quad + \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \gamma(s-t) \gamma(s-t+k-h) \\ &\quad + \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \gamma(s-t+k) \gamma(s-t-h) \\ &= \sum_{r=-n+1}^{n-1} \left(1 - \frac{|r|}{n} \right) T_r \end{aligned}$$

mit

$$\begin{aligned} T_r &= (\eta - 3) \sigma_e^4 \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} b_{\nu+r} b_{\nu+r+k} \quad (10.23) \\ &\quad + \gamma(r) \gamma(r+k-h) + \gamma(r+k) \gamma(r-h) \end{aligned}$$

unter Verwendung von (A.11) aus Anhang A.1.

Aus $\sum_j |b_j| < \infty$ folgt wegen Bemerkung 10.2 iv) zunächst die absolute Summierbarkeit von $\gamma(\cdot)$ und deshalb auch $\sum_r |T_r| < \infty$. Aufgrund von Beispiel A.2 erhalten wir damit

$$nCov(\tilde{\gamma}_n(h), \tilde{\gamma}_n(k)) \rightarrow \sum_{r=-\infty}^{\infty} T_r. \quad (10.24)$$

Wegen

$$\sum_{r=-\infty}^{\infty} \sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} b_{\nu+r} b_{\nu+r+k} = \left(\sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+h} \right) \left(\sum_{\nu=-\infty}^{\infty} b_{\nu} b_{\nu+k} \right) \quad (10.25)$$

folgt nun zusammen mit (10.20) die Aussage (10.15).

Schließlich folgern wir aus (10.15) mit $h = k$ die Aussage $Var \tilde{\gamma}_n(h) = \mathcal{O}(n^{-1})$, woraus mit (10.14) $\sqrt{n}(\tilde{\gamma}_n(h) - \gamma(h)) = \mathcal{O}_P(1)$, also die gewünschte \sqrt{n} -Konsistenz folgt. Mit einer Standardargumentation erhalten wir die gleiche Aussage auch für $\tilde{\varrho}(h)$. \square

Anmerkung 10.5

Im Falle normalverteilter Innovationen (e_t) in Satz 10.4 gilt

$$Ee_t^4 = 3\sigma_e^4 \quad \text{bzw.} \quad \eta = 3, \quad (10.26)$$

sodass dann in Gleichung (10.16) der Term mit $(\eta - 3)$ wegfällt. \square

10.2 Schwache Abhängigkeit

Nachdem wir in Abschnitt 10.1 erste Konsistenzaussagen für das empirische Mittel und die empirische Autokovarianz abgeleitet haben, sollen nun zentrale Grenzwertaussagen diese Untersuchungen abrunden. Hierzu verwenden wir ein wahrscheinlichkeitstheoretisches Konzept, welches uns in einem angemessenen Rahmen derartige zentrale Grenzwertsätze für abhängige Zufallsvariablen zur Verfügung stellt. Es existieren eine ganze Reihe von möglichen Vorgehensweisen, die erfolgreich im Zeitreihenkontext angewandt wurden. Bereits das klassische Central Limit Theorem (CLT) für Martingaldifferenzschemata (vgl. Brown (1971) oder auch Gänssler/Stute (1977), Kap. IX) kann zum Nachweis der asymptotischen Normalität von Parameterschätzern im linearen autoregressiven Modell bekannter Ordnung verwendet werden. Auf Basis eines CLT für streng stationäre m -abhängige Prozesse entwickeln Brockwell/Davis (1991), vergleiche Chapter 6.4, zentrale Grenzwertaussagen für lineare Prozesse im Sinne von Definition 10.3. Schließlich erlauben die diversen mixing-Konzepte ebenfalls asymptotische Verteilungsaussagen für stochastische Prozesse (vgl. etwa Doukhan (1994) oder auch Bosq (1996)).

Während mixing-Bedingungen den Nachteil haben, dass sie aufgrund ihrer recht hohen technischen Ebene in der Regel schwierig zu überprüfen sind (vgl. Doukhan (1994)), bleiben die Ansätze über Martingalschemata oder

m -Abhängigkeitsannahmen beschränkt auf gewisse Situation. Martingaldifferenzen eignen sich besonders gut für lineare wie nicht lineare autoregressive Zeitreihen mit unabhängigen Innovationen. Für m -Abhängigkeit sehen wir kaum Möglichkeiten, dieses Konzept zum Nachweis zentraler Grenzwertaussagen weit über lineare Prozesse hinaus auszudehnen.

In diesem Text wollen wir ein von Doukhan/Louhichi (1991) entwickeltes Konzept der sogenannten schwachen Abhängigkeit aufgreifen, da wir auf dieser Basis alle intendierten Modellsituationen einheitlich behandeln können. Außerdem basiert die wesentliche Forderung der schwachen Abhängigkeit auf einer für die Zeitreihenanalyse recht gängigen Konzeption, nämlich der der Kovarianz. Grob gesprochen wird gefordert, dass für geeignete Funktionen g und h die Kovarianz

$$\text{Cov}(g(\text{Vergangenheit}), h(\text{Zukunft})) \quad (10.27)$$

für wachsenden zeitlichen Abstand zwischen „Vergangenheit“ und „Zukunft“ gegen Null konvergiert.

Wir wollen diese Idee zunächst am Beispiel der linearen Prozesse gemäß Definition 10.3 demonstrieren, wobei wir zusätzlich die Kausalität der Zeitreihe voraussetzen wollen. Dazu bezeichnen wir mit

$$|(x_1, \dots, x_u)| := (x_1^2 + \dots + x_u^2)^{1/2} \quad (10.28)$$

die Euklidische Norm auf \mathbb{R}^u und führen für eine beliebige Funktion $h : \mathbb{R}^u \rightarrow \mathbb{R}$ mit $\text{Lip}(h) := \sup_{x \neq y} (|h(x) - h(y)|/|x - y|)$ die Lipschitz-Konstante ein.

Satz 10.6 (Schwache Abhängigkeit für lineare Prozesse)

Für einen kausalen linearen Prozess $X = (X_t : t \in \mathbb{Z})$ wie in Definition 10.3 mit der zusätzlichen Koeffizientenbedingung $\sum_{k=1}^{\infty} k |b_k| < \infty$ ist die folgende schwache Abhängigkeitsaussage erfüllt:

Für $s_1 < \dots < s_u < s_u + r \leq t_1 < \dots < t_v$ und Funktionen $g : \mathbb{R}^u \rightarrow \mathbb{R}$ $h : \mathbb{R}^v \rightarrow \mathbb{R}$ mit $Eg^2(X_{s_1}, \dots, X_{s_u}), Eh^2(X_{t_1}, \dots, X_{t_v}) < \infty$ gilt:

$$\begin{aligned} & |\text{Cov}(g(X_{s_1}, \dots, X_{s_u}), h(X_{t_1}, \dots, X_{t_v}))| \\ & \leq \sqrt{Eg^2(X_{s_1}, \dots, X_{s_u})} \cdot \text{Lip}(h) \cdot \sqrt{v} \cdot \vartheta_r, \end{aligned} \quad (10.29)$$

wobei $(\vartheta_r : r \geq 1)$ eine monoton nichtsteigende summierbare Folge ist.

Beweis: Zum Nachweis von (10.29) ersetzen wir die X_{t_j} zunächst durch die von X_{s_1}, \dots, X_{s_u} unabhängigen Größen (man beachte, dass aufgrund der Kausalität in der $MA(\infty)$ -Darstellung der Zeitreihe $b_j = 0$ für alle $j < 0$ gilt) $\tilde{X}_t := \sum_{\nu=0}^{t-s_u-1} b_\nu e_{t-\nu}$, $t > s_u$, und erhalten

$$\begin{aligned} & |\text{Cov}(g(X_{s_1}, \dots, X_{s_u}), h(X_{t_1}, \dots, X_{t_v}))| \\ & = \left| \text{Cov}(g(X_{s_1}, \dots, X_{s_u}), h(X_{t_1}, \dots, X_{t_v}) - h(\tilde{X}_{t_1}, \dots, \tilde{X}_{t_v})) \right|. \end{aligned} \quad (10.30)$$

Mit Hilfe von $|Cov(U, V)| \leq \sqrt{EU^2EV^2}$ erhalten wir für die obige Kovarianz die folgende obere Schranke

$$\sqrt{Eg^2(X_{s_1}, \dots, X_{s_u})} \cdot Lip(h) \cdot \left(E \sum_{j=1}^v (X_{t_j} - \tilde{X}_{t_j})^2 \right)^{1/2}. \quad (10.31)$$

Einsetzen der MA-Darstellung für X_{t_j} und $\tilde{X}_{t_j}, j = 1, \dots, v$, ergibt dann

$$\begin{aligned} & \left(E \sum_{j=1}^v (X_{t_j} - \tilde{X}_{t_j})^2 \right)^{1/2} \\ &= \left(\sum_{j=1}^v \sum_{k=t_j-s_u}^{\infty} b_k^2 E e_{t_j-k}^2 \right)^{1/2} \leq \sqrt{v E e_1^2} \left(\sum_{k=r}^{\infty} b_k^2 \right)^{1/2} \\ &\leq \sqrt{v E e_1^2} \sum_{k=r}^{\infty} |b_k| =: \sqrt{v} \cdot \vartheta_r, \end{aligned} \quad (10.32)$$

wegen $t_j - s_u \geq r$ für alle j .

Damit gilt also (10.29). Die behauptete Summierbarkeit der ϑ_r folgt wegen

$$\sum_{r=1}^{\infty} \sum_{k=r}^{\infty} |b_k| = \sum_{k=1}^{\infty} |b_k| \sum_{r=1}^{\infty} 1\{k \geq r\} = \sum_{k=1}^{\infty} k |b_k| \quad (10.33)$$

direkt aus der Annahme $\sum_{k=1}^{\infty} k |b_k| < \infty$. \square

Die Eigenschaft (10.29) ist der Urprungsdefinition von Doukhan/Louhichi (1991) sehr ähnlich. Wir verzichten auf ihre exakte Formulierung, da wir im nachfolgenden zentralen Grenzwertsatz eine darauf zugeschnittene Variante des schwachen Abhängigkeitsbegriffs verwenden.

Bemerkung 10.7

i) Setzt man in Satz 10.6 zusätzlich voraus, dass die MA-Koeffizienten b_r in geometrischer Rate gegen Null konvergieren, d.h. es gibt ein $\varrho > 1$ und ein $c < \infty$ mit $|b_r| \leq c\varrho^{-r}$ für alle $r \in \mathbb{N}_0$, dann konvergiert auch der schwache Abhängigkeitskoeffizient ϑ_r in geometrischer Rate gegen Null.

ii) Insbesondere gilt i) für kausale $ARMA(p, q)$ -Zeitreihen mit i.i.d. (e_t) . \square

10.3 Ein zentraler Grenzwertsatz für schwach abhängige Zufallsvariable

In diesem Abschnitt formulieren und beweisen wir einen Zentralen Grenzwertsatz für schwach abhängige Zeitreihen. Das Resultat ist einer Arbeit von Neumann/Paparoditis (2003) entnommen. Konvergenzen beziehen sich, falls nichts anderes gesagt wird, auf $n \rightarrow \infty$.

Satz 10.8 (CLT für schwach abhängige Zufallsvariable)

Sei $X_n = (X_{n,k} : k \in \mathbb{Z})$ für jedes $n \in \mathbb{N}$ eine streng stationäre, reelle Zeitreihe mit $EX_{n,k} = 0$ sowie $EX_{n,k}^2 \leq C < \infty \forall k, n$. Außerdem gelte eine Lindeberg-Bedingung der Form

$$\frac{1}{n} \sum_{k=1}^n E(X_{n,k}^2 1\{|X_{n,k}| \geq \varepsilon \sqrt{n}\}) \rightarrow 0, \quad \forall \varepsilon > 0, \quad (10.34)$$

sowie für ein $\sigma^2 \in [0, \infty)$ die Konvergenzaussage

$$\begin{aligned} \sigma_n^2 &:= \sum_{k \in \mathbb{Z}} E(X_{n,0} X_{n,k}) \\ &= EX_{n,0}^2 + 2 \cdot \sum_{k=1}^{\infty} E(X_{n,0} X_{n,k}) \rightarrow \sigma^2. \end{aligned} \quad (10.35)$$

Es existiere ferner eine monoton nichtsteigende und summierbare nichtnegative Folge $(\vartheta_r : r \in \mathbb{N})$ derart, dass für alle $n \geq n_0$ und für alle $s_1 < \dots < s_u < s_u + r = t_1 \leq t_2$ die beiden folgenden „schwachen Abhängigkeitsbedingungen“ gelten

(i) für alle messbaren und quadratintegrierbaren Funktionen $f : \mathbb{R}^u \rightarrow \mathbb{R}$

$$|Cov(f(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1})| \leq \sqrt{Ef^2(X_{n,s_1}, \dots, X_{n,s_u})} \cdot \vartheta_r, \quad (10.36)$$

(ii) für alle messbaren und beschränkten Funktionen $f : \mathbb{R}^u \rightarrow \mathbb{R}$

$$|Cov(f(X_{n,s_1}, \dots, X_{n,s_u}), X_{n,t_1} \cdot X_{n,t_2})| \leq \|f\|_{\infty} \cdot \vartheta_r, \quad (10.37)$$

wobei $\|\cdot\|_{\infty}$ die Supremumsnorm bezeichne.

Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{n,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2). \quad (10.38)$$

Anmerkung 10.9

Es existieren weitere zentrale Grenzwertaussagen für schwach abhängige Zufallsfolgen in der Literatur (etwa in Doukhan/Louhichi (1991) oder Coulon-Prieur/Doukhan (2000)). Wir verwenden die angegebene Version des CLT, da sie uns für die hier benötigten Zwecke besonders adäquat erscheint. \square

Der Beweis von Satz 10.8 kann bei einem ersten Lesen übersprungen werden.

Beweis: (zu Satz 10.8) Zunächst gilt aufgrund der Stationarität

$$\begin{aligned}
& \left| Var\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{n,t}\right) - \sigma^2 \right| \\
&= \left| \frac{1}{n} \sum_{s,t=1}^n E(X_{n,s} X_{n,t}) - \sigma^2 \right| \\
&\leq \left| EX_{n,0}^2 + 2 \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) E(X_{n,0} X_{n,k}) - \sigma_n^2 \right| \quad (10.39) \\
&\quad + |\sigma_n^2 - \sigma^2|, \quad \text{man beachte (A.11)},
\end{aligned}$$

$$\begin{aligned}
&\leq 2 \sum_{k=n}^{\infty} |E(X_{n,0} X_{n,k})| + 2 \sum_{k=1}^{n-1} \frac{k}{n} |E(X_{n,0} X_{n,k})| \\
&\quad + |\sigma_n^2 - \sigma^2|, \quad \text{man beachte (10.35)}. \quad (10.40)
\end{aligned}$$

Zunächst sichert die erste schwache Abhängigkeitsbedingung (10.36) mit $u = 1$ und $f(x) = x$ die Konvergenz des ersten Summanden gegen Null, da ϑ_r summierbar ist. (10.35) impliziert die Konvergenz des letzten Summanden gegen Null. Für den mittleren Ausdruck erhalten wir erneut mit (10.36) wegen $EX_{n,k}^2 \leq C$ die folgende obere Schranke

$$2\sqrt{C} \sum_{k=1}^{n-1} \frac{k}{n} \vartheta_k \rightarrow 0. \quad (10.41)$$

Die letzte Aussage folgt aufgrund der Summierbarkeit der ϑ_k zusammen mit dem Kroneckerlemma Beispiel A.2. Wir haben also gezeigt

$$\tilde{\sigma}_n^2 := Var\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{n,t}\right) = \frac{1}{n} E(X_{n,1} + \cdots + X_{n,n})^2 \rightarrow \sigma^2. \quad (10.42)$$

Im Falle $\sigma^2 = 0$ liefert dies bereits die Behauptung, sodass wir im Folgenden $\sigma^2 > 0$ annehmen werden. Wegen (10.42) reicht es nun zu zeigen, dass für $Y_{n,t} := X_{n,t}/(\tilde{\sigma}_n \sqrt{n})$ gilt

$$\sum_{t=1}^n Y_{n,t} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (10.43)$$

Wir haben aufgrund der Stationarität von $\{X_{n,k} : k \in \mathbb{Z}\}$

$$\begin{aligned}
v_{n,k} &:= \text{Var}\left(\sum_{t=1}^k Y_{n,t}\right) - \text{Var}\left(\sum_{t=1}^{k-1} Y_{n,t}\right) \\
&= \text{Var} Y_{n,k} + 2 \sum_{t=1}^{k-1} \text{Cov}(Y_{n,t}, Y_{n,k}) \\
&= \left(EX_{n,0}^2 + 2 \sum_{t=1}^{k-1} E(X_{n,k-t} X_{n,0}) \right) / (n\tilde{\sigma}_n^2) \quad (10.44) \\
&= \left(EX_{n,0}^2 + 2 \sum_{t=1}^{k-1} E(X_{n,t} X_{n,0}) \right) / (n\tilde{\sigma}_n^2) \\
&= \left(\sigma_n^2 - 2 \sum_{t=k}^{\infty} E(X_{n,t} X_{n,0}) \right) / (n\tilde{\sigma}_n^2).
\end{aligned}$$

Mit (10.35), (10.42) und der Tatsache, dass wegen (10.36)

$$\sup_n \sum_{t=k}^{\infty} |EX_{n,0} X_{n,t}| \leq \sqrt{C} \sum_{t=k}^{\infty} \vartheta_t \rightarrow 0$$

für $k \rightarrow \infty$ gilt, ergibt sich

$$nv_{n,k} \rightarrow 1 \text{ für } n, k \rightarrow \infty. \quad (10.45)$$

Deshalb existieren $n_0, k_0 \in \mathbb{N}$ mit $v_{n,k} > 0$ für alle $n \geq n_0$ und $k_0 \leq k \leq n$. Ohne Einschränkung nehmen wir $v_{n,k} \geq 0 \forall n, k$ an.

Außerdem gilt für $n \rightarrow \infty$

$$\max_{1 \leq k \leq n} v_{n,k} \rightarrow 0, \quad (10.46)$$

denn zum einen haben wir, dass $v_{n,k}$ wegen (10.45) für großes n und großes k hinreichend klein ist und zum anderen folgt für jedes feste k aus der Tatsache, dass $n\tilde{\sigma}_n^2 \rightarrow \infty$ mit (10.44) direkt $v_{n,k} \rightarrow 0$ für $n \rightarrow \infty$.

Nun bezeichnen wir für jedes $n \in \mathbb{N}$ mit $Z_{n,k}, k = 1, \dots, n$, stochastisch unabhängige und gemäß $\mathcal{N}(0, v_{n,k})$ -verteilte Zufallsvariable, die ebenfalls unabhängig von allen $Y_{n,j}, j = 1, \dots, n$, seien.

Wegen $\text{Var}(\sum_{t=1}^n Y_{n,t}) = 1$ und der Definition von $v_{n,k}$ in (10.44) gilt

$$v_{n,1} + \dots + v_{n,n} = 1. \quad (10.47)$$

Gemäß Theorem 7.1 aus Billingsley (1968) reicht es hin, die folgende Aussage

$$Eh(Y_{n,1} + \dots + Y_{n,n}) \rightarrow Eh(Z_{n,1} + \dots + Z_{n,n}) \quad (10.48)$$

für alle reellwertigen Funktionen h zu zeigen, die dreimal stetig differenzierbar mit beschränkten Ableitungen nullter bis dritter Ordnung sind. Wir werden

nun gemäß der klassischen Lindeberg-Methode ein $Y_{n,j}$ nach dem anderen durch $Z_{n,j}$ ersetzen. Dazu bezeichnen wir

$$S_{n,k} = \sum_{t=1}^{k-1} Y_{n,t} \quad \text{und} \quad T_{n,k} = \sum_{t=k+1}^n Z_{n,t}, \quad (10.49)$$

wobei hier und im Folgenden leere Summen als 0 definiert seien. Dann zerlegen wir mit

$$\Delta_{n,k} := E[h(S_{n,k} + Y_{n,k} + T_{n,k}) - h(S_{n,k} + Z_{n,k} + T_{n,k})] \quad (10.50)$$

den zu untersuchenden Ausdruck wie folgt

$$Eh(Y_{n,1} + \cdots + Y_{n,n}) - Eh(Z_{n,1} + \cdots + Z_{n,n}) = \sum_{k=1}^n \Delta_{n,k}. \quad (10.51)$$

$\Delta_{n,k}$ wird mit

$$\begin{aligned} \Delta_{n,k}^{(1)} &:= Eh(S_{n,k} + Y_{n,k} + T_{n,k}) - Eh(S_{n,k} + T_{n,k}) \\ &\quad - \frac{v_{n,k}}{2} Eh''(S_{n,k} + T_{n,k}) \end{aligned} \quad (10.52)$$

und

$$\begin{aligned} \Delta_{n,k}^{(2)} &:= Eh(S_{n,k} + Z_{n,k} + T_{n,k}) - Eh(S_{n,k} + T_{n,k}) \\ &\quad - \frac{v_{n,k}}{2} Eh''(S_{n,k} + T_{n,k}) \end{aligned} \quad (10.53)$$

gemäß $\Delta_{n,k} = \Delta_{n,k}^{(1)} - \Delta_{n,k}^{(2)}$ noch weiter zerlegt und wir zeigen für $i = 1, 2$

$$\sum_{k=1}^n \Delta_{n,k}^{(i)} \rightarrow 0. \quad (10.54)$$

Zunächst zu $\Delta_{n,k}^{(2)}$. Mit Hilfe der Taylorschen Formel erhalten wir für ein zufälliges $\varrho_{n,k} \in (0, 1)$ zunächst

$$\begin{aligned} \Delta_{n,k}^{(2)} &= E[Z_{n,k} h'(S_{n,k} + T_{n,k})] \\ &\quad + E \frac{Z_{n,k}^2}{2} [h''(S_{n,k} + \varrho_{n,k} Z_{n,k} + T_{n,k}) - h''(S_{n,k} + T_{n,k})]. \end{aligned} \quad (10.55)$$

Da $EZ_{n,k} = 0$ und $Z_{n,k}$ unabhängig von $S_{n,k}$ und $T_{n,k}$ ist, verschwindet der erste Summand in (10.55) und wir können mit dem Mittelwertsatz wie folgt abschätzen

$$\begin{aligned}
 \left| \sum_{k=1}^n \Delta_{n,k}^{(2)} \right| &\leq \mathcal{O}(1) \sum_{k=1}^n E |Z_{n,k}|^3 \\
 &= \mathcal{O}(1) \sum_{k=1}^n v_{n,k}^{3/2}, \text{ wegen } Z_{n,k} \sim \mathcal{N}(0, v_{n,k}), \quad (10.56) \\
 &= \mathcal{O}(1) \cdot \max_{1 \leq k \leq n} v_{n,k}^{1/2}, \text{ man verwende (10.47),} \\
 &\rightarrow 0, \text{ wegen (10.46).}
 \end{aligned}$$

Die Abschätzungen für $\Delta_{n,k}^{(1)}$ sind umfangreicher. Die Taylorsche Formel ergibt diesmal für ein zufälliges $\tau_{n,k} \in (0, 1)$

$$\begin{aligned}
 \Delta_{n,k}^{(1)} &= E [Y_{n,k} h'(S_{n,k} + T_{n,k})] \\
 &\quad + E \left[\frac{Y_{n,k}^2}{2} h''(S_{n,k} + \tau_{n,k} Y_{n,k} + T_{n,k}) \right] \quad (10.57) \\
 &\quad - \frac{v_{n,k}}{2} E h''(S_{n,k} + T_{n,k}).
 \end{aligned}$$

Da $Y_{n,k}$ und $S_{n,k}$ nicht unabhängig sein müssen, verschwindet der erste Term auf der rechten Seite von (10.57) nicht notwendig. Deshalb müssen wir erneut die Taylorsche Formel bemühen und erhalten für zufälliges $\delta_{n,k,j} \in (0, 1)$

$$\begin{aligned}
 &E [Y_{n,k} h'(S_{n,k} + T_{n,k})] \\
 &= \sum_{j=1}^{k-1} E Y_{n,k} [h'(S_{n,j+1} + T_{n,k}) - h'(S_{n,j} + T_{n,k})] \quad (10.58) \\
 &= \sum_{j=1}^{k-1} E [Y_{n,k} Y_{n,j} h''(S_{n,j} + \delta_{n,k,j} Y_{n,j} + T_{n,k})].
 \end{aligned}$$

Für das erste Gleichheitszeichen beachte man, dass $E(Y_{n,k} h'(S_{n,1} + T_{n,k})) = E(Y_{n,k} h'(T_{n,k})) = 0$ gilt, da $EY_{n,k} = 0$ und $Y_{n,k}$ und $T_{n,k}$ stochastisch unabhängig sind. Setzen wir (10.58) in den obigen Ausdruck für $\Delta_{n,k}^{(1)}$ ein und beachten die Gleichung $v_{n,k} = EY_{n,k}^2 + 2 \sum_{j=1}^{k-1} E(Y_{n,k} Y_{n,j})$ (vgl. (10.44), so erhalten wir

$$\begin{aligned}
 \Delta_{n,k}^{(1)} &= \sum_{j=1}^{k-1} E Y_{n,k} Y_{n,j} [h''(S_{n,j} + \delta_{n,k,j} Y_{n,j} + T_{n,k}) - E h''(S_{n,k} + T_{n,k})] \\
 &\quad + \frac{1}{2} E Y_{n,k}^2 [h''(S_{n,k} + \tau_{n,k} Y_{n,k} + T_{n,k}) - E h''(S_{n,k} + T_{n,k})] \quad (10.59) \\
 &=: \Delta_{n,k}^{(1,1)} + \frac{1}{2} \Delta_{n,k}^{(1,2)}.
 \end{aligned}$$

Wir behandeln die beiden Ausdrücke $\Delta_{n,k}^{(1,1)}$ und $\Delta_{n,k}^{(1,2)}$ getrennt und zerlegen zunächst $\Delta_{n,k}^{(1,2)}$ in die folgenden drei Terme

$$\Delta_{n,k}^{(1,2,1)} := EY_{n,k}^2 [h''(S_{n,k} + \tau_{n,k}Y_{n,k} + T_{n,k}) - h''(S_{n,k-d} + T_{n,k})] \quad (10.60)$$

$$\Delta_{n,k}^{(1,2,2)} := EY_{n,k}^2 [h''(S_{n,k-d} + T_{n,k}) - Eh''(S_{n,k-d} + T_{n,k})] \quad (10.61)$$

$$\Delta_{n,k}^{(1,2,3)} := EY_{n,k}^2 [Eh''(S_{n,k-d} + T_{n,k}) - Eh''(S_{n,k} + T_{n,k})] . \quad (10.62)$$

Dabei stellt d einen nicht von n abhängenden Wert dar, der im Rahmen der nun folgenden Abschätzung von $\Delta_{n,k}^{(1,2,2)}$ genauer spezifiziert wird. Unter erstmaliger Anwendung der zweiten schwachen Abhängigkeitsbedingung (10.37) erhalten wir die folgende Schranke für $\Delta_{n,k}^{(1,2,2)}$

$$\begin{aligned} |\Delta_{n,k}^{(1,2,2)}| &= \frac{1}{n\tilde{\sigma}_n^2} |E(X_{n,k}^2 [h''(S_{n,k-d} + T_{n,k}) - Eh''(S_{n,k-d} + T_{n,k})])| \\ &= \frac{1}{n\tilde{\sigma}_n^2} |Cov(h''(S_{n,k-d} + T_{n,k}), X_{n,k}^2)| \\ &\leq \frac{1}{n\tilde{\sigma}_n^2} \|h''\|_\infty \cdot \vartheta_d . \end{aligned} \quad (10.63)$$

Wegen $\tilde{\sigma}_n^2 \rightarrow \sigma^2$, vgl. (10.42), erreichen wir zu gegebenem $\eta > 0$ für alle d hinreichend groß

$$\sum_{k=1}^n \Delta_{n,k}^{(1,2,2)} = \mathcal{O}(\vartheta_d) \leq \eta . \quad (10.64)$$

Von nun an sei d mit dieser Eigenschaft fest gewählt.

Mit Hilfe der Lindeberg-Bedingung (10.34) und der Beschränktheit von h'' und h''' erhalten wir weiter für beliebiges $\varepsilon > 0$

$$\begin{aligned} \left| \sum_{k=1}^n \Delta_{n,k}^{(1,2,1)} \right| &\leq \sum_{k=1}^n E[Y_{n,k}^2 1\{|X_{n,k}| \geq \varepsilon\sqrt{n}\} \cdot \mathcal{O}(1)] \\ &\quad + \sum_{k=1}^n E\left[Y_{n,k}^2 1\{|X_{n,k}| < \varepsilon\sqrt{n}\} \sum_{j=k-d}^k |Y_{n,j}| \cdot \mathcal{O}(1)\right] \\ &= \frac{\mathcal{O}(1)}{n\tilde{\sigma}_n^2} \cdot \sum_{k=1}^n E[X_{n,k}^2 1\{|X_{n,k}| \geq \varepsilon\sqrt{n}\}] \\ &\quad + \frac{\mathcal{O}(\varepsilon)}{\tilde{\sigma}_n} \cdot \sum_{k=1}^n E\left[|Y_{n,k}| \sum_{j=k-d}^k |Y_{n,j}|\right] \\ &= o(1) + \frac{\mathcal{O}(\varepsilon)}{\tilde{\sigma}_n} \sum_{k=1}^n \sum_{j=k-d}^k (EY_{n,k}^2 EY_{n,j}^2)^{1/2} , \text{ man beachte (10.34) ,} \\ &= o(1) + \frac{\mathcal{O}(\varepsilon)}{\tilde{\sigma}_n^3} (d+1) \sup_{n,j} EX_{n,j}^2 \\ &= o(1) + d\mathcal{O}(\varepsilon) \leq \eta , \end{aligned} \quad (10.65)$$

für genügend großes n und genügend kleines $\varepsilon > 0$. Wegen

$$\begin{aligned} & |Eh''(S_{n,k-d} + T_{n,k}) - Eh''(S_{n,k} + T_{n,k})| \\ & \leq \mathcal{O}(1) \cdot E \sum_{j=k-d}^{k-1} |Y_{n,j}| = d \mathcal{O}\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (10.66)$$

folgt schließlich für $\Delta_{n,k}^{(1,2,3)}$

$$\left| \Delta_{n,k}^{(1,2,3)} \right| = d \mathcal{O}\left(n^{-\frac{3}{2}}\right). \quad (10.67)$$

Indem wir zunächst $n \rightarrow \infty$ und danach $\eta \rightarrow 0$ betrachten, ergibt sich insgesamt $\sum_{k=1}^n \Delta_{n,k}^{(1,2)} \xrightarrow{n \rightarrow \infty} 0$.

Beim abschließenden Nachweis der Konvergenz von $\sum_{k=1}^n \Delta_{n,k}^{(1,1)}$ gegen Null gehen wir in gewisser Weise analog vor. Erschwert wird die Untersuchung dadurch, dass noch eine zusätzliche Summe über den Index j auftritt. Wir müssen hier erneut aufsplitten, wobei der erste Teil der Summe mit Hilfe der ersten schwachen Abhängigkeitsbedingung (10.36) wie folgt abgeschätzt werden kann

$$\begin{aligned} & \left| \sum_{j=1}^{k-d} EY_{n,k} Y_{n,j} [h''(S_{n,j} + \delta_{n,k,j} Y_{n,j} + T_{n,k}) - Eh''(S_{n,k} + T_{n,k})] \right| \\ & = \mathcal{O}\left(\frac{1}{n}\right) \sum_{j=1}^{k-d} \vartheta_{k-j} = \mathcal{O}\left(\frac{1}{n}\right) \sum_{j=d}^{\infty} \vartheta_j. \end{aligned} \quad (10.68)$$

Umfangreicher ist die Untersuchung des Anteils

$$\sum_{j=k-d+1}^{k-1} EY_{n,k} Y_{n,j} [h''(S_{n,j} + \delta_{n,k,j} Y_{n,j} + T_{n,k}) - Eh''(S_{n,k} + T_{n,k})]. \quad (10.69)$$

Analog zur Untersuchung von $\Delta_{n,k}^{(1,2)}$ zerlegen wir auch diesen Ausdruck in drei Teilsummen. Exemplarisch betrachten wir hier nur den Term

$$\left| \sum_{j=k-d+1}^{k-1} EY_{n,k} Y_{n,j} [h''(S_{n,j-d} + T_{n,k}) - Eh''(S_{n,j-d} + T_{n,k})] \right|. \quad (10.70)$$

Unter Anwendung der zweiten schwachen Abhängigkeitsbedingung (10.37) erhalten wir hierfür die folgende obere Schranke (vgl. mit der Rechnung bei (10.63))

$$\mathcal{O}(n^{-1} d \vartheta_d). \quad (10.71)$$

Ohne auf weitere Details einzugehen erhalten wir insgesamt

$$\left| \sum_{k=1}^n \Delta_{n,k}^{(1,1)} \right| \leq \mathcal{O} \left(\sum_{j=d}^{\infty} \vartheta_j \right) + \mathcal{O}(d \vartheta_d) . \quad (10.72)$$

Mit $d \rightarrow \infty$ konvergiert der erste Summand der Schranke aufgrund der Summierbarkeit der Folge $\{\vartheta_r\}$ gegen Null. Diese Summierbarkeit in Verbindung mit der Monotonie von $\{\vartheta_r\}$ sichert abschließend auch $d \cdot \vartheta_d \rightarrow_{n \rightarrow \infty} 0$ (vgl. Aufgabe 10.1).

Damit ist Satz 10.8 bewiesen. \square

Bevor wir uns den angestrebten statistischen Untersuchungen zuwenden, wollen wir als Beispiel zeigen, dass für sogenannte m -abhängige Zeitreihen die schwachen Abhängigkeitsvoraussetzungen des zentralen Grenzwertsatzes Satz 10.8 erfüllt sind.

Definition 10.10 (*m -Abhängigkeit*)

Eine Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt m -abhängig ($m \in \mathbb{N}_0$), falls die beiden Klassen von Zufallsvariablen $(X_s : s \leq t)$ und $(X_s : s \geq t + m + 1)$ für alle $t \in \mathbb{Z}$ stochastisch unabhängig sind.

Bemerkung 10.11

Für eine streng stationäre Zeitreihe $X = (X_t : t \in \mathbb{Z})$ folgt die m -Abhängigkeit bereits, falls für ein $t_0 \in \mathbb{Z}$ die Klassen $(X_s : s \leq t_0)$ und $(X_s : s \geq t_0 + m + 1)$ stochastisch unabhängig sind. \square

Offensichtlich gelten die beiden schwachen Abhängigkeitsbedingungen (10.36) und (10.37) aus Satz 10.8 für m -abhängige Zeitreihen, da Unabhängigkeit bei existierenden entsprechenden Momenten die Unkorreliertheit impliziert. Wir erhalten demnach sogar $\vartheta_r = 0$ für alle $r > m$.

Als direkte Konsequenz ergibt sich, dass $MA(q)$ -Zeitreihen mit stochastisch unabhängigem weißen Rauschen stets schwach abhängig sind, da sie q -abhängig sind.

10.4 Asymptotische Normalität des Stichprobenmittels und der Stichprobenautokovarianz

Wir beginnen diesen Abschnitt mit der Untersuchung der asymptotischen Eigenschaften des arithmetischen Mittels $\bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t$ als Schätzer für den Erwartungswert $EX_t = \mu$ einer stationären Zeitreihe. In Abschnitt 10.1 haben wir bereits gesehen, dass unter recht allgemeinen und schwachen Bedingungen an die Autokovarianzfunktion der unterliegenden Zeitreihe Konsistenz, d.h. $\bar{X}_n \rightarrow \mu$ nach Wahrscheinlichkeit, gilt. Der folgende Satz ergänzt diese Aussage in Richtung auf die asymptotische Verteilung von \bar{X}_n .

Satz 10.12 (*Asymptotische Normalität des Stichprobenmittels*)

Sei $X = (X_t : t \in \mathbb{Z})$ ein kausaler linearer Prozess mit $X_t = \mu + \sum_{j=0}^{\infty} b_j e_{t-j}$ wie in Definition 10.3 mit der zusätzlichen Voraussetzung $\sum_{j=1}^{\infty} j|b_j| < \infty$. Dann gilt

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (X_t - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \gamma_X(h)\right), \quad (10.73)$$

wobei γ_X die Autokovarianzfunktion der Zeitreihe X bezeichnet.

Beweis: Wir weisen im Folgenden die Voraussetzungen des CLT für schwach abhängige Zufallsvariablen (Satz 10.8) nach. Zunächst gilt unter den gemachten Voraussetzungen $E(X_t - \mu) = 0$ und $E(X_t - \mu)^2 \leq C < \infty$ für alle $t \in \mathbb{Z}$. Nach Bemerkung 10.2 iv) sind die Autokovarianzen $\gamma_X(h)$ absolut summierbar, sodass (10.35) erfüllt ist mit $\sigma^2 = \sum_{h \in \mathbb{Z}} \gamma_X(h)$. Die Forderung (10.34), d.h.

$$\frac{1}{n} \sum_{t=1}^n E(X_t - \mu)^2 \mathbf{1}\{|X_t - \mu| \geq \varepsilon \sqrt{n}\} \xrightarrow{n \rightarrow \infty} 0, \quad (10.74)$$

folgt direkt aus der strengen Stationarität (deshalb sind alle Summanden gleich) und der Quadratintegrierbarkeit der X_t mit Hilfe des Satzes von Lebesgue. Die erste Abhängigkeitsbedingung (10.36) folgt direkt aus Satz 10.6 mit $v = 1$ und der Identität $h(x) = x$. Der Beweis der zweiten Abhängigkeitsbedingung (10.37) folgt prinzipiell dem Beweisverlauf von Satz 10.6 (siehe auch den späteren Beweis von Satz 10.15) und ist als Aufgabe 10.2 gestellt, sodass damit Satz 10.12 bewiesen ist. \square

In Aufgabe 10.5 wird eine Methode angegeben, wie man die Aussage (10.73) im obigen Satz ohne die zusätzliche Voraussetzung $\sum_j j|b_j| < \infty$ beweisen kann.

Bemerkung 10.13 (Konfidenzintervalle für den Erwartungswert)

Wir können für kausale lineare Zeitreihen $X_t = \mu + \sum_{j=0}^{\infty} b_j e_{t-j}$ wie in Satz 10.12 die Varianz der asymptotischen Verteilung des arithmetischen Mittels (vgl. (10.73)) auch wie folgt ausdrücken. Da $\gamma_X(h) = \sigma_e^2 \sum_{j=-\infty}^{\infty} b_j b_{j+h}$ (man setze dabei stets $b_k = 0$ für $k < 0$) folgt einerseits

$$\begin{aligned} \sum_{h=-\infty}^{\infty} \gamma_X(h) &= \sigma_e^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j b_{j+h} \\ &= \sigma_e^2 \sum_{j=-\infty}^{\infty} b_j \sum_{h=-\infty}^{\infty} b_{j+h} \\ &= \sigma_e^2 \left(\sum_{j=0}^{\infty} b_j \right)^2 \end{aligned} \quad (10.75)$$

und andererseits mit (3.14) aus der Inversionsformel

$$\sum_{h=-\infty}^{\infty} \gamma_X(h) = 2\pi f_X(0) \quad (10.76)$$

wobei f_X die Spektraldichte der Zeitreihe X bezeichnet.

Die asymptotische Normalität aus (10.73) ergibt dann schließlich folgende Konfidenzintervalle für $EX_t = \mu$ zum asymptotischen Niveau $1 - \alpha$

$$\bar{X}_n \pm \frac{\sqrt{2\pi f_X(0)}}{\sqrt{n}} z_{1-\alpha/2} \quad (10.77)$$

beziehungsweise

$$\bar{X}_n \pm \frac{\left| \sum_{j=0}^{\infty} b_j \right|}{\sqrt{n}} \sigma_e z_{1-\alpha/2}, \quad (10.78)$$

wobei $z_{1-\alpha/2}$ wie üblich das $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung bezeichnet. In der konkreten Anwendung müssen die relevanten Größen $f_X(0)$, $\sum_{j=0}^{\infty} b_j$ sowie σ_e^2 natürlich durch konsistente Schätzer ersetzt werden, was unter Berücksichtigung des Lemmas von Slutsky (siehe Anhang A.5, Bemerkung A.26 (ii)) das asymptotische Niveau $1 - \alpha$ nicht beeinflusst. \square

Beispiel 10.14

Gemäß (10.78) erhalten wir für den nicht zentrierten $AR(1)$ -Prozeß $X_t + \mu$ mit $X_t + aX_{t-1} = e_t$, $t \in \mathbb{Z}$, wobei $\mu \in \mathbb{R}$, $|a| < 1$ und (e_t) i.i.d., das asymptotische $(1 - \alpha)$ -Konfidenzintervall für $EX_t = \mu$

$$\bar{X}_n \pm \frac{\sigma_e}{|1 + a|} \frac{z_{1-\alpha/2}}{\sqrt{n}} \quad (10.79)$$

Zur Verifikation beachte man die zugehörige kausale $MA(\infty)$ -Darstellung (2.16) einer $AR(1)$ -Zeitreihe.

Man erkennt, dass das Konfidenzintervall für $EX_t = \mu$ für a in der Nähe von -1 sehr groß und der Schätzer \bar{X}_n damit weniger vertrauenswürdig wird. \square

Nun wollen wir die asymptotische Verteilung der Stichprobenautokovarianz, auch empirische Autokovarianz genannt, herleiten. Als Folgerung erhalten wir später auch die asymptotische Verteilung der empirischen Autokorrelation und der schon in Kapitel 8 eingeführten Yule-Walker-Schätzer im autoregressiven Modell.

Satz 10.15 (Asymptotische Normalität der empirischen Autokovarianz)

$X = (X_t : t \in \mathbb{Z})$ sei wie in Satz 10.12 eine kausale lineare Zeitreihe, nun jedoch unter der zusätzlichen Annahme endlicher vierter Momente des weißen

Rauschens, d.h. mit $Ee_t^4 < \infty$, und erneut mit der Koeffizientenbedingung $\sum_{j=1}^{\infty} j |b_j| < \infty$. Dann gilt für jedes $M \in \mathbb{N}_0$ mit der $(M+1) \times (M+1)$ -Kovarianzmatrix $\mathbf{S}_M = (c(h, k))_{h,k=0,\dots,M}$, mit $c(h, k)$ aus (10.16)

$$\sqrt{n} (\tilde{\gamma}_n(h) - \gamma_X(h))_{h=0,\dots,M} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{S}_M). \quad (10.80)$$

Dabei ist $\tilde{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n)$, $h \geq 0$, die empirische Autokovarianzfunktion und γ_X die tatsächliche Autokovarianzfunktion.

Beweis: Wir beschränken uns auf den Fall $EX_t = 0$. Wegen (10.14) reicht es nach dem Lemma von Slutsky (siehe Bemerkung A.26) die Aussage des Satzes für $\tilde{\gamma}_n$ zu zeigen. Dazu wenden wir die Cramér-Wold Technik an (siehe Satz A.25) und zeigen für jedes $\mathbf{c} = (c_0, c_1, \dots, c_M)^T \in \mathbb{R}^{M+1}$ (nicht zu verwechseln mit den obigen Größen $c(h, k)$!)

$$\sqrt{n} \sum_{h=0}^M c_h (\tilde{\gamma}_n(h) - \gamma_X(h)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{c}^T \mathbf{S}_M \mathbf{c}). \quad (10.81)$$

Zunächst gilt

$$\begin{aligned} & \sqrt{n} \sum_{h=0}^M c_h (\tilde{\gamma}_n(h) - \gamma_X(h)) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \sum_{h=0}^M c_h (X_t X_{t+h} - \gamma_X(h)) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathcal{Z}_t \end{aligned} \quad (10.82)$$

mit

$$\mathcal{Z}_t := \sum_{h=0}^M c_h (X_t X_{t+h} - \gamma_X(h)). \quad (10.83)$$

Wir wollen nun Satz 10.8 auf die streng stationäre Zeitreihe $(\mathcal{Z}_t : t \in \mathbb{Z})$ anwenden. Die Existenz der vierten Momente von e_t sichert über die $MA(\infty)$ -Darstellung die Endlichkeit von $\sup_t E\mathcal{Z}_t^2$. Mit Hilfe des Satzes von Lebesgue kann man dann die Lindeberg-Bedingung (10.34) nachweisen, worauf wir an dieser Stelle verzichten wollen. Die Größe

$$\tau^2 := \sum_{t=-\infty}^{\infty} E\mathcal{Z}_0 \mathcal{Z}_t \quad (10.84)$$

ist endlich. Wir werden nun zeigen, dass $\tau^2 = \mathbf{c}^T \mathbf{S}_M \mathbf{c}$ gilt. Insbesondere ist dieser Ausdruck nichtnegativ, sodass die Bezeichnung τ^2 gerechtfertigt ist. Mit $\sigma_n^2 = \sigma^2 = \tau^2$ folgt dann (10.35). Wir formen nun τ^2 um. Zunächst gilt

$$\begin{aligned} E\mathcal{Z}_0 \mathcal{Z}_t &= \sum_{h=0}^M \sum_{k=0}^M c_h c_k E(X_0 X_h - \gamma_X(h))(X_t X_{t+k} - \gamma_X(k)) \\ &= \mathbf{c}^T \mathbf{A}_t \mathbf{c} \end{aligned} \quad (10.85)$$

mit der Matrix $A_t = (A_t(h, k))_{h, k=0, \dots, M}$ gemäß

$$A_t(h, k) = EX_0 X_h X_t X_{t+k} - \gamma_X(h) \gamma_X(k). \quad (10.86)$$

Aufgrund von (10.21) erhalten wir

$$\begin{aligned} A_t(h, k) &= (\eta - 3) \sigma_e^4 \sum_{\nu=-\infty}^{\infty} b_\nu b_{\nu+h} b_{\nu+t} b_{\nu+t+k} \\ &\quad + \gamma_X(t) \gamma_X(t+k-h) + \gamma_X(t+k) \gamma_X(t-h). \end{aligned} \quad (10.87)$$

Summation über $t \in \mathbb{Z}$ ergibt schließlich unter Verwendung von (10.20) die gewünschte Gleichung $\tau^2 = \mathbf{c}^T \mathbf{S}_M \mathbf{c}$.

Kommen wir jetzt zur Überprüfung der beiden schwachen Abhängigkeitsbedingungen (10.36) und (10.37). Zum Nachweis von (10.36) sei $f: \mathbb{R}^u \rightarrow \mathbb{R}$ eine messbare, quadratintegrierbare Funktion, d.h. $E f^2(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}) < \infty$, sowie $s_1 < \dots < s_u < s_u + r = t$. Wir zeigen, dass für eine geeignete Konstante $C > 0$ gilt

$$|Cov(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_t)| \leq C \sqrt{E f^2(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u})} \sum_{j=r}^{\infty} |b_j|. \quad (10.88)$$

Die im Satz 10.8 geforderte Summierbarkeit der schwachen Abhängigkeitskoeffizienten folgt dann aus (10.33).

Zum Beweis von (10.88) setze man

$$\tilde{\mathcal{Z}}_t := \sum_{h=0}^M c_h \left(\sum_{j,k=0}^{r-1} b_j b_k e_{t-j} e_{t+h-k} - \gamma_X(h) \right) \quad (10.89)$$

und beachte, dass \mathcal{Z}_t gemäß (10.83) und der Linearität des Prozesses X eine analoge Darstellung für $r = \infty$ erfüllt. Zunächst sichert dann die stochastische Unabhängigkeit von $(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u})$ und $\tilde{\mathcal{Z}}_t$ die folgende Abschätzung

$$\begin{aligned} &|Cov(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_t)| \\ &= \left| Cov\left(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_t - \tilde{\mathcal{Z}}_t\right) \right| \\ &\leq \sqrt{E f^2(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u})} \cdot \sqrt{Var(\mathcal{Z}_t - \tilde{\mathcal{Z}}_t)}. \end{aligned} \quad (10.90)$$

Wegen $Var(\mathcal{Z}_t - \tilde{\mathcal{Z}}_t) \leq E(\mathcal{Z}_t - \tilde{\mathcal{Z}}_t)^2$ betrachten wir lediglich $E(\mathcal{Z}_t - \tilde{\mathcal{Z}}_t)^2$. Die expliziten Darstellungen für $\tilde{\mathcal{Z}}_t$ bzw. \mathcal{Z}_t (vgl. (10.89)) ergeben

$$\begin{aligned}
E(\mathcal{Z}_t - \tilde{\mathcal{Z}}_t)^2 &\leq E\left(\sum_{h=0}^M c_h \sum_{\max(j,k) \geq r} b_j b_k e_{t-j} e_{t+h-k}\right)^2 \quad (10.91) \\
&\leq \sum_{h=0}^M c_h^2 E\left(\sum_{\max(j,k) \geq r} b_j b_k e_{t-j} e_{t+h-k}\right)^2 \\
&\leq \sum_{h=0}^M c_h^2 E e_1^4 \left(\sum_{\max(j,k) \geq r} |b_j| |b_k|\right)^2 \\
&\leq 4 \|\mathbf{c}\|^2 E e_1^4 \left(\sum_{j=r}^{\infty} |b_j|\right)^2 \left(\sum_{k=0}^{\infty} |b_k|\right)^2,
\end{aligned}$$

wobei sich die zweite Ungleichung aus der Cauchy–Schwarz–Ungleichung für Folgen und die dritte Ungleichung mit Hilfe von $E|e_{j_1} e_{j_2} e_{j_3} e_{j_4}| \leq E e_1^4$ ergibt. Aus (10.90) und (10.91) folgt unmittelbar (10.88).

Vollständig analog behandeln wir die zweite schwache Abhängigkeitsbedingung (10.37) aus Satz 10.8. Wir erhalten für $s_1 < \dots < s_u < s_u + r = t_1 \leq t_2$ und beschränktes f

$$Cov(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_{t_1} \mathcal{Z}_{t_2}) \leq \mathcal{O}(1) \|f\|_{\infty} \cdot \sum_{j=r}^{\infty} |b_j|. \quad (10.92)$$

Zum Nachweis von (10.92) schreiben wir zunächst

$$\begin{aligned}
&|Cov(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_{t_1} \mathcal{Z}_{t_2})| \\
&= |Cov(f(\mathcal{Z}_{s_1}, \dots, \mathcal{Z}_{s_u}), \mathcal{Z}_{t_1} \mathcal{Z}_{t_2} - \tilde{\mathcal{Z}}_{t_1} \tilde{\mathcal{Z}}_{t_2})|. \quad (10.93)
\end{aligned}$$

Unter Verwendung von $|Cov(U, V)| \leq 2\|U\|_{\infty} E|V|$ für beschränktes U kann die rechte Seite weiter nach oben abgeschätzt werden durch

$$2\|f\|_{\infty} E|\mathcal{Z}_{t_1} \mathcal{Z}_{t_2} - \tilde{\mathcal{Z}}_{t_1} \tilde{\mathcal{Z}}_{t_2}|,$$

wobei der Erwartungswert unter Verwendung der Dreiecksungleichung und danach der Cauchy–Schwarz–Ungleichung weiter nach oben abgeschätzt werden kann durch

$$\begin{aligned}
&E|\mathcal{Z}_{t_1} - \tilde{\mathcal{Z}}_{t_1}| |\mathcal{Z}_{t_2}| + E|\mathcal{Z}_{t_2} - \tilde{\mathcal{Z}}_{t_2}| |\tilde{\mathcal{Z}}_{t_1}| \quad (10.94) \\
&\leq \left(E(\mathcal{Z}_{t_1} - \tilde{\mathcal{Z}}_{t_1})^2 E \mathcal{Z}_{t_2}^2\right)^{1/2} + \left(E(\mathcal{Z}_{t_2} - \tilde{\mathcal{Z}}_{t_2})^2 E \tilde{\mathcal{Z}}_{t_1}^2\right)^{1/2}.
\end{aligned}$$

Die Rechnung in (10.91) mit $r = 0$ zeigt, dass $E \mathcal{Z}_{t_2}^2$ und ähnlich $E \tilde{\mathcal{Z}}_{t_1}^2$ durch die rechte Seite von (10.91) mit $r = 0$ beschränkt sind, sodass die Behauptung direkt aus (10.91) folgt. \square

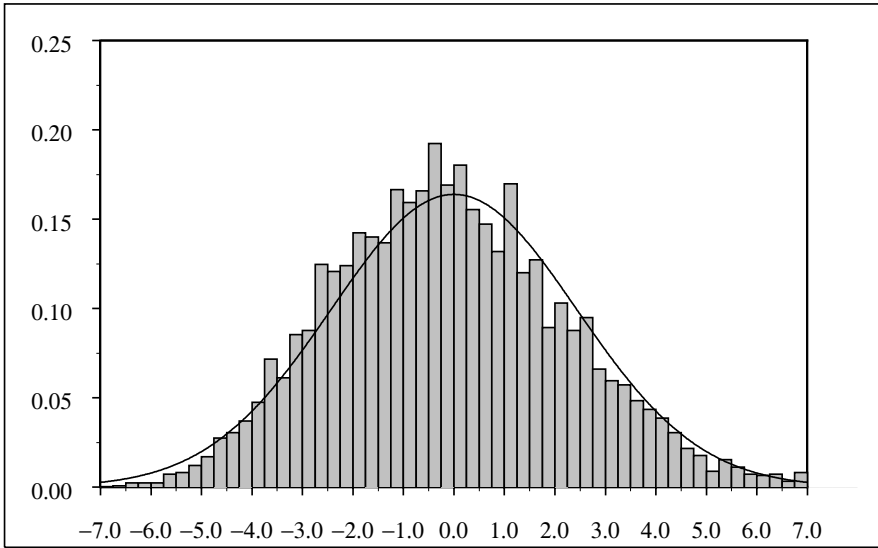


Abb. 10.1 Simuliertes Histogramm (5000 Wiederholungen) der empirischen Varianz $\sqrt{n}(\hat{\gamma}(0) - \gamma(0))$ im $AR(1)$ -Modell mit $a = 0.5$ und $n = 200$ zusammen mit der Dichte der asymptotischen Normalverteilung

Beispiel 10.16

In Abbildung 10.1 ist für eine $AR(1)$ -Zeitreihe mit Parameter $a = 0.5$ ein simuliertes Histogramm der Verteilung von $\sqrt{n}(\hat{\gamma}(0) - \gamma(0))$ angegeben. Das dargestellte Histogramm basiert auf 5000 Zeitreihen jeweils mit der Länge $n = 200$. Die zum Vergleich eingezeichnete Dichte der asymptotischen Normalverteilung $\mathcal{N}(0, 2(1+a^2)/(1-a^2)^3)$ zeigt, dass die Approximation durch die asymptotische Verteilung durchaus akzeptabel ist. Es wird aber auch deutlich, dass die grundsätzliche Streubreite beim Schätzen der Varianz der Zeitreihe recht hoch ist. Als Anhaltspunkt kann die Varianz der asymptotischen Normalverteilung betrachtet werden, die für Parameter a in der Nähe von ± 1 beliebig groß werden kann. Für $a = 0.5$ ergibt sich immerhin schon die asymptotische Varianz 5.93. Für andere Zeitreihen, etwa MA -Zeitreihen ergibt sich beim Schätzen der Varianz der Zeitreihen ein vergleichbares Bild. Grundsätzlich kann festgehalten werden, dass die Varianz einer stationären Zeitreihe in vielen Fällen nicht gut geschätzt werden kann. Wir werden später sehen (vgl. Beispiel 10.19), dass die Situation für die Autokorrelationen ganz anders aussieht.

□

Zum Abschluß dieses Kapitels wollen wir die asymptotische Normalität der empirischen Autokovarianzen $\hat{\gamma}_n(h)$ aus Satz (10.15) auf die für die Anwendung wichtigen empirischen Autokorrelationen $\hat{\varrho}_n(h)$ übertragen. Wegen

$\widehat{\varrho}_n(h) = \widehat{\gamma}_n(h)/\widehat{\gamma}_n(0)$, $h = 1, 2, \dots$, gelingt dies mit der sogenannten Delta-Methode (vgl. Satz (A.31)). Das Resultat lautet

Korollar 10.17 (*Asymptotische Normalität der emp. Autokorrelation*)

Unter den Voraussetzungen von Satz (10.15) gilt für jedes $M \in \mathbb{N}$ mit der $M \times M$ -Kovarianzmatrix $\mathbf{V}_M = (v(j, k))_{j, k=1, \dots, M}$ wobei

$$\begin{aligned} v(j, k) = & \sum_{r=-\infty}^{\infty} \{ \varrho(r+k)\varrho(r-j) \\ & + \varrho(r+j)\varrho(r+k) + 2\varrho^2(r)\varrho(j)\varrho(k) \\ & - 2\varrho(r)\varrho(r+j)\varrho(k) - 2\varrho(r)\varrho(r+k)\varrho(j) \}, \end{aligned} \quad (10.95)$$

$$\sqrt{n}(\widehat{\varrho}_n(h) - \varrho_X(h))_{h=1, \dots, M} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V}_M). \quad (10.96)$$

Beweis: Mit $f_i(x_0, x_1, \dots, x_n) = x_i/x_0$, $i = 1, \dots, M$ gilt

$$(\widehat{\varrho}_n(1), \dots, \widehat{\varrho}_n(M)) = (f_i(\widehat{\gamma}_n(0), \dots, \widehat{\gamma}_n(M)) : i = 1, \dots, M).$$

und entsprechend für ϱ_X und γ_X anstelle von $\widehat{\varrho}_n$ und $\widehat{\gamma}_n$.

Wegen $\partial f_i / \partial x_0 = -x_i/x_0^2$, $\partial f_i / \partial x_i = 1/x_0$ ($i = 1, \dots, M$) und $\partial f_i / \partial x_j = 0$ sonst, erhalten wir mit Hilfe von Satz (A.31) aus (10.80)

$$\sqrt{n}(\widehat{\varrho}_n(h) - \widehat{\varrho}_X(h))_{h=1, \dots, M} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{A}\mathbf{S}_M\mathbf{A}^T),$$

wobei $\mathbf{A} = (\partial f_i / \partial x_j(\gamma(0), \dots, \gamma(M)))_{i=1, \dots, M; j=0, \dots, M}$. Mit der Definition von $\mathbf{S}_M = (c(h, k))_{h, k=0, \dots, M}$ gemäß (10.16) erhalten wir für das (j, k) -Element von $\mathbf{A}\mathbf{S}_M\mathbf{A}^T$ den folgenden Ausdruck

$$\begin{aligned} & \frac{c(0, 0)\gamma(j)\gamma(k)}{\gamma(0)^4} - \frac{c(j, 0)\gamma(k)}{\gamma(0)^3} - \frac{c(0, k)\gamma(j)}{\gamma(0)^3} + \frac{c(j, k)}{\gamma(0)^2} \\ = & \sum_{r=-\infty}^{\infty} \{ \varrho(r+k)\varrho(r-j) + \varrho(r)\varrho(r+k-j) + 2\varrho^2(r)\varrho(j)\varrho(k) \\ & - 2\varrho(r)\varrho(r-j)\varrho(k) - 2\varrho(r)\varrho(r+k)\varrho(j) \}. \end{aligned}$$

Nun kann man durch einfache Verschiebung der Summationsvariablen r den Summanden $\varrho(r)\varrho(r+k-j)$ durch $\varrho(r+j)\varrho(r+k)$ und $-2\varrho(r)\varrho(r-j)\varrho(k)$ durch $-2\varrho(r+j)\varrho(r)\varrho(k)$ ersetzen, um die behauptete Darstellung (10.95) zu erhalten. \square

Bemerkung 10.18 (Konfidenzintervalle für die Autokorrelation)

Mit Hilfe von Korollar (10.17) können wir für jedes $h \in \mathbb{N}$ Konfidenzintervalle

für $\varrho(h)$ zum asymptotischen Niveau $1 - \alpha$ angeben. Mit dem $(1 - \alpha)$ -Quantil $z_{1-\alpha}$ der Standardnormalverteilung lauten diese

$$\hat{\varrho}_n(h) \pm \frac{\sqrt{v(h, h)}}{\sqrt{n}} z_{1-\alpha/2}, \quad h = 1, 2, \dots, \quad (10.97)$$

mit

$$\begin{aligned} v(h, h) = & \sum_{r=-\infty}^{\infty} \{ \varrho(r+h)\varrho(r-h) + \varrho^2(r) \\ & + 2\varrho^2(r)\varrho^2(h) - 4\varrho(r)\varrho(r+h)\varrho(h) \}. \end{aligned} \quad (10.98)$$

Mit Hilfe der asymptotischen Verteilungsaussage (10.96) kann man über die gerade behandelten Konfidenzintervalle für einzelne Autokorrelationen hinaus, die lediglich das entsprechende Diagonalelement der Kovarianzmatrix der asymptotischen Verteilung verwenden, auch simultane Konfidenzellipsoide für einen endlichen Vektor von Autokorrelationen angeben. \square

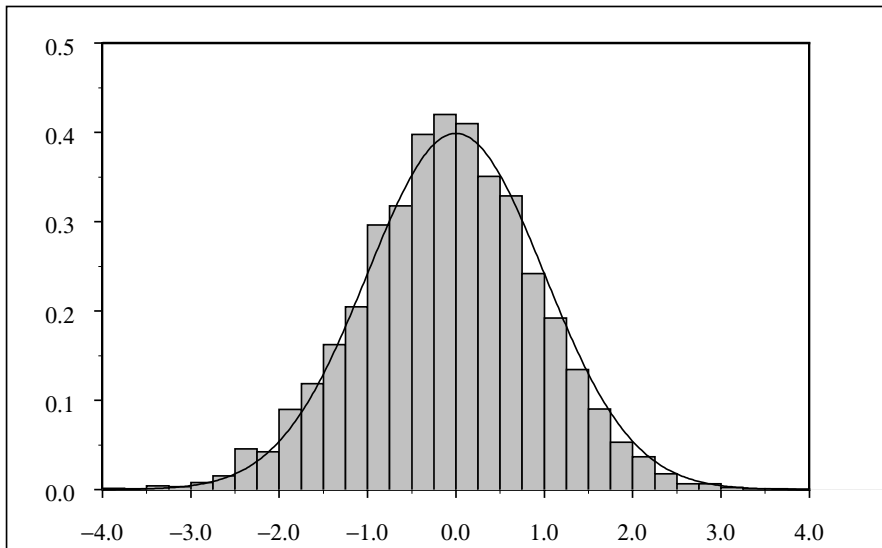


Abb. 10.2 Simuliertes Histogramm (5000 Wiederholungen) der empirischen Autokorrelation $\sqrt{n}(\hat{\varrho}(1) - \varrho(1))$ für ein weißes Rauschen der Länge $n = 200$ zusammen mit der Dichte der asymptotischen Normalverteilung

Beispiel 10.19

Für das weiße Rauschen (e_t) mit e_t i.i.d. erhalten wir nach Bemerkung (10.18) die folgenden asymptotischen $(1 - \alpha)$ -Konfidenzintervalle für $\varrho(h)$, $h \geq 1$,

$$\hat{\varrho}_n(h) \pm z_{1-\alpha/2}/\sqrt{n} \quad (10.99)$$

Mit Hilfe dieser Konfidenzintervalle können wir später für eine beobachtete Zeitreihe testen, ob es sich um ein weißes Rauschen (mit i.i.d. Zufallsvariablen) handelt.

Um beurteilen zu können, wie gut die Approximation der Verteilung der Au-

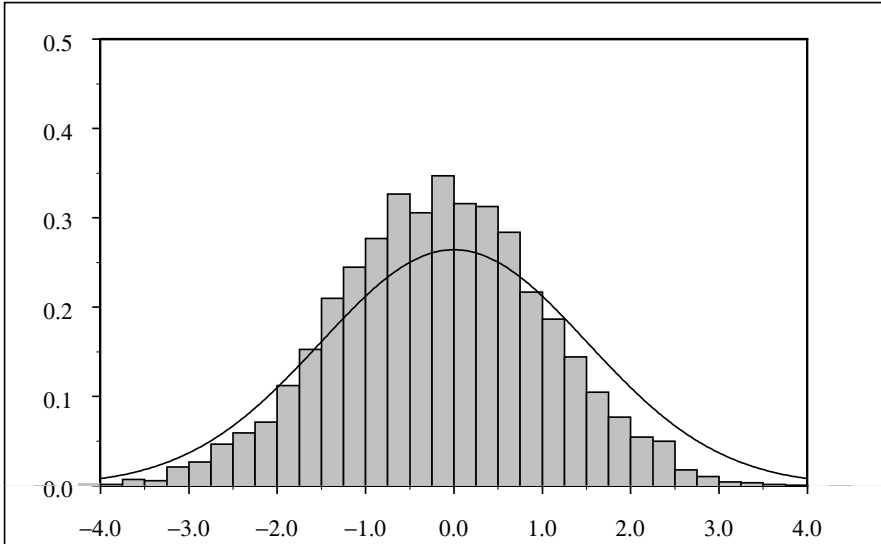


Abb. 10.3 Simuliertes Histogramm (5000 Wiederholungen) der empirischen Autokorrelation $\sqrt{n}\hat{\varrho}(2)$ für eine $MA(1)$ -Zeitreihe der Länge $n = 200$ mit Parameter $b = 0.8$ und standard normalverteiltem weißem Rauschen. Zum Vergleich ist die Dichte der asymptotischen Normalverteilung eingezeichnet

tokorrelationen durch die asymptotische Normalverteilung im Falle eines weißen Rauschens ist, geben wir in Abbildung 10.2 ein simuliertes Histogramm der Verteilung von $\sqrt{n}(\hat{\varrho}(1) - \varrho(1))$ für die Länge $n = 200$ eines weißen Rauschens mit Varianz 1 an. Die zum Vergleich eingezeichnete Dichte der zugehörigen asymptotischen Standardnormalverteilung zeigt eine gute Approximation.

Für eine $MA(q)$ -Zeitreihe mit i.i.d. weißem Rauschen erhalten wir für alle $h > q$ das folgende asymptotische $(1 - \alpha)$ -Konfidenzintervall für $\varrho(h) = 0$

$$\hat{\varrho}_n(h) \pm \sqrt{1 + 2 \sum_{r=1}^q \varrho^2(r)} \frac{z_{1-\alpha/2}}{\sqrt{n}} \quad (10.100)$$

(vgl. Aufgabe 10.3).

Auch für den MA -Fall wollen wir eine kleine Simulation betrachten. Abbildung 10.3 zeigt für das $MA(1)$ -Modell mit Parameter $b = 0.8$ und der Länge $n = 200$ die simulierte Verteilung von $\sqrt{n}(\hat{\varrho}(2) - \varrho(2))$ in Form eines Histogramms zusammen mit der Dichte der asymptotischen Normalverteilung. Für eine $AR(1)$ -Zeitreihe mit $|a| < 1$ und i.i.d. weißem Rauschen erhalten wir schließlich das asymptotische $(1 - \alpha)$ -Konfidenzintervall

$$\hat{\varrho}_n(1) \pm \sqrt{1 - a^2} \frac{z_{1-\alpha/2}}{\sqrt{n}}. \quad (10.101)$$

Sowohl in (10.100) wie in (10.101) sind $\varrho(r), r = 1, \dots, q$, bzw. a für die tatsächliche Anwendung noch durch konsistente Schätzer zu ersetzen.

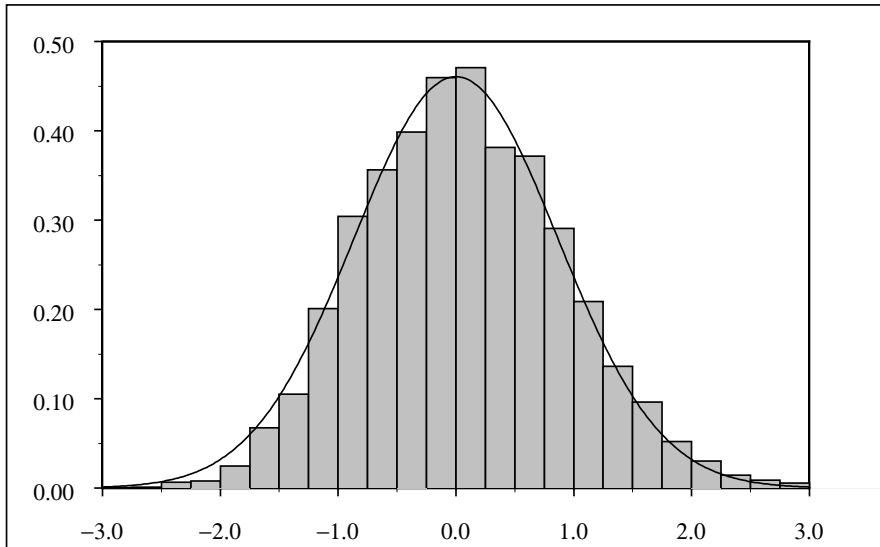


Abb. 10.4 Simuliertes Histogramm (5000 Wiederholungen) der empirischen Autokorrelation $\sqrt{n}(\hat{\varrho}(1) - \varrho(1))$ für eine $AR(1)$ -Zeitreihe der Länge $n = 200$ und mit Parameter $a = 0.5$ zusammen mit der Dichte der asymptotischen Normalverteilung $\mathcal{N}(0, 0.75)$

Abschließend wird auch noch für das $AR(1)$ -Modell Simulationsergebnisse angegeben. In Abbildung 10.4 erkennt man, dass die Autokorrelation einer $AR(1)$ -Zeitreihe weit präziser geschätzt werden kann als die Varianz (vgl. mit Beispiel 10.16). Glücklicherweise hängen geeignete Parameterschätzer im autoregressiven Modell nur von der Autokorrelationsfunktion, nicht aber von der Autokovarianzfunktion ab, wie wir im nächsten Kapitel sehen werden. Wir können also begründet hoffen, dass die Parameter autoregressiver Zeitreihen brauchbar geschätzt werden können.

Aufgaben

Aufgabe 10.1 Sei ϑ_n , $n \geq 1$, eine monoton nichtsteigende Folge nicht-negativer reeller Zahlen mit $\sum_{n=1}^{\infty} \vartheta_n < \infty$. Man zeige, dass dann auch $\lim_{n \rightarrow \infty} n\vartheta_n = 0$ gilt.

Aufgabe 10.2 Man beweise die zweite schwache Abhängigkeitsbedingung (10.37) in der Situation von Satz 10.12.

Aufgabe 10.3 (X_t) sei eine $MA(q)$ -Zeitreihe mit i.i.d. weißem Rauschen. Man bestimme die Varianz der asymptotischen Normalverteilung der empirischen Autokorrelationen $\hat{\varrho}(h)$ für $h > q$ und zeige, dass das Konfidenzintervall (10.100) asymptotisch das Konfidenzniveau einhält.

Aufgabe 10.4 Seien $Z_n, Z, Y_k, S_{kn}, D_{kn}$ Zufallsvariable mit $Z_n = S_{kn} + D_{kn} \forall k, n \in \mathbb{N}$, für die gilt

$$D_{kn} \rightarrow 0 \text{ n.W. für } k \rightarrow \infty \text{ gleichmäßig in } n \in \mathbb{N}$$

$$S_{kn} \xrightarrow{\mathcal{D}} Y_k \text{ für } n \rightarrow \infty, Y_k \xrightarrow{\mathcal{D}} Z \text{ für } k \rightarrow \infty.$$

Man zeige: $Z_n \xrightarrow{\mathcal{D}} Z$ für $n \rightarrow \infty$.

Aufgabe 10.5 Seien $e_t, t \in \mathbb{Z}$, i.i.d Zufallsvariable mit $Ee_t = 0$ und $Ee_t^2 = \sigma_e^2 \in (0, \infty)$. Man betrachte den $MA(\infty)$ -Prozess $X_t = \sum_{j=0}^{\infty} b_j e_{t-j}$ mit $\sum_{j=0}^{\infty} |b_j| < \infty$, $c := \sum_{j=0}^{\infty} b_j \neq 0$ und zeige direkt

$$Z_n := n^{-1/2} \sum_{t=1}^n X_t \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_e^2 c^2).$$

Hinweis: (1) Zunächst betrachte man den Spezialfall $b_i = 0 \forall i > k$.

(2) Für den allgemeinen Fall spalte man Z_n auf in $S_{kn} + D_{kn}$ mit

$$S_{kn} := n^{-1/2} \sum_{t=1}^n Y_{tk}, \quad D_{kn} := n^{-1/2} \sum_{t=1}^n W_{tk}$$

$$Y_{tk} := \sum_{j=0}^k b_j e_{t-j}, \quad W_{tk} := \sum_{j=k+1}^{\infty} b_j e_{t-j}$$

und verwende Aufgabe 10.4 sowie den Satz von Lindeberg-Lévy.

Aufgabe 10.6 Nach Formel (10.7) gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \sum_{k=-\infty}^{\infty} \gamma_X(k) \text{ für } n \rightarrow \infty$$

bei absolut konvergenter Kovarianzfunktion γ_X . Damit zeige man

a) Bei einer $AR(1)$ -Reihe gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \gamma_X(0) \frac{1 + \rho_X(1)}{1 - \rho_X(1)}.$$

b) Bei einer $MA(1)$ -Reihe gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \gamma_X(0)(1 + 2\rho_X(1)).$$

c) Bei einer $MA(2)$ -Reihe gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \gamma_X(0)(1 + 2\rho_X(1) + 2\rho_X(2)).$$

d) Bei einer $ARMA(1, 1)$ -Reihe gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \gamma_X(0) \left[1 + \frac{2\varrho_X^2(1)}{\varrho_X(1) - \varrho_X(2)} \right].$$

e) Bei einer $AR(2)$ -Reihe gilt

$$n \operatorname{Var} \bar{X}_n \rightarrow \gamma_X(0) \frac{(1 + \varrho_X(1))(1 - 2\varrho_X^2(1) + \varrho_X(2))}{(1 - \varrho_X(1))(1 - \varrho_X(2))}.$$

Hinweis zu e): Man verwende (8.21) für $q = 0$. Summation dieser Formel über $r = 2, 3, \dots$ gibt einen Ansatzpunkt zur Darstellung von $\sum_{k=1}^{\infty} \gamma_X(k)$ als Funktion von a_1, a_2 .

Parameterschätzung in ARMA-Modellen

In diesem Kapitel wenden wir uns der Schätzung von Modellparametern in ARMA-Zeitreihen zu. Für den Spezialfall einer rein autoregressiven Zeitreihe ist die Herleitung von Parameterschätzern und auch die Untersuchung ihrer Eigenschaften recht übersichtlich. Deshalb behandeln wir diese auch für die Anwendungen besonders wichtige Klasse in eigenen Abschnitten. Zunächst untersuchen wir in Abschnitt 11.1 die bereits mehrfach aufgetauchten *Yule-Walker*- und sogenannte *Kleinste-Quadrate (LS)*-Schätzer, bevor wir in Abschnitt 11.2 die Maximum-Likelihood-Methode für AR-Zeitreihen vorstellen. Es wird sich zeigen, dass unter der Annahmen, dass tatsächlich ein autoregressives Modell mit bekannter Ordnung vorliegt, alle drei genannten Schätzer ein asymptotisch gleiches Verhalten aufweisen, im asymptotischen Sinne also äquivalent sind.

Auch der Fall eines zugrunde liegenden autoregressiven Modells mit unendlicher Ordnung wird behandelt (Abschnitt 11.3). Hier können wir auf der Basis von Beobachtungen X_1, \dots, X_n , die wir in diesem Kapitel übrigens immer als reell voraussetzen werden, lediglich eine Parameterschätzung in einem angepassten AR-Modell mit endlicher Ordnung vornehmen und das Verhalten eines solchen Vorgehens untersuchen. Beachtet man, dass alle invertiblen ARMA-Zeitreihen eine $AR(\infty)$ -Darstellung besitzen, so studieren wir also unter anderem die Schätzung eines an das vorliegende ARMA-Modell angepassten autoregressiven Modells. Es wird sich in Abschnitt 11.4 zeigen, dass die Betrachtung der $AR(\infty)$ -Zeitreihen uns auch zu einer Möglichkeit der Schätzung der Modellparameter in ARMA-Zeitreihen selbst führt. Darüberhinaus stellen wir auch für ARMA-Zeitreihen einen Maximum-Likelihood-Ansatz vor. Es sei darauf hingewiesen, dass die Darstellungen in den Abschnitten 11.3 und 11.4, insbesondere auch was die Beweise angeht, knapp gehalten sind. Im Mittelpunkt steht nicht eine vollständige Behandlung der jeweiligen Fragestellungen sondern vielmehr eine grundlegende Einführung in die Ideen und Konstruktionsprinzipien für Parameterschätzer in autoregressiven Modellen unendlicher Ordnung sowie in ARMA-Modellen.

Bevor wir uns sinnvoll mit der statistischen Frage der Parameterschätzung auseinandersetzen können, ist grundsätzlich die Frage der Identifizierbarkeit von ARMA-Zeitreihen, d.h. die Frage, ob ein und dieselbe Zeitreihe verschiedenen ARMA-Darstellungen genügen kann, zu beantworten. Dies ist bereits in Abschnitt 8.3 geschehen, so dass wir direkt mit der Betrachtung von Schätzverfahren in autoregressiven Zeitreihen beginnen können.

11.1 Parameterschätzung für autoregressive Zeitreihen

Autoregressive Zeitreihen stellen eine in den Anwendungen weit verbreitete Klasse von Modellen dar, die zur Beschreibung von realen Daten verwendet wird. Ein wichtiger Grund hierfür liegt in der Tatsache, dass die Modellparameter von AR-Zeitreihen ausgesprochen stabil geschätzt werden können. Bereits im Abschnitt 8.3 haben wir die Yule-Walker-Gleichungen (vgl. (8.26)) kennen gelernt. Unter Verwendung der Stichprobenautokovarianz $\hat{\gamma}(h)$ (vgl. (8.27)) liegt es nahe, aus den Yule-Walker-Gleichungen mit geschätzten Autokovarianzen sogenannte Yule-Walker-Schätzer für die Parameter eines autoregressiven Zeitreihenmodells abzuleiten (vgl. mit Abschnitt 8.3). Mit leicht geänderten Bezeichnungen ergibt sich das folgende lineare Gleichungssystem (für $p \in \mathbb{N}$)

$$\hat{\Gamma}(p)\hat{\mathbf{a}}(p) + \hat{\gamma}(p) = \mathbf{0}, \quad (11.1)$$

mit $\hat{\mathbf{a}}(p) = (\hat{a}_1(p), \dots, \hat{a}_p(p))^T$, $\hat{\gamma}(p) = (\hat{\gamma}(1), \dots, \hat{\gamma}(p))^T$ und

$$\hat{\Gamma}(p) := \begin{pmatrix} \hat{\gamma}(|r-s|) \\ r, s = 1, \dots, p \end{pmatrix}. \quad (11.2)$$

Das Gleichungssystem (11.1) resultiert zwar aus Überlegungen zum AR(p)-Modell, ist aber davon losgelöst für beliebige Zufallsvariable X_1, \dots, X_n bei vorgegebenem $p \in \mathbb{N}$ einfach ein Gleichungssystem zur Ermittlung eines Vektors $\hat{\mathbf{a}}(p)$.

Falls die Matrix $\hat{\Gamma}(p)$ invertierbar ist, erhalten wir eine eindeutig bestimmte Lösung $\hat{\mathbf{a}}(p)$ des linearen Gleichungssystems (11.1). Die Frage nach der Invertierbarkeit von $\hat{\Gamma}(p)$ können wir mit der Hilfe von Satz 8.5 leicht beantworten. Es gilt nämlich (vgl. Lemma 2.11)

Lemma 11.1

Für gegebene Beobachtungen X_1, \dots, X_n bezeichne $\hat{\gamma}(h)$ gemäß (10.9) die Stichprobenautokovarianz. $\hat{\Gamma}(p)$ ist für jedes $p \in \mathbb{N}$ genau dann nicht ausgeartet, wenn $\hat{\gamma}(0) > 0$. Dabei gilt $\hat{\gamma}(0) = 0$ genau dann, wenn $X_1 = \dots = X_n$.

Beweis: Wegen $\hat{\gamma}(h) = 0$ für alle $h > n$ folgt die Invertierbarkeit von $\hat{\Gamma}(p)$ aus $\hat{\gamma}(0) > 0$ sofort mit Hilfe von Satz 8.5.

Dass $\hat{\gamma}(0) = 0$ äquivalent mit $X_1 = \dots = X_n$ ist, sieht man direkt ein. Schließlich folgt unter der Voraussetzung $X_1 = \dots = X_n$ für jedes $h \geq 0$, dass $\hat{\gamma}(h) = 0$ gilt. $\hat{\Gamma}(p)$ ist dann stets die Nullmatrix und damit ausgeartet. \square

Lemma 11.1 besagt, dass für alle in der Realität vorkommenden relevanten Fälle das Gleichungssystem (11.1) eindeutig lösbar ist. Für den Fall $\hat{\gamma}(0) > 0$, d.h. nicht alle Beobachtungen sind gleich, können wir gemäß

$$\hat{\mathbf{a}}(p) := -\hat{\Gamma}(p)^{-1}\hat{\underline{\gamma}}(p) \quad (11.3)$$

für jedes $p \in \mathbb{N}$ einen Schätzer $\hat{\mathbf{a}}(p)$ definieren. Für $\hat{\gamma}(0) = 0$ ist $\hat{\Gamma}(p)$ die Nullmatrix und $\hat{\underline{\gamma}}(p)$ der Nullvektor, sodass jeder p -Vektor $\hat{\mathbf{a}}(p)$ (11.1) löst. Zur eindeutigen Festlegung definieren wir in diesem Fall $\hat{\mathbf{a}}(p) = 0$.

Der Schätzer (11.3) ist uns schon in Kapitel 8 begegnet und heißt *Yule-Walker-Schätzer* der Ordnung p .

Um eine geschlossene Darstellung für $\hat{\mathbf{a}}(p)$ zu erhalten ist es günstig, den Begriff der verallgemeinerten Inversen zu benutzen. Wir verwenden stets die eindeutig bestimmte sogenannte Moore-Penrose-Inverse A^+ einer Matrix A (siehe Anhang A.6). A^+ stimmt mit der üblichen inversen Matrix A^{-1} überein, wenn diese existiert. Die Bildung der Moore-Penrose-Inversen ist zudem stetig auf der Menge der invertierbaren Matrizen.

In unserer speziellen Situation gilt

$$\hat{\Gamma}(p)^+ = \begin{cases} \hat{\Gamma}(p)^{-1}, & \text{falls } \hat{\gamma}(0) > 0 \\ 0, & \text{falls } \hat{\gamma}(0) = 0 \end{cases}. \quad (11.4)$$

Damit wird

$$\hat{\mathbf{a}}(p) = -\hat{\Gamma}(p)^+\hat{\underline{\gamma}}(p). \quad (11.5)$$

Sobald $\hat{\gamma}(h)$ ein konsistenter Schätzer für $\gamma(h)$ ist, z.B. nach Satz 10.15 unter den dortigen Voraussetzungen, folgt

$$\hat{\mathbf{a}}(p) \rightarrow -\Gamma(p)^{-1}\underline{\gamma}(p) \text{ n.W.} \quad \text{für } n \rightarrow \infty. \quad (11.6)$$

Aus (11.6) können wir den folgenden Satz ableiten.

Satz 11.2 (*Konsistenz der Yule-Walker-Schätzer bei AR-Reihen*)

Sei $X = (X_t : t \in \mathbb{Z})$ eine kausale $\text{AR}(p)$ -Zeitreihe mit Parametern a_1, \dots, a_p und i.i.d.weißem Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $\text{Var } e_t = \sigma_e^2 \in (0, \infty)$ und endlichen vierten Momenten $Ee_t^4 < \infty$.

Dann gilt für den Yule-Walker-Schätzer $\hat{\mathbf{a}}(p')$ gemäß (11.3) mit Ordnung $p' \geq p$

$$\hat{\mathbf{a}}(p') \rightarrow (a_1, \dots, a_p, 0, \dots, 0)^T \in \mathbb{R}^{p'} \quad \text{n. W.}$$

Der Yule-Walker-Schätzer ist also ein konsistenter Parameterschätzer für autoregressive Modelle, solange $p' \geq p$ gewählt wird. Für σ_e^2 verwenden wir den Schätzer

$$\hat{\sigma}_e^2(p') = \hat{\gamma}(0) + \sum_{j=1}^{p'} \hat{a}_j(p') \hat{\gamma}(j). \quad (11.7)$$

Für $p' \geq p$ ist dieser Schätzer ebenfalls konsistent für σ_e^2 .

Beweis: Die $MA(\infty)$ -Darstellung von X besitzt nach Bemerkung 7.8 Koeffizienten, die in geometrischer Rate abfallen, insbesondere die Koeffizientenbedingung in Satz 10.15 erfüllen, sodass als Folgerung aus diesem Satz alle $\hat{\gamma}(h)$ konsistent sind. Nach (8.3) fallen auch die $\gamma(h)$ in geometrischer Rate ab, konvergieren insbesondere gegen 0, so dass nach Satz 8.5 die Invertierbarkeit der Matrizen $\Gamma(p')$ folgt. Man beachte dabei, dass $\gamma(0) > 0$ gilt, da X nicht der Nullprozess sein kann, weil sonst nach der AR-Gleichung auch das weiße Rauschen der Nullprozess wäre, was aber definitionsgemäß ausgeschlossen ist. Insgesamt erhalten wir $\hat{\mathbf{a}}(p') \rightarrow -\Gamma(p')^{-1} \underline{\gamma}(p')$ n.W..

Nun ist $-\Gamma(p')^{-1} \underline{\gamma}(p')$ die eindeutig bestimmte Lösung des Gleichungssystems

$$\Gamma(p') \mathbf{a}(p') + \underline{\gamma}(p') = \mathbf{0}.$$

Aufgrund der Yule-Walker-Gleichungen im $AR(p)$ -Modell ist aber stets

$$\mathbf{a}(p') = (a_1, \dots, a_p, 0, \dots, 0)^T$$

ebenfalls eine Lösung. Die Eindeutigkeit liefert die Behauptung.

Zum Nachweis der Konsistenz von $\hat{\sigma}_e^2(p')$ beachte (8.25) und die Konsistenz von $\hat{\gamma}$ sowie $\hat{\mathbf{a}}(p')$. □

Nun wollen wir eine zentrale Grenzwertaussage für den Yule-Walker-Schätzer beweisen. Zum einen interessiert uns natürlich die Frage, wie der Yule-Walker-Schätzer verteilt ist, wenn tatsächlich ein autoregressives Modell vorliegt. Da jedes Zeitreihenmodell aber immer nur eine mehr oder weniger gute Approximation an die Realität darstellt, sollte man sich zum anderen auch fragen, wie sich ein etwaiger Schätzer verhält, wenn die angedachte Modellvorstellung nicht erfüllt ist. Wir werden gleich mit diesem Schritt beginnen und anschließend auf den Spezialfall einer autoregressiven Zeitreihe reduzieren.

Satz 11.3 (*Asymptotische Normalität des Yule–Walker–Schätzers*)

Sei $X = (X_t : t \in \mathbb{Z})$ eine $MA(\infty)$ -Zeitreihe mit i.i.d. weißem Rauschen $e = (e_t : t \in \mathbb{Z})$. Es gelte $Ee_t^4 < \infty$ und die Summationsbedingung $\sum_{j=0}^{\infty} j |b_j| < \infty$ für die $MA(\infty)$ -Koeffizienten. Bis auf die zusätzliche Annahme, dass der Mittelwert $EX_t = \mu = 0$ ist, sind dies dieselben Voraussetzungen wie in Satz 10.15. Außerdem sei $\gamma(0) > 0$, also X nicht der Nullprozess. Dann gilt für den Yule–Walker–Schätzer $\hat{\mathbf{a}}(p)$ zur Ordnung $p \in \mathbb{N}$

$$\sqrt{n}(\hat{\mathbf{a}}(p) - \mathbf{a}(p)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, K(p)). \quad (11.8)$$

Dabei ist $\mathbf{a}(p) := -\Gamma(p)^{-1}\underline{\gamma}(p)$ und

$$K(p) = \Gamma(p)^{-1}A(p)\mathbf{S}_pA(p)^T\Gamma(p)^{-1} \quad (11.9)$$

mit \mathbf{S}_p aus Satz 10.15 und der $p \times (p+1)$ -Matrix

$$A(p) = \begin{pmatrix} a_1(p) & 1 & 0 & 0 & \dots & 0 \\ a_2(p) & a_1(p) & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{p-1}(p) & a_{p-2}(p) & \dots & a_1(p) & 1 & 0 \\ a_p(p) & a_{p-1}(p) & \dots & \dots & a_1(p) & 1 \end{pmatrix} \quad (11.10)$$

$$+ \begin{pmatrix} 0 & a_2(p) & a_3(p) & \dots & a_{p-1}(p) & a_p(p) & 0 \\ 0 & a_3(p) & a_4(p) & \dots & a_p(p) & 0 & 0 \\ \vdots & \vdots & & & & \vdots & \vdots \\ 0 & a_p(p) & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Beweis: Die Überlegungen zu Beginn des Beweises von Satz 11.2 zeigen sowohl die Invertierbarkeit der Matrizen $\Gamma(p)$ als auch die Konsistenz der $\hat{\gamma}(h)$, sodass auch $\hat{\Gamma}(p) \rightarrow \Gamma(p)$ n.W. und $\hat{\Gamma}^+(p) \rightarrow \Gamma(p)^{-1}$ n.W. gelten. Dann gilt auf $\{\hat{\gamma}(0) > 0\}$

$$\begin{aligned} U_n &:= \sqrt{n}(\hat{\mathbf{a}}(p) - \mathbf{a}(p)) \\ &= -\sqrt{n}\hat{\Gamma}(p)^{-1}[\hat{\underline{\gamma}}(p) + \hat{\Gamma}(p)\mathbf{a}(p)] \\ &= -\sqrt{n}\hat{\Gamma}(p)^{-1}[\hat{\underline{\gamma}}(p) - \underline{\gamma}(p) + (\hat{\Gamma}(p) - \Gamma(p))\mathbf{a}(p)], \end{aligned} \quad (11.11)$$

denn es gilt $\underline{\gamma}(p) + \Gamma(p)\mathbf{a}(p) = \mathbf{0}$.

Setzt man die Komponenten der Vektoren bzw. der Matrizen explizit ein, so kann man wie folgt fortfahren

$$\begin{aligned} &= -\hat{\Gamma}(p)^{-1}\sqrt{n} \left(\hat{\gamma}(r) - \gamma(r) + \sum_{s=1}^p a_s(p)(\hat{\gamma}(|r-s|) - \gamma(|r-s|)) \right) \\ &\quad r = 1, \dots, p \\ &= -\hat{\Gamma}(p)^{-1}A(p)\sqrt{n}(\hat{\gamma}(0) - \gamma(0), \dots, \hat{\gamma}(p) - \gamma(p))^T =: V_n. \end{aligned}$$

Satz 10.15 zusammen mit dem Lemma von Slutsky, siehe Anhang A.5, ergibt wegen $\hat{\Gamma}(p)^{-1} \rightarrow \Gamma(p)^{-1}$ n.W. die Aussage, dass der Zufallsvektor V_n nach Verteilung gegen die in (11.8) angegebene Normalverteilung konvergiert. Der uns interessierende Zufallsvektor U_n unterscheidet sich nach obiger Rechnung von V_n nur auf der Menge $\{\hat{\gamma}(0) = 0\}$, für die

$$P\{\hat{\gamma}(0) = 0\} \leq P\{|\hat{\gamma}(0) - \gamma(0)| \geq \gamma(0)/2\} \rightarrow 0$$

gilt. Wiederum nach dem Lemma von Slutsky haben dann U_n und V_n dieselbe Limesverteilung. □

Um die asymptotische Kovarianzmatrix $K(p)$ für den Fall eines zugrunde liegenden AR-Modells zu berechnen, ziehen wir einen mit dem Yule-Walker-Schätzer eng zusammenhängenden Schätzer heran, nämlich den aus der Methode der kleinsten Quadrate gewonnenen sogenannten *LS* (Least-Squares)-Schätzer $\hat{\mathbf{a}}^{LS}(p) = (\hat{a}_1^{LS}, \dots, \hat{a}_p^{LS})^T$. Dieser ist definiert durch

$$\hat{\mathbf{a}}^{LS}(p) = - \left(\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s} \right)_{r,s=1,\dots,p}^+ \left(\frac{1}{n} \sum_{t=1}^n X_t X_{t-j} \right)_{j=1,\dots,p}. \quad (11.12)$$

In Abschnitt 11.2 werden wir diesen Schätzer aus dem Maximum-Likelihood-Prinzip ableiten und außerdem in Satz 11.9 nachweisen, dass er unter den Voraussetzungen von Satz 11.4 dieselbe asymptotische Normalverteilung wie der Yule-Walker-Schätzer besitzt. Die Berechnung der asymptotischen Kovarianzmatrix $K(p)$ für den Fall eines zugrunde liegenden AR-Modells ist jedoch für den *LS*-Schätzer viel einfacher als für den Yule-Walker-Schätzer, weshalb wir ihn schon hier verwenden.

Satz 11.4 (*Asym. Normalität des Yule-Walker-Schätzers bei AR-Reihen*)

$X = (X_t : t \in \mathbb{Z})$ sei eine kausale $AR(p)$ -Zeitreihe mit i.i.d. weißem Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t^2 = \sigma_e^2$ und $Ee_t^4 < \infty$. Dann gilt für jedes $p_0 \geq p$ für den Yule-Walker-Schätzer $\hat{\mathbf{a}}(p_0)$ zur Ordnung p_0

$$\sqrt{n}(\hat{\mathbf{a}}(p_0) - \mathbf{a}(p_0)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \sigma_e^2 \Gamma(p_0)^{-1}), \quad (11.13)$$

wobei $\mathbf{a}(p_0) = (a_1, \dots, a_p, 0, \dots, 0)^T \in \mathbb{R}^{p_0}$. Für $p_0 > p$ gilt für das unterste Diagonalelement der asymptotischen Kovarianzmatrix $\sigma_e^2 \Gamma(p_0)^{-1}_{p_0, p_0} = 1$, also

$$\sqrt{n} \hat{a}_{p_0}(p_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (11.14)$$

Beweis: Da die Voraussetzungen von Satz 11.4 alle Annahmen von Satz 11.3 implizieren, ist die grundsätzliche asymptotische Normalität bereits bewiesen. Es ist lediglich die asymptotische Kovarianzmatrix nachzurechnen.

Wie schon erwähnt, zeigt der spätere Satz 11.9, dass es reicht, die asymptotische Kovarianzmatrix des LS -Schätzers $\hat{\mathbf{a}}^{LS}(p_0)$ aus (11.12) zu bestimmen. Für den LS -Schätzer gilt auf $\{\det \tilde{\Gamma}(p_0) \neq 0\}$ mit der Abkürzung $\tilde{\Gamma}(p_0) = (\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s} : r, s = 1, \dots, p_0)$

$$\begin{aligned} & \sqrt{n}(\hat{\mathbf{a}}^{LS}(p_0) - \mathbf{a}(p_0)) \\ &= -\sqrt{n}\tilde{\Gamma}(p_0)^+ \left(\frac{1}{n} \sum_{t=1}^n X_t (X_{t-1}, \dots, X_{t-p_0})^T + \tilde{\Gamma}(p_0) \mathbf{a}(p_0) \right). \end{aligned} \quad (11.15)$$

Mit demselben Beweis wie in Satz 10.4 gilt mit $h := |r - s|$

$$n \operatorname{Var} \left(\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s} \right) \rightarrow c(h, h),$$

sodass $\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s} \rightarrow \gamma(|r - s|)$ n.W., also auch $\tilde{\Gamma}(p_0) \rightarrow \Gamma(p_0)$ n.W. folgt. Insbesondere ergibt sich wegen der Stetigkeit der Determinantenfunktion $P\{\det \tilde{\Gamma}(p_0) = 0\} \rightarrow 0$, sodass nach dem Lemma von Slutsky die rechte Seite in (11.15) dieselbe asymptotische Normalverteilung wie die linke Seite besitzt. Durch direktes Einsetzen von $\tilde{\Gamma}(p_0)$ können wir auf der rechten Seite in (11.15) fortfahren mit

$$= -\tilde{\Gamma}(p_0)^+ \frac{1}{\sqrt{n}} \sum_{t=1}^n \underbrace{\left(X_t + \sum_{\nu=1}^p a_\nu X_{t-\nu} \right)}_{=e_t} (X_{t-1}, \dots, X_{t-p_0})^T. \quad (11.16)$$

Da $\tilde{\Gamma}(p_0) \rightarrow \Gamma(p_0)$ n.W., folgt zunächst auch die asymptotische Normalität von $\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t (X_{t-1}, \dots, X_{t-p_0})^T$. Die Kovarianzmatrix dieser asymptotischen Normalverteilung können wir gemäß

$$\lim_{n \rightarrow \infty} \left(E \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t X_{t-r} \frac{1}{\sqrt{n}} \sum_{\nu=1}^n e_\nu X_{\nu-s} : r, s = 1, \dots, p_0 \right) \quad (11.17)$$

bestimmen, falls $E \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t (X_{t-1}, \dots, X_{t-p_0})^T \right\|^4$ endlich ist (vgl. Bemerkung A.26 (v), Anhang A.5, angewendet auf die jeweiligen Erwartungswerte in (11.17) und mit Abschätzung dieser nach oben durch vierte Momente mittels der Cauchy-Schwarz-Ungleichung). Die Ausführung der Überlegungen ist als Übungsaufgabe 11.2 gestellt.

Da für $r, s > 0$ die Gleichung

$$\begin{aligned} & E \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t X_{t-r} \frac{1}{\sqrt{n}} \sum_{\nu=1}^n e_\nu X_{\nu-s} \\ &= \frac{1}{n} \sum_{t=1}^n E(e_t^2 X_{t-r} X_{t-s}) = \sigma_e^2 \gamma(r - s) \end{aligned}$$

gilt, wobei man beachte, dass X_{t-r} und X_{t-s} aufgrund der $MA(\infty)$ -Darstellung der $AR(p)$ -Zeitreihe unabhängig von e_t sind, folgt, dass die asymptotische Kovarianzmatrix gleich $\sigma_e^2 \Gamma(p_0)$ ist. Mit dem Lemma von Slutsky ergibt sich wegen (11.16) sofort $\sqrt{n}(\hat{\mathbf{a}}^{LS}(p_0) - \mathbf{a}(p_0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_e^2 \Gamma(p_0)^{-1})$. Wie bereits oben erwähnt besitzt $\sqrt{n}(\hat{\mathbf{a}}(p_0) - \mathbf{a}(p_0))$ schließlich die gleiche asymptotische Kovarianzmatrix.

Zur Berechnung des untersten Diagonalelementes $\sigma_e^2 \Gamma(p_0)^{-1}_{p_0, p_0}$ im Fall $p_0 > p$ dieser Kovarianzmatrix gehen wir wie folgt vor. Zunächst gilt

$$\Gamma(p_0) = \left(\begin{array}{ccc|ccc} \gamma(0) & \cdots & \gamma(p-1) & \gamma(p) & \cdots & \gamma(p_0-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \cdots & \gamma(0) & \gamma(1) & \cdots & \gamma(p_0-p) \\ \hline \gamma(p) & \cdots & \gamma(1) & \gamma(0) & \cdots & \gamma(p_0-p-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(p_0-1) & \cdots & \gamma(p_0-p) & \gamma(p_0-p-1) & \cdots & \gamma(0) \end{array} \right).$$

Aufgrund der Yule-Walker-Gleichungen

$$\gamma(h) + a_1 \gamma(h-1) + \cdots + a_p \gamma(h-p) = 0, \quad h = 1, 2, \dots, p,$$

und der Symmetrie von γ kann durch elementare Zeilen- und Spaltenumformungen die folgende Darstellung erreicht werden

$$\left(\begin{array}{c|ccc} \Gamma(p) & & & \mathbf{0} \\ \hline & \gamma(0) + \sum_{t=1}^p a_t \gamma(t) & & \mathbf{0} \\ \mathbf{0} & \vdots & \ddots & \gamma(0) + \sum_{t=1}^p a_t \gamma(t) \\ & \mathbf{0} & & \end{array} \right).$$

Aus der ersten Yule-Walker-Gleichung erhalten wir (vgl. (8.25))

$$\gamma(0) + \sum_{t=1}^p a_t \gamma(t) = \sigma_e^2.$$

Damit folgt

$$\det \Gamma(p_0) = \det \Gamma(p) \cdot \sigma_e^{2(p_0-p)}.$$

Exakt die gleiche Überlegung führt zu

$$\det \Gamma(p_0 - 1) = \det \Gamma(p) \cdot \sigma_e^{2(p_0-p-1)}.$$

Da nun nach der Cramerschen Regel

$$\Gamma(p_0)^{-1}_{p_0, p_0} = \frac{\det \Gamma(p_0 - 1)}{\det \Gamma(p_0)} = \frac{1}{\sigma_e^2}$$

gilt, wobei man beachte, dass durch Streichung der letzten Zeile und letzten Spalte in $\Gamma(p_0)$ exakt $\Gamma(p_0 - 1)$ entsteht, folgt auch (11.14). \square

Bemerkung 11.5 (Asymptot. Konfidenzbereiche für die AR-Parameter)

Sowohl mit Hilfe von Satz 11.3 wie Satz 11.4 können wir nun asymptotische $(1 - \alpha)$ -Konfidenzbereiche für die Parameter eines geschätzten AR-Modells erhalten. Konkret ergibt sich etwa in der Situation von Satz 11.4, also für den Fall, dass tatsächlich ein autoregressives Modell zugrunde liegt, das folgende Konfidenzellipsoid für den Parametervektor $\mathbf{a} = (a_1, \dots, a_p)^T$ eines AR(p)-Modells

$$\left\{ \mathbf{a} \in \mathbb{R}^p : n \cdot (\hat{\mathbf{a}} - \mathbf{a})^T \hat{\Gamma}(p) (\hat{\mathbf{a}} - \mathbf{a}) \leq \hat{\sigma}_e^2 \cdot \chi_{p;1-\alpha}^2 \right\}. \quad (11.18)$$

Dabei bezeichnet $\hat{\Gamma}(p)$ einen konsistenten Schätzer für die Kovarianzmatrix $\Gamma(p)$. Es sei daran erinnert, dass die Ersetzung von $\Gamma(p)$ durch einen konsistenten Schätzer in (11.18) die asymptotische Verteilung auf Grund des Lemmas von Slutsky (vgl. Anhang A.5, (A.26)) nicht beeinflusst. Für die Schätzung von $\Gamma(p)$ kommen zum einen die empirischen Autokovarianzen in Frage. Zum anderen können wir aber auch die Autokovarianzen $\gamma(h)$ explizit mit Hilfe der Yule-Walker-Gleichungen (vgl. (8.25) und (8.26)) durch die AR-Parameter \mathbf{a} ausdrücken und dann die unbekannten Parameter durch die Schätzer $\hat{a}_1, \dots, \hat{a}_p$ ersetzen. In der Situation aus Satz 11.3 können entsprechende Konfidenzbereiche angegeben werden.

Für den Spezialfall eines AR(1)-Modells verweisen wir auf Aufgabe 11.7. \square

Bemerkung 11.6 (Asympt. Normalität der emp. part. Autokorrelation)

Aus Kapitel 2 (2.40) und (2.41) wissen wir, dass die empirische partielle Autokorrelation $\hat{\pi}(h)$ zum Lag h jeweils mit der letzten Komponente des Yule-Walker-Parameterschätzers $\hat{\mathbf{a}}(h)$ übereinstimmt, also dass

$$\hat{\pi}(h) = \hat{a}_h(h)$$

gilt. Da für kausale und stationäre autoregressive Zeitreihen der Ordnung p außerdem $\pi(h) = 0$ für $h > p$ erfüllt ist (vgl. Aufgabe 7.7), folgt unter den Voraussetzungen von Satz 11.4 aus (11.14) unmittelbar

$$\sqrt{n} \hat{\pi}(h) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \forall h > p. \quad (11.19)$$

\square

Bemerkung 11.7 (Levinson-Rekursion für Yule-Walker-Schätzer)

Wir haben eingangs gesehen, dass für die Bestimmung des Yule-Walker-Parameterschätzers $\hat{\mathbf{a}}(p)$ das lineare Gleichungssystem (11.1) zu lösen ist. Falls

$\widehat{\Gamma}(p)$ nicht singulär ist, ist die Lösung eindeutig bestimmt.

Betrachtet man die Berechnung der Koeffizienten der 1-Schritt-Vorhersage bei stationären Zeitreihen gleich zu Beginn von Kapitel 4 mit dieser Kenntnis erneut, so erkennt man, dass für die Bestimmung der Koeffizienten der Vorhersage ein identisches Gleichungssystem (4.3) zu lösen ist. Lediglich die tatsächliche Autokovarianzfunktion γ in Kapitel 4 ist durch die Stichprobenautokovarianz $\widehat{\gamma}$ (vgl. (8.28)), n ist durch p und schließlich der Vektor d_n^1 durch $-\widehat{\mathbf{a}}(p)$ zu ersetzen. Die Verwendung der Stichprobenautokovarianz ist unproblematisch, da diese ja die charakterisierenden Eigenschaften der positiven Semidefinitheit einer Autokovarianzfunktion besitzt (vgl. Aufgabe 8.3). Für die Berechnung der Koeffizienten der 1-Schritt-Vorhersage haben wir in Abschnitt 4.2 den Levinson-Algorithmus angegeben. Aufgrund der vollkommen identischen Fragestellung können wir den Levinson-Algorithmus nun auch zur effizienten Berechnung des Yule-Walker-Schätzer verwenden. Beachtet man noch, dass für den Vorhersagefehler v_p aus Abschnitt 4.2 (vgl. (4.15)) in der hiesigen Notation

$$v_p = \widehat{\gamma}(0) + \sum_{j=1}^p \widehat{a}_j(p) \widehat{\gamma}(j) = \widehat{\sigma}^2(p)$$

(vgl. (11.7)) gilt, so können wir die Ausführungen aus Abschnitt 4.2 Wort für Wort übertragen und erhalten den folgenden Algorithmus.

Mit $\widehat{\mathbf{a}}(1) = \widehat{a}_1(1) = -\widehat{\gamma}(1)/\widehat{\gamma}(0)$ und $v_0 = \widehat{\sigma}^2(0) = \widehat{\gamma}(0)$ gilt die Levinson-Rekursion zur Berechnung des Yule-Walker-Schätzers

$$\widehat{a}_p(p) = -\frac{\widehat{\gamma}(p) + \sum_{j=1}^{p-1} \widehat{a}_j(p-1) \widehat{\gamma}(p-j)}{\widehat{\sigma}^2(p-1)} \quad (11.20)$$

$$\widehat{a}_j(p) = \widehat{a}_j(p-1) + \widehat{a}_p(p) \cdot \widehat{a}_{p-j}(p-1), \quad j = 1, \dots, p-1,$$

sowie

$$\widehat{\sigma}^2(p) = \widehat{\sigma}^2(p-1) \cdot (1 - \widehat{a}_p^2(p)).$$

Damit berechnet man rekursiv

$$\widehat{\sigma}^2(0); \widehat{a}_1(1), \widehat{\sigma}^2(1); \widehat{a}_2(2), \widehat{a}_1(2), \widehat{\sigma}^2(2); \widehat{a}_3(3), \widehat{a}_2(3), \widehat{a}_1(3), \widehat{\sigma}^2(3); \dots \text{ usw.}$$

□

Beispiel 11.8

Für eine $AR(1)$ -Zeitreihe mit Parameter $a = 0.85$ wissen wir aus Satz 11.4, dass der Yule-Walker-Schätzer \widehat{a} eine asymptotische Normalverteilung besitzt. Für 5000 simulierte Zeitreihen jeweils mit der Länge $n = 200$ ist in Abbildung 11.1 ein Histogramm der sich aus den Simulationen ergebenden und mit dem tatsächlichen Wert zentrierten sowie mit \sqrt{n} standardisierten Schätzer $\sqrt{200}(\widehat{a}_i - 0.85)$, $i = 1, \dots, 5000$, zusammen mit der Dichte der

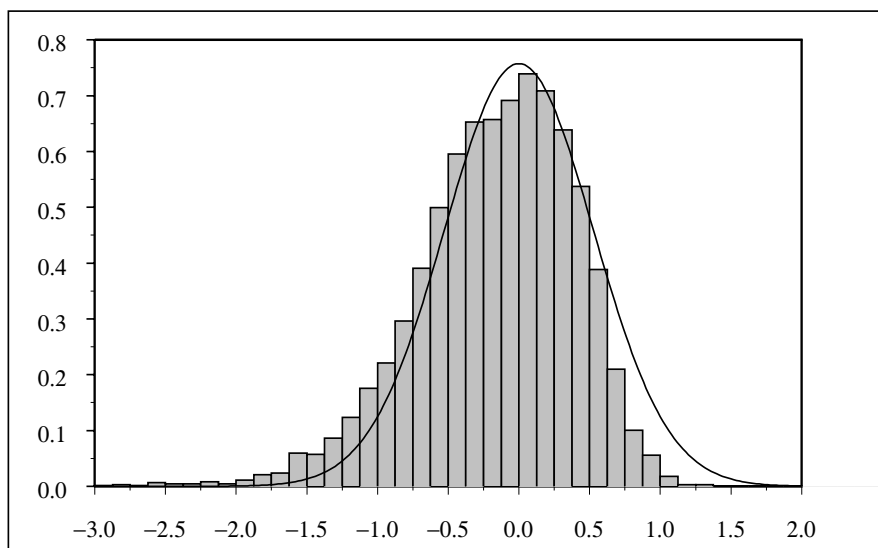


Abb. 11.1 Simuliertes Histogramm (5000 Wiederholungen) des zentrierten und standardisierten Yule–Walker-Schätzers $\sqrt{n}(\hat{a} - a)$ im $AR(1)$ -Modell mit $a = 0.85$ und $n = 200$ zusammen mit der Dichte der asymptotischen Normalverteilung $\mathcal{N}(0, 1 - a^2)$

asymptotischen Normalverteilung dargestellt. Man erkennt eine gewisse Schiefe des Histogramms und deshalb eine moderate Abweichung von der asymptotischen Normalverteilung. Die Schiefe des Histogramms resultiert aus einer Tendenz des Yule–Walker-Schätzers, den Parameter a zu unterschätzen. Dieses grundsätzliche Problem des Yule–Walker-Verfahrens und ist uns auch schon in Abbildung 2.9 begegnet.

Abbildung 11.2 enthält ein entsprechendes Simulationsergebnis für den Schätzer \hat{a}_2 eines an die Daten des obigen $AR(1)$ -Modells angepassten $AR(2)$ -Modells. Wegen (11.14) ist die asymptotische Verteilung hier eine Standardnormalverteilung.

□

11.2 Maximum–Likelihood Schätzer im autoregressiven Modell

Eine in der Statistik häufig angewandte Methode zur Konstruktion von Parameterschätzern ist die Maximum–Likelihood (*ML*)-Methode. Im Gegensatz zu stochastisch unabhängigen Zufallsvariablen, deren gemeinsame Dichte Produktgestalt besitzt, ist die explizite Angabe einer Wahrscheinlichkeitsdichte

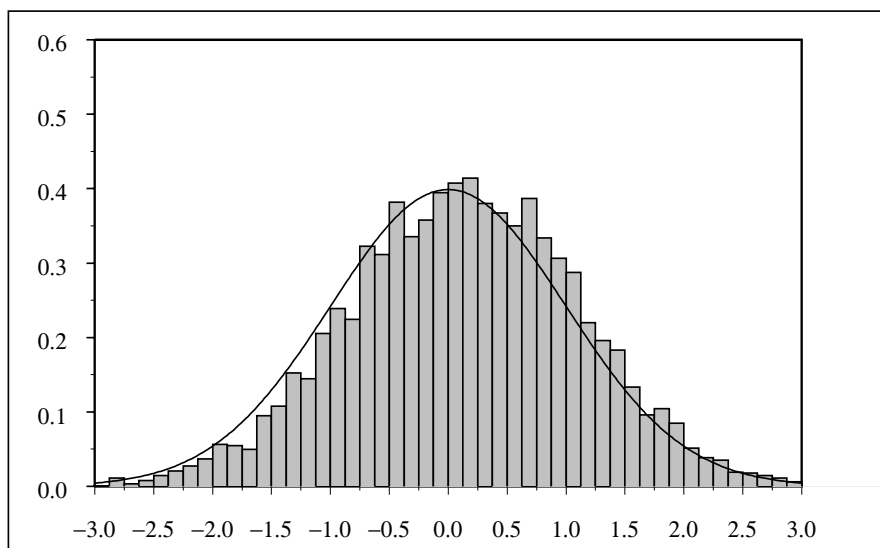


Abb. 11.2 Simuliertes Histogramm (5000 Wiederholungen) des zentrierten und standardisierten Yule-Walker-Schätzers $\sqrt{n}\hat{a}_2$ für ein zugrunde liegendes AR(1)-Modell mit $a = 0.85$ und $n = 200$ zusammen mit der Dichte der asymptotischen Standardnormalverteilung

für abhängige Zufallsgrößen komplizierter. Unterstellt man für den beobachteten Ausschnitt X_1, \dots, X_n der stationären Zeitreihe eine multivariate Normalverteilung, dann werden der zugehörige Erwartungswertvektor und die Kovarianzmatrix durch $EX_t = \mu$ und die Autokovarianzfunktion festgelegt. In diesem Fall gilt

$$(X_1, \dots, X_n)^T \sim \mathcal{N}((\mu, \dots, \mu)^T, (\gamma(r-s) : r, s = 1, \dots, n)) . \quad (11.21)$$

Setzen wir für eine kausale autoregressive Zeitreihe X voraus, dass das weiße Rauschen normalverteilt ist, so gilt (11.21). Dies folgt aus der existierenden $MA(\infty)$ -Darstellung der Zeitreihe (vgl. Aufgabe 11.1).

Ferner gilt für autoregressive Zeitreihen $\mu = 0$ und $\gamma(h)$, $h = 0, \dots, n-1$, ist eine Funktion der autoregressiven Parameter a_1, \dots, a_p und der Varianz σ_e^2 des weißen Rauschens. Damit können wir für diesen Fall eines normalverteilten autoregressiven Modells die Likelihood-Funktion aufstellen und für gegebene Beobachtungen mit einem numerischen Verfahren die Maximalstellen, d.h. den Maximum-Likelihood Schätzer bestimmen. Aufgrund der nichtlinearen Abhängigkeit der Autokovarianzfunktion von den Parametern ist eine explizite Angabe der ML -Schätzer in der Regel nicht möglich.

Viel einfacher gestaltet sich für autoregressive Zeitreihen der Ordnung p die Berechnungen von sogenannten bedingten ML -Schätzern, d.h. von ML -Schätzern bzgl. der bedingten Verteilung von X_1, \dots, X_n bei gegebenen Wer-

ten X_{1-p}, \dots, X_0 . Grundsätzlich gilt

$$\mathcal{L}(X_n, \dots, X_{1-p}) = \bigotimes_{t=1}^n \mathcal{L}(X_t | X_{t-1}, \dots, X_{1-p}) \otimes \mathcal{L}(X_0, \dots, X_{1-p}), \quad (11.22)$$

wobei $\mathcal{L}(X_t | X_{t-1}, \dots, X_{1-p})$ die bedingte Verteilung von X_t gegeben X_{t-1}, \dots, X_{1-p} bezeichnet und \otimes die Koppelungsbildung für Maße bedeutet. Für $AR(p)$ -Zeitreihen mit normalverteiltem weißem Rauschen gilt wegen

$$X_t = e_t - \sum_{j=1}^p a_j X_{t-j},$$

dass $\mathcal{L}(X_t | X_{t-1}, \dots, X_{1-p})$ die Dichte

$$\frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} \left(X_t + \sum_{j=1}^p a_j X_{t-j}\right)^2\right)$$

besitzt, woraus man direkt $\mathcal{L}(X_t | X_{t-1}, \dots, X_{1-p}) = \mathcal{L}(X_t | X_{t-1}, \dots, X_{t-p})$ für $t \geq 1$ einsieht. (11.22) besagt damit, dass

$$\begin{aligned} \mathcal{L}((X_n, \dots, X_1) | (X_0, \dots, X_{1-p})) &= \bigotimes_{t=1}^n \mathcal{L}(X_t | X_{t-1}, \dots, X_{1-p}) \\ &= \bigotimes_{t=1}^n \mathcal{L}(X_t | X_{t-1}, \dots, X_{t-p}) \end{aligned}$$

die Dichte

$$\prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{1}{2\sigma_e^2} \left(X_t + \sum_{j=1}^p a_j X_{t-j}\right)^2\right) \quad (11.23)$$

besitzt. Ein *bedingter ML-Schätzer* kann also als Minimalstelle von

$$\sum_{t=1}^n \left(X_t + \sum_{j=1}^p a_j X_{t-j}\right)^2 \quad (11.24)$$

erklärt werden. Die Suche nach Minimalstellen von (11.24) führt aber zu den bereits in Abschnitt 11.1 (vgl. (11.12)) eingeführten *Kleinste-Quadrate* oder auch *Least-Squares* (kurz: *LS*)-Schätzern für a_1, \dots, a_p . Durch partielle Differentiation nach a_1, \dots, a_p erhalten wir nämlich das zur Suche nach Minimalstellen zugehörige lineare Gleichungssystem

$$\left(\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s}\right) \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} + \begin{pmatrix} \frac{1}{n} \sum_{t=1}^n X_t X_{t-r} \\ r = 1, \dots, p \end{pmatrix} = 0. \quad (11.25)$$

Im Gegensatz zum linearen Gleichungssystem (11.1), ist dieses Gleichungssystem auch für den Fall, dass nicht alle X_t gleich sind, nicht immer eindeutig lösbar.

Man kann sich jedoch überlegen, dass eine Lösung von (11.25) die Koeffizienten der Projektion des Vektors $(-X_1, \dots, -X_n)^T$ auf den von den p Vektoren $(X_{1-j}, \dots, X_{n-j})^T, j = 1, \dots, p$, aufgespannten linearen Teilraums im \mathbb{R}^n angibt. Da diese orthogonalen Projektionen stets existieren, besitzt (11.25) in jedem Fall eine – wenn auch nicht eindeutig bestimmte – Lösung. Der mit der verallgemeinerten Moore–Penrose–Inversen gebildete Ausdruck $\hat{\mathbf{a}}^{LS}(p)$ gemäß (11.12) ist dann stets eine Lösung des Gleichungssystems (11.25).

Da unter den Voraussetzungen von Satz 11.4 auch $\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s}, r, s = 1, \dots, p$, für $n \rightarrow \infty$ n.W. gegen $\gamma(r-s)$ konvergiert (vgl. Beweis zu Satz 11.4), impliziert die Invertierbarkeit von $\Gamma(p)$ die Konsistenzaussage

$$\hat{\mathbf{a}}^{LS}(p) \xrightarrow[n \rightarrow \infty]{} -\Gamma(p)^{-1} \underline{\gamma}(p) = \mathbf{a}(p) \quad \text{n.W.} \quad (11.26)$$

in voller Analogie zu (11.6).

Für die weitere Untersuchung der Eigenschaften des LS - bzw. bedingten ML -Schätzers wollen wir nicht auf die spezielle Situation einer normalverteilten $AR(p)$ -Zeitreihe zurückgreifen, unter der wir den bedingten ML -Schätzer abgeleitet haben. Vielmehr interessiert uns, wie sich dieser Schätzer in einer möglichst allgemeinen Situation verhält. Der folgende Satz zeigt zusammen mit Satz 11.3, dass sich LS - (bzw. mit anderen Worten bedingte ML -) und Yule–Walker–Parameterschätzer bezüglich ihrer asymptotischen Normalverteilung nicht unterscheiden.

Satz 11.9 (*Asymptotische Normalität des LS -Schätzers*)

Unter den in Satz 11.3 genannten Voraussetzungen gilt für den LS -Schätzer $\hat{\mathbf{a}}^{LS}(p)$ gemäß (11.12)

$$\sqrt{n} (\hat{\mathbf{a}}^{LS}(p) - \mathbf{a}(p)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, K(p)), \quad (11.27)$$

wobei $\mathbf{a}(p)$ und $K(p)$ wie in Satz 11.3 definiert sind.

Beweis: Wie im Beweis zu Satz 11.3 erhalten wir für $\sqrt{n}(\hat{\mathbf{a}}^{LS}(p) - \mathbf{a}(p))$ eine zu (11.11) äquivalente Darstellung, wobei lediglich in den dort auftretenden Ausdrücken $\hat{\underline{\gamma}}(p)$ bzw. $\hat{\Gamma}(p)$ einerseits $\hat{\underline{\gamma}}(r)$ durch

$$\frac{1}{n} \sum_{t=1}^n X_t X_{t-r}$$

und andererseits $\hat{\underline{\gamma}}(r-s)$ durch

$$\frac{1}{n} \sum_{t=1}^n X_{t-r} X_{t-s}$$

zu ersetzen sind. Wegen

$$\hat{\gamma}(r) - \frac{1}{n} \sum_{t=1}^n X_t X_{t-r} = \mathcal{O}_P\left(\frac{1}{n}\right)$$

und analog für $\hat{\gamma}(r-s)$ erhalten wir aus Satz 10.15 zusammen mit dem Lemma von Slutsky, siehe Bemerkung A.26, die gewünschte Aussage. \square

Bemerkung 11.10

Aus Satz 11.9 folgt sofort, dass der *LS*-Schätzer für den Fall, dass tatsächlich ein autoregressives Zeitreihenmodell zugrunde liegt die Aussage von Satz 11.4 erfüllt. Es sei an dieser Stelle darin erinnert, dass wir den Nachweis zu Satz 11.4 tatsächlich bereits für den *LS*-Schätzer geführt haben.

Auch die Aussagen über asymptotische Konfidenzintervalle für die Parameter in Bemerkung 11.6 behalten damit im Kontext von Kleinste-Quadrate-Schätzern ihre Gültigkeit. \square

In der realen Anwendung wird man die Beobachtungen zunächst am arithmetischen Beobachtungsmittel \bar{X}_n zentrieren und zur Definition eines *LS*-Schätzers nach Nullstellen $(a_1, \dots, a_p)^T$ von

$$\left(\begin{array}{c} \frac{1}{n} \sum_{t=1}^n (X_{t-r} - \bar{X}_n)(X_{t-s} - \bar{X}_n) \\ r, s = 1, \dots, p \end{array} \right) \begin{pmatrix} a_1 \\ \vdots \\ a_p \end{pmatrix} + \left(\begin{array}{c} \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)(X_{t-r} - \bar{X}_n) \\ r = 1, \dots, p \end{array} \right)$$

suchen.

In Bezug auf (unbedingte) *ML*-Schätzer stellt sich nun die Frage, ob auch diese Schätzer das gleiche asymptotische Verhalten wie der Yule-Walker-Schätzer und der bedingte *ML*-Schätzer (bzw. der *LS*-Schätzer) aufweisen. Dies ist tatsächlich der Fall, wie sich auf dem im Folgenden skizzierten Weg (der sog. *Delta-Methode*, vgl. Satz A.31) ableiten lässt. Auf einen formalen Beweis verzichten wir allerdings.

Zunächst folgt unter der Normalverteilungsannahme, dass die Verteilung $\mathcal{L}(X_0, \dots, X_{1-p})$ für $\mathbf{a} = (a_1, \dots, a_p)^T$ folgende Dichte besitzt

$$f(X_0, \dots, X_{1-p}; \mathbf{a}) = \frac{1}{\sqrt{(2\pi)^p |\det \Gamma(p)|}} \exp \left(-\frac{1}{2} (X_0, \dots, X_{1-p}) \Gamma(p)^{-1} \begin{pmatrix} X_0 \\ \vdots \\ X_{1-p} \end{pmatrix} \right).$$

Um Missverständnisse zu vermeiden sei darauf hingewiesen, dass die Kovarianzmatrix $\Gamma(p)$ auf der rechten Seite des obigen Ausdrucks natürlich vom Parameter \mathbf{a} abhängt. Mit (11.22) und (11.23) erhalten wir – ebenfalls aufgrund der Normalverteilungsannahme – die gemeinsame Dichte von X_{1-p}, \dots, X_n zu

$$\frac{f(X_0, \dots, X_{1-p}; \mathbf{a})}{(2\pi\sigma_e^2)^{n/2}} \cdot \exp\left(-\frac{1}{2\sigma_e^2} \sum_{t=1}^n \left(X_t + \sum_{j=1}^p a_j X_{t-j}\right)^2\right). \quad (11.28)$$

Maximum-Likelihood (ML)-Schätzer suchen wir dann über Nullstellen der Ableitung der sogenannten *Log-Likelihood-Funktion*, d.h. der Ableitung der logarithmierten Dichte. Mit anderen Worten suchen wir nach Lösungen der Gleichung

$$\Delta_n(\mathbf{a}) = \mathbf{0} \quad (11.29)$$

mit

$$\begin{aligned} \sqrt{n}\Delta_n(\mathbf{a}) := \\ \sigma_e^2 \left(\ln f(X_0, \dots, X_{1-p}; \mathbf{a}) \right)' - \sum_{t=1}^n \left(X_t + \sum_{j=1}^p a_j X_{t-j} \right) (X_{t-1}, \dots, X_{t-p})^T, \end{aligned}$$

wobei der Ableitungsstrich ' den Vektor der partiellen Ableitungen nach den Komponenten von \mathbf{a} bezeichne. Der zusätzliche Faktor \sqrt{n} auf der linken Seite dieser Definition wird sich im Folgenden als günstig erweisen.

Falls wir auf diese Weise einen Parameterschätzer $\hat{\mathbf{a}}_{ML}$ gewinnen können, der mit einer Taylor-Entwicklung erster Ordnung von Δ_n um den wahren Parameter \mathbf{a} die folgende Eigenschaften besitzt

$$\begin{aligned} \Delta_n(\hat{\mathbf{a}}_{ML}) &= \Delta_n(\mathbf{a}) + \Delta'_n(\mathbf{a})(\hat{\mathbf{a}}_{ML} - \mathbf{a}) + o_P(1) \\ &= o_P(1), \end{aligned} \quad (11.30)$$

wobei

$$\begin{aligned} \Delta'_n(\mathbf{a}) &= -\frac{1}{\sqrt{n}} \left(\frac{\partial}{\partial a_l} \sum_{t=1}^n (X_t + \sum_{j=1}^p a_j X_{t-j}) X_{t-k} \right)_{k,l=1,\dots,p} + o_P(1) \\ &= -\frac{1}{\sqrt{n}} \left(\sum_{t=1}^n X_{t-l} X_{t-k} : k, l = 1, \dots, p \right) + o_P(1) \end{aligned}$$

die Jacobi-Matrix bezeichnet, dann erhalten wir

$$\begin{aligned} \sqrt{n}(\hat{\mathbf{a}}_{ML} - \mathbf{a}) \\ = - \left(\frac{1}{n} \sum_{t=1}^n X_{t-l} X_{t-k} : k, l = 1, \dots, p \right)^{-1} \Delta_n(\mathbf{a}) + o_P(1). \end{aligned} \quad (11.31)$$

Man beachte, dass der relevante Anteil von $\Delta_n(\mathbf{a})$, nämlich

$$-\frac{1}{\sqrt{n}} \sum_{t=1}^n (X_t + \sum_{j=1}^p a_j X_{t-j})(X_{t-1}, \dots, X_{t-p})^T$$

als lineare Funktion in \mathbf{a} fehlerfrei durch ein Taylorpolynom erster Ordnung approximiert werden kann. Im verbleibenden Anteil von $\Delta_n(\mathbf{a})$ tritt stets \sqrt{n} im Nenner auf und deshalb wird dieser Ausdruck nach Wahrscheinlichkeit gegen Null konvergieren, wenn $\hat{\mathbf{a}}_{ML}$ sich zumindest asymptotisch um einen festen Wert stabilisiert, so dass die auftretenden Ableitungen der Dichten nicht über alle Schranken wachsen. Der wesentliche Teil der Forderung in (11.30) ist also die Aussage $\Delta_n(\hat{\mathbf{a}}_{ML}) = o_P(1)$, was für einen plausiblen Schätzer eine natürliche Forderung ist.

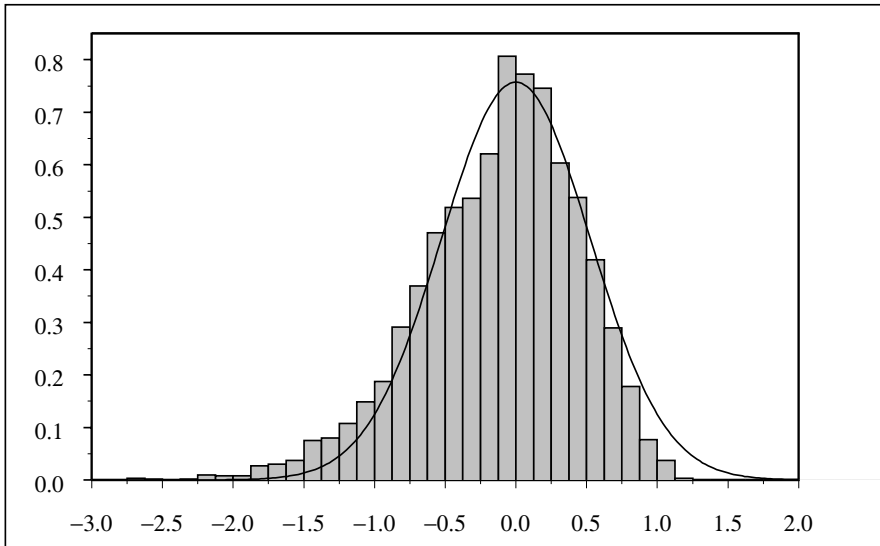


Abb. 11.3 Simuliertes Histogramm (5000 Wiederholungen) des zentrierten und standardisierten *QML*-Schätzers $\sqrt{n}(\hat{a} - a)$ im *AR*(1)-Modell mit $a = 0.85$ und $n = 200$ zusammen mit der Dichte der asymptotischen Normalverteilung

Unter den Voraussetzungen von Satz 11.3 ergibt sich dann einerseits

$$\frac{1}{\sqrt{n}} \Delta'_n(\mathbf{a}) \longrightarrow -\Gamma(p) \quad n.W. \quad (11.32)$$

und andererseits mit den Bezeichnungen aus diesem Satz

$$\Delta_n(\mathbf{a}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, A(p) \mathbf{S}_p A(p)^T). \quad (11.33)$$

Damit haben wir für den ML -Schätzer die identische asymptotische Verteilung wie für den Yule-Walker- und den LS -Schätzer sowohl unter der Annahme der Gültigkeit des autoregressiven Modells als auch unter der Annahme eines $MA(\infty)$ -Modells (wie in Satz 11.3) erhalten. Es sei ausdrücklich darauf hingewiesen, dass die eingangs gemachte Normalverteilungsannahme nur zur Herleitung der Likelihood-Funktion und damit zur Einführung von ML -Schätzern herangezogen wurde. Für den oben skizzierten Nachweis der asymptotischen Normalität aber – wie bei den anderen betrachteten Parameterschätzverfahren auch – wurde eine Normalverteilungsannahme nicht benutzt. Aus diesem Grunde nennt man einen aufgrund einer hypothetisch unterstellten Normalverteilungsannahme lediglich abgeleiteten ML -Schätzer in der Literatur auch *Quasi-Maximum-Likelihood (QML)-Schätzer*.

Ganz wesentlich für die Gültigkeit der oben beschriebenen Vorgehensweise ist die Eigenschaft (11.30) des QML -Schätzers. Einen solchen Schätzer erhalten wir, wenn wir ausgehend vom Yule-Walker-Schätzer $\hat{\mathbf{a}}_{YW}$ mit dem Newton-Verfahren numerisch eine Nullstelle der Log-Likelihood-Funktion approximieren. Es ist so, dass bereits eine 1-Schritt-Korrektur ausreicht, d.h. der Schätzer

$$\hat{\mathbf{a}}_{ML}(p) := \hat{\mathbf{a}}_{YW}(p) - (\Delta'_n(\hat{\mathbf{a}}_{YW}))^{-1} \Delta_n(\hat{\mathbf{a}}_{YW}) \quad (11.34)$$

erfüllt nun einerseits wegen

$$(\Delta'_n(\hat{\mathbf{a}}_{YW}))^{-1} \Delta_n(\hat{\mathbf{a}}_{YW}) = \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} \Delta'_n(\hat{\mathbf{a}}_{YW}) \right)^{-1} \Delta_n(\hat{\mathbf{a}}_{YW}) \rightarrow 0 \quad n.W.$$

die Aussage $\hat{\mathbf{a}}_{ML}(p) \rightarrow \mathbf{a}(p) \ n.W.$ und andererseits $\Delta_n(\hat{\mathbf{a}}_{ML}(p)) = o_P(1)$, was man durch direktes Einsetzen in die Taylor-Approximation (erste Gleichung von (11.30)) und die erneute Anwendung dieser Taylor-Reihe auf $\hat{\mathbf{a}}_{YW}$ einsehen kann. Weitere Approximationsschritte mit dem Newton-Verfahren führen möglicherweise zu einer besseren Approximation an eine Nullstelle der Log-Likelihood-Funktion, aber sie verändern das asymptotische Verhalten der resultierenden Parameterschätzer nicht.

Zusammenfassend kann schließlich das folgende Resultat auf Basis der obenstehenden Überlegungen formuliert werden.

Satz 11.11 (*Asymptotische Normalität des QML-Schätzers*)

Unter den Voraussetzungen von Satz 11.3 gilt für einen QML -Schätzer $\hat{\mathbf{a}}_{ML}(p)$ mit (11.30) – also etwa für den in (11.34) definierten Schätzer – die asymptotische Verteilungsaussage (11.8).

Analog gilt unter den Voraussetzungen des Satzes 11.4 die asymptotische Verteilungsaussage (11.13).

Beispiel 11.12

Für das gleiche $AR(1)$ -Modell wie in Beispiel 11.8 ist in Abbildung 11.3 ebenfalls basierend auf 5000 simulierten Zeitreihen mit der Länge $n = 200$ ein Histogramm des QML -Schätzers für den Parameter a (zusammen mit der Dichte der asymptotischen Normalverteilung) wiedergegeben. Im Vergleich zu Abbildung 11.1 erkennt man eine bessere Anpassung der asymptotischen Normalverteilung für den QML -Schätzer. Die einzelnen Zeitreihen wurden mit Hilfe der Routine `arima.sim` und die Schätzer mit der Routine `arima.mle` von S-PLUS® berechnet.

□

11.3 Parameterschätzung in autoregressiven Modellen mit wachsender Ordnung

Grundsätzlich wird man in der Anwendung jede Annahme eines Zeitreihenmodells lediglich als Approximation an die tatsächlich zugrunde liegende Struktur verstehen. Man möchte durch eine angemessene Modellannahme etwa erreichen, dass möglichst viele wichtige Eigenschaften eines realen Datensatzes mit dem gewählten Modell dargestellt werden können. Es ist eine Tatsache, dass autoregressive Zeitreihenmodelle in der Anwendung eine ganz wichtige, wenn nicht sogar eine dominierende Rolle spielen. Ein Grund dafür liegt sicher darin, dass die Parameterschätzung für autoregressive Zeitreihenmodelle ausgesprochen stabil verläuft und bei Anwendung der Yule–Walker–Methode immer Eindeutigkeit des Parameterschätzers gegeben ist und darüberhinaus das geschätzte Modell stets zur Klasse der stationären und kausalen autoregressiven Zeitreihen gehört. Mit dem Levinson–Algorithmus (vgl. Bemerkung 11.7) steht außerdem ein schnelles und numerisch wenig aufwändiges Berechnungsverfahren für Yule–Walker–Schätzer zur Verfügung. Selbst wenn das angenommene autoregressive Zeitreihenmodell dem vorliegenden Datensatz nicht zugrunde liegt, verhalten sich die üblichen Parameterschätzer gutmütig und führen zu gut interpretierbaren Ergebnissen (vgl. hierzu die Sätze 11.3, 11.9 und 11.11).

Da man mit wachsendem Stichprobenumfang auch größere Modellordnungen in Betracht ziehen wird, wollen wir uns in diesem Abschnitt mit der Parameterschätzung in derartigen Situationen auseinandersetzen. Dieser Zugang bereitet auch die deutlich aufwändigere Parameterschätzung in $ARMA(p, q)$ -Modellen vor, die wir im folgenden Abschnitt in Angriff nehmen werden. Damit die Übersichtlichkeit der Darstellung erhalten bleibt, nehmen wir als datengenerierenden Prozess eine autoregressive Zeitreihe unendlicher Ordnung an, d.h. es gelte

$$X_t + \sum_{j=1}^{\infty} a_j X_{t-j} = e_t, \quad \forall t \in \mathbb{Z}, \quad (11.35)$$

für ein aus i.i.d. Zufallsvariablen e_t bestehendes weißes Rauschen mit Varianz σ_e^2 . Die Koeffizienten $(a_j : j \in \mathbb{N})$ mögen die Voraussetzung

$$1 + \sum_{j=1}^{\infty} a_j z^j \neq 0, \quad \forall |z| \leq 1 + \delta, \quad (11.36)$$

für ein geeignetes $\delta > 0$ erfüllen. Hieraus folgt sofort für ein geeignetes $\eta \in (0, 1)$ und $C < \infty$

$$|a_j| \leq C \cdot \eta^j, \quad \forall j \in \mathbb{N}, \quad (11.37)$$

und die Existenz einer $MA(\infty)$ -Darstellung

$$X_t = e_t + \sum_{j=1}^{\infty} \alpha_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (11.38)$$

mit $(1 + \sum_{j=1}^{\infty} a_j z^j)^{-1} = 1 + \sum_{j=1}^{\infty} \alpha_j z^j$, $|z| \leq 1$.

Basierend auf Beobachtungen (X_1, \dots, X_n) der beschriebenen $AR(\infty)$ -Zeitreihe betrachten wir exemplarisch das Verhalten des Yule-Walker-Schätzers

$$\hat{\mathbf{a}}(p) = -\hat{\Gamma}(p)^+ \hat{\gamma}(p)$$

(vgl. (Satz 11.3)) für wachsende Modellordnungen $p \in \mathbb{N}$. Satz 11.3 beschreibt das Verhalten dieses Schätzers für jedes fest gewählte $p \in \mathbb{N}$. Bevor wir mit Satz 11.14 ein erstes Resultat über das Verhalten des Yule-Walker-Schätzers bei wachsender Modellordnung angeben und beweisen können, geben wir im Folgenden eine explizite Darstellung für $\Gamma(p)^{-1}$ an.

Lemma 11.13

Sei $X = (X_t : t \in \mathbb{Z})$ eine $AR(\infty)$ -Zeitreihe mit (11.35) und (11.36). $\mathbf{a}(p)$ sei die eindeutige Minimalstelle von $E(X_t + \sum_{j=1}^p a_j X_{t-j})^2$, d.h.

$$\mathbf{a}(p) = (a_1(p), \dots, a_p(p))^T = -\Gamma(p)^{-1} \gamma(p)$$

und

$$\sigma_e^2(p) = E\left(X_t + \sum_{j=1}^p a_j(p) X_{t-j}\right)^2$$

bezeichne das Minimum. Mit

$$A(p) = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ a_1(p-1) & 1 & \ddots & & \vdots \\ a_2(p-1) & a_1(p-2) & \ddots & 0 & \vdots \\ \vdots & \vdots & & 1 & 0 \\ a_{p-1}(p-1) & a_{p-2}(p-2) & & a_1(1) & 1 \end{pmatrix}$$

gilt die Darstellung

$$\Gamma(p)^{-1} = A(p) \operatorname{diag}(\sigma_e^2(p-1), \dots, \sigma_e^2(0))^{-1} A(p)^T. \quad (11.39)$$

Bezeichnen wir die Einträge von $\Gamma(p)^{-1}$ mit $\overline{\gamma_p(r, s)}$, $r, s = 1, \dots, p$, so bedeutet (11.39) für $r \geq s$ und unter Beachtung von $a_0(k) = 1 \quad \forall k \in \mathbb{N}$

$$\overline{\gamma_p(r, s)} = \sum_{j=1}^s a_{r-j}(p-j) a_{s-j}(p-j) / \sigma_e^2(p-j). \quad (11.40)$$

Mit $\eta \in (0, 1)$ aus (11.37) gilt für $\mathbf{a}(p)$ weiter

$$\sum_{j=0}^p \left| \frac{a_j(p)}{\sigma_e^2(p)} - \frac{a_j}{\sigma_e^2} \right| = \mathcal{O}(\eta^p). \quad (11.41)$$

Hieraus erhält man für die Einträge von $\Gamma(p)^{-1}$ die folgende hilfreiche Abschätzung

$$\left| \overline{\gamma_p(r, s)} \right| \leq C(\eta^{|r-s|} + \eta^p), \quad \forall r, s = 1, \dots, p \quad \forall p \in \mathbb{N}, \quad (11.42)$$

und $C < \infty$.

Beweis: Die explizite Darstellung (11.39) der Inversen der Autokovarianzmatrix zeigt man per Induktion. Der Induktionsanfang für $p = 1$ ist wegen $\sigma_e^2(0) = \gamma(0)$ offensichtlich. Bezüglich des Induktionsschritts verweisen wir auf Aufgabe 11.3. An dieser Stelle sei angemerkt, dass die Dissertation von Kromer (1970) ein sehr ähnliches Resultat enthält, welches auch in Berk (1974), Formel (2.6), angegeben wird. Die hier verwendete Darstellung (11.39) wird sich später als vorteilhafter erweisen. (11.40) ist nichts anderes als die explizite Umsetzung von (11.39).

Zum Nachweis von (11.41) greifen wir auf ein Resultat von Baxter (1962), Theorem 2.2, zurück, welches besagt, dass

$$\sum_{j=0}^p \left| \frac{a_j(p)}{\sigma_e^2(p)} - \frac{a_j}{\sigma_e^2} \right| \leq \operatorname{Const} \cdot \sum_{j=p+1}^{\infty} \frac{|a_j|}{\sigma_e^2}$$

gilt. Da die Annahme (11.36) ein geometrisches Abklingen der $AR(\infty)$ -Koeffizienten a_j impliziert (vgl. (11.37)), ergibt sich aus Baxters Abschätzung (11.41).

Aufgrund der Tatsache $\sigma_e^2(0) \geq \sigma_e^2(1) \geq \sigma_e^2(2) \geq \dots \geq \sigma_e^2$ folgt aus (11.41) insbesondere die gleichmäßige Beschränktheit der $a_j(p)$ in j und p , so dass aus (11.40) mit Hilfe von (11.41) und (11.37) schließlich auch (11.42) gefolgert werden kann.

Die Beschränktheit der Folge $\sigma_e^2(p)$ impliziert auch

$$\sum_{j=0}^p |a_j(p)| \leq \sigma_e^2(0) \left\{ \sum_{j=0}^p \left| \frac{a_j(p)}{\sigma_e^2(p)} - \frac{a_j}{\sigma_e^2} \right| + \sum_{j=0}^p |a_j|/\sigma_e^2 \right\} = \mathcal{O}(1), \quad (11.43)$$

ein Resultat, das wir später noch benötigen werden. \square

Von Interesse ist natürlich, wie gut die empirische Kovarianzmatrix $\hat{\Gamma}(p)$ (vgl. (11.2)), die tatsächliche Kovarianzmatrix $\Gamma(p)$ approximiert. Für jedes feste p erhalten wir aufgrund von Satz 10.15 sofort die \sqrt{n} -Konsistenz von $\hat{\Gamma}(p)$. Nachfolgend behandeln wir den Fall wachsender Ordnungen $p(n)$.

Satz 11.14

Gegeben sei ein autoregressives Modell unendlicher Ordnung gemäß (11.35) und (11.36). Das zugehörige weiße Rauschen $(e_t : t \in \mathbb{Z})$ bestehe aus i.i.d. Zufallsvariablen mit $Ee_1^4 < \infty$. $(p(n) : n \in \mathbb{N})$ sei eine Folge natürlicher Zahlen mit $p(n) \rightarrow \infty$ und $p^3(n)/n$ beschränkt für $n \rightarrow \infty$. Dann gilt

$$\frac{\sqrt{n}}{p(n)} \left\| \hat{\Gamma}(p(n)) - \Gamma(p(n)) \right\| = \mathcal{O}_P(1), \quad (11.44)$$

wobei für eine $p \times p$ -Matrix A die Norm $\|A\| = \sup(\|Ax\| : x \in \mathbb{R}^p, \|x\| = 1)$ und für einen p -Vektor x die Euklidische Norm $\|x\|$ verwendet wird. Weiterhin haben wir für die Moore–Penrose–Inverse

$$\frac{\sqrt{n}}{p(n)} \left\| \hat{\Gamma}(p(n))^+ - \Gamma(p(n))^{-1} \right\| = \mathcal{O}_P(1) \quad (11.45)$$

sowie

$$\sqrt{\frac{n}{p(n)}} \left\| \hat{\mathbf{a}}(p(n)) - \mathbf{a}(p(n)) \right\| = \mathcal{O}_P(1). \quad (11.46)$$

wobei $\mathbf{a}(p(n)) = \Gamma(p(n))^{-1} \gamma(p(n))$, vergleiche Satz 11.3. Interessieren wir uns nur für das Verhalten eines festen Abschnitts der Länge K des Vektors $\hat{\mathbf{a}}(p(n))$, also $\hat{\mathbf{a}}_K(p(n)) := (\hat{a}_1(p(n)), \dots, \hat{a}_K(p(n)))^T$ und $\mathbf{a}_K(p(n))$ entsprechend, so gilt

$$\sqrt{n} \left\| \hat{\mathbf{a}}_K(p(n)) - \mathbf{a}_K(p(n)) \right\| = \mathcal{O}_P(1). \quad (11.47)$$

In (11.46) kann $\mathbf{a}(p(n))$ durch $\mathbf{a}_{p(n)} = (a_1, \dots, a_{p(n)})^T$ und in (11.47) kann $\mathbf{a}_K(p(n))$ durch $\mathbf{a}_K = (a_1, \dots, a_K)^T$ ersetzt werden.

Bevor wir dieses Resultat beweisen, wollen wir in der folgenden Bemerkung auf den Unterschied zwischen der Anpassung eines AR-Modells mit fester bzw. wachsender Ordnung eingehen.

Bemerkung 11.15

Es sei ausdrücklich darauf hingewiesen, dass die Vektoren $\hat{\mathbf{a}}_K(p(n))$ und $\hat{\mathbf{a}}(K)$ sowie $\mathbf{a}_K(p(n))$ und $\mathbf{a}(K)$ im Allgemeinen nicht übereinstimmen. Auch stimmt $\mathbf{a}(K)$ nicht mit \mathbf{a}_K überein!

$\hat{\mathbf{a}}_K(p(n))$ stellt die ersten K Komponenten eines Schätzers für die Koeffizienten der $AR(\infty)$ -Entwicklung der zugrunde liegenden Zeitreihe dar. Dafür ist es ganz wesentlich, dass die Ordnung $p(n)$ mit wachsendem Stichprobenumfang n gegen unendlich konvergiert. Da die ersten K Koeffizienten a_1, \dots, a_K der $AR(\infty)$ -Darstellung aber keineswegs mit den Parametern $a_1(K), \dots, a_K(K)$ eines best angepassten AR -Modells mit der festen Ordnung K übereinstimmen, sind auch die Schätzer für die beiden verschiedenen K -dimensionalen Parametervektoren, nämlich $\hat{\mathbf{a}}_K(p(n))$ und $\hat{\mathbf{a}}(K)$ grundsätzlich unterschiedlich und schätzen tatsächlich unterschiedliche Zielgrößen.

Der Unterschied zwischen $\mathbf{a}_K(p(n))$ und \mathbf{a}_K verschwindet erst mit wachsendem n , wie Lemma 11.13, (11.41) zeigt.

Nur im Fall $a_j = 0 \quad \forall j > K$ – d.h. der datengenerierende Prozess ist tatsächlich eine $AR(p)$ -Zeitreihe mit endlicher Ordnung $p \leq K$ – stimmen $\mathbf{a}(K)$, \mathbf{a}_K und $\mathbf{a}_K(p(n))$ (falls $p(n) \geq K$) alle drei überein. In diesem Fall können sich die zugehörigen Schätzer $\hat{\mathbf{a}}(K)$ und $\hat{\mathbf{a}}_K(p(n))$ zwar immer noch zahlenmäßig unterscheiden, aber sie unterscheiden sich dann nicht mehr grundsätzlich, da sie beide die gleiche Zielgröße schätzen. Auch ist das asymptotische Verhalten der beiden Schätzer identisch, wie wir noch sehen werden. Festzuhalten bleibt, dass die beiden Situationen

Anpassung eines $AR(K)$ -Modells (K fest)

und

Anpassung eines $AR(p(n))$ -Modells ($p(n)$ wachsend)

an einen Datensatz strikt auseinander zuhalten sind, und zwar selbst dann, wenn wir uns nur für einen Anfangsabschnitt der Parameterschätzer mit fester Länge interessieren!

□

Beweis: (zu Satz 11.14)

Durch direkte Abschätzung erhalten wir zunächst

$$\begin{aligned}
 & \|\hat{\Gamma}(p(n)) - \Gamma(p(n))\|^2 \\
 &= \sup_{\|\mathbf{x}\|=1} \sum_{r=1}^{p(n)} ((\hat{\gamma}(r-s) - \gamma(r-s) : s = 1, \dots, p(n))^T \mathbf{x})^2 \\
 &\leq \sum_{r,s=1}^{p(n)} (\hat{\gamma}(r-s) - \gamma(r-s))^2.
 \end{aligned}$$

Da $E(\widehat{\gamma}(h) - \gamma(h))^2$ unabhängig von h nach oben durch $\mathcal{O}(n^{-1})$ abgeschätzt werden kann (diese Aussage kann aus Satz 10.4, insbesondere (10.15) und (10.16) abgeleitet werden), folgt (11.44). Mit (11.42), Lemma 11.13, erhalten wir wie folgt die Beschränktheit der Folge $\{\|\Gamma(p)^{-1}\| : p \in \mathbb{N}\}$.

$$\begin{aligned}
& \|\Gamma(p)^{-1}\|^2 \\
&= \sup_{\|\mathbf{x}\|=1} \sum_{r=1}^p \left(\sum_{s=1}^p \overline{\gamma_p(r, s)} x_s \right)^2 \\
&\leq \sup_{\|\mathbf{x}\|=1} \sum_{r=1}^p \left(\sum_{s=1}^p |\overline{\gamma_p(r, s)}| x_s^2 \right) \left(\sum_{s=1}^p |\overline{\gamma_p(r, s)}| \right) \\
&\leq \max_{r=1, \dots, p} \left(\sum_{s=1}^p |\overline{\gamma_p(r, s)}| \right)^2 \cdot \sup_{\|\mathbf{x}\|=1} \sum_{s=1}^p x_s^2 \\
&= \mathcal{O}(1),
\end{aligned}$$

da $\sum_{s=1}^p |\overline{\gamma_p(r, s)}|$ gleichmäßig in r und p mit Hilfe von (11.42) nach oben beschränkt werden kann.

Die Beschränktheit von $\|\Gamma(p)^{-1}\|$ erlaubt nun die Übertragung der Konvergenzrate von $\widehat{\Gamma}(p(n))$ auf die inverse empirische Autokovarianzmatrix. Da allgemein $A^{-1} - B^{-1} = (A^{-1} - B^{-1} + B^{-1})(B - A)B^{-1}$ gilt, kann mit der Dreiecksungleichung und wegen $\|AB\| \leq \|A\|\|B\|$ wie folgt abgeschätzt werden

$$\begin{aligned}
& \|\widehat{\Gamma}(p(n))^{-1} - \Gamma(p(n))^{-1}\| \\
&\leq \frac{\|\Gamma(p(n))^{-1}\|^2 \cdot \|\widehat{\Gamma}(p(n)) - \Gamma(p(n))\|}{1 - \|\widehat{\Gamma}(p(n)) - \Gamma(p(n))\| \cdot \|\Gamma(p(n))^{-1}\|}.
\end{aligned}$$

Die erwähnte Beschränktheit von $\|\Gamma(p(n))^{-1}\|$ und (11.44) ergeben nun (11.45). Nach diesen Vorbereitungen können wir nun (11.46) und (11.47) zeigen. Dazu beachte die bereits im Beweis von Satz 11.3 eingesetzte Zerlegung (vgl. (11.11)), mit der folgt

$$\begin{aligned}
\|\widehat{\mathbf{a}}(p(n)) - \mathbf{a}(p(n))\| &\leq \|\widehat{\Gamma}(p(n))^{-1}(\widehat{\underline{\gamma}}(p(n)) - \underline{\gamma}(p(n)))\| \\
&\quad + \|\widehat{\Gamma}^{-1}(p(n))\| \|\widehat{\Gamma}(p(n)) - \Gamma(p(n))\| \|\mathbf{a}(p(n))\|.
\end{aligned} \tag{11.48}$$

Der erste Summand wird wie folgt nach oben abgeschätzt

$$\begin{aligned}
& \|\widehat{\Gamma}(p(n))^{-1} - \Gamma(p(n))^{-1}\| \|\widehat{\underline{\gamma}}(p(n)) - \underline{\gamma}(p(n))\| \\
&\quad + \|\Gamma(p(n))^{-1}\| \|\widehat{\underline{\gamma}}(p(n)) - \underline{\gamma}(p(n))\| \\
&\leq \mathcal{O}_P\left(\frac{p(n)^{3/2}}{n}\right) + \left(\sum_{r=1}^{p(n)} \left[\sum_{s=1}^{p(n)} \overline{\gamma_{p(n)}(r, s)} (\widehat{\gamma}(s) - \gamma(s)) \right]^2 \right)^{-1/2}. \tag{11.49}
\end{aligned}$$

Hierbei beachte man, dass $\hat{\gamma}(s) - \gamma(s) = \mathcal{O}_P(n^{-1/2})$ und zwar gleichmäßig in s gilt, sowie die Tatsache, dass bei Abschätzung der Euklidischen Norm eines $p(n)$ -Vektors mit beschränkten Komponenten ein Faktor $\sqrt{p(n)}$ entsteht. Die erste dieser Aussagen in Verbindung mit der Beschränktheit von $\gamma_p(r, s)$ aus (11.42) sichert, dass der zweite Summand in der letzten Formel gleich $\mathcal{O}_P((p(n)/n)^{1/2})$ ist, wobei der Ausdruck $[\dots]^2$ mit Hilfe der Cauchy-Schwarz-Ungleichung für Vektoren abgeschätzt wird, wobei einer der Vektoren $(\gamma_p(r, s))^{1/2} : s = 1, \dots, p(n))^T$ sei. Wegen $p(n)^3/n = \mathcal{O}(1)$ folgt schließlich, dass der Ausdruck (11.49) gleich $\mathcal{O}_P((p(n)/n)^{1/2})$ ist.

Auf genau die gleiche Art und Weise – diesmal wegen der Summierbarkeit von $a_j(p)$ (vgl. (11.44)) – folgt für den zweiten Summanden in (11.48) die Schranke $\mathcal{O}_P((p(n)/n)^{1/2})$ und damit (11.46).

Bei genauer Inspektion der obigen Beweisschritte erhalten wir auf genau dem gleichen Wege, dass für jede einzelne Komponente $\hat{a}_j(p(n))$ von $\hat{\mathbf{a}}(p(n))$, $j = 1, \dots, p(n)$ die Aussage

$$\sqrt{n}(\hat{a}_j(p(n)) - a_j(p(n))) = \mathcal{O}_P(1) \quad (11.50)$$

gilt, woraus sofort (11.47) folgt. Die Aussage $\sum_{j=0}^p |a_j(p) - a_j| = \mathcal{O}(\eta^p)$, die aus (11.41) folgt (beachte, dass (11.41) wegen $a_0(p) = a_0 = 1$ insbesondere $|\sigma_e^2(p) - \sigma_e^2| = \mathcal{O}(\eta^p)$ ergibt), erlaubt abschließend die Ersetzung von $a_j(p(n))$ durch a_j , $j = 1, \dots, p(n)$, sowohl in (11.46) wie (11.47). Damit ist Satz 11.14 vollständig bewiesen. \square

Mit den erzielten Ergebnissen ist nun sichergestellt, dass wir in der Lage sind, eine mit dem Stichprobenumfang wachsende Anzahl $p(n)$ von Koeffizienten einer $AR(\infty)$ -Zeitreihe konsistent zu schätzen. Für jeden einzelnen Koeffizienten a_j der $AR(\infty)$ -Darstellung haben wir sogar die übliche Konvergenzrate $1/\sqrt{n}$ nachgewiesen (vgl. (11.50) und beachte, dass $\sqrt{n}(a_j(p(n)) - a_j) = \mathcal{O}(1)$). Es stellt sich die Frage, ob für einen festen Abschnitt der Länge K , d.h. für $\hat{\mathbf{a}}_K(p(n))$, sogar ein zentraler Grenzwertsatz (CLT) mit einer asymptotischen Normalverteilung gilt. Dies ist tatsächlich der Fall, wie der folgende und abschließende Satz dieses Abschnitts besagt.

Satz 11.16 (CLT für den Yule-Walker-Schätzer im $AR(\infty)$ -Modell)

Sei $X = (X_t : t \in \mathbb{Z})$ ein autoregressives Modell unendlicher Ordnung gemäß (11.35). Es gebe $C < \infty$ und $\eta \in (0, 1)$, sodass für die AR -Koeffizienten $(a_j : j \in \mathbb{N})$

$$|a_j| \leq C \cdot \eta^j, \quad \forall j \in \mathbb{N},$$

gilt. Das zugehörige weiße Rauschen $e = (e_t : t \in \mathbb{Z})$ bestehe aus i.i.d. Zufallsvariablen mit Erwartungswert Null und Varianz σ_e^2 . Ferner gelte $Ee_1^4 < \infty$ und $(p(n) : n \in \mathbb{N})$ bezeichne eine Folge natürlicher Zahlen mit

$$p(n) \longrightarrow \infty \quad \text{und} \quad p(n)^3/n \longrightarrow 0 \quad \text{für } n \rightarrow \infty. \quad (11.51)$$

Für festes $K \in \mathbb{N}$ gilt dann für den Yule-Walker-Schätzer $\hat{\mathbf{a}}_K(p(n))$

$$\sqrt{n} \begin{pmatrix} \hat{a}_1(p(n)) - a_1 \\ \vdots \\ \hat{a}_K(p(n)) - a_K \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma) \quad (11.52)$$

mit

$$\Sigma(r, s) = \sum_{j=1}^{\min(r, s)} a_{r-j} a_{s-j}, \quad r, s = 1, \dots, K. \quad (11.53)$$

Dabei wurde $a_0 = 1$ verwendet.

Mit

$$A := \begin{pmatrix} 1 & 0 & \dots & 0 \\ a_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{K-1} & \dots & a_1 & 1 \end{pmatrix}$$

gilt

$$\Sigma = AA^T. \quad (11.54)$$

Beweis: Um den Beweis übersichtlich zu halten betrachten wir nur den einfachsten Fall, nämlich $K = 1$. Aufgrund von Satz 11.3 wissen wir zunächst, dass zu jedem festen $p \in \mathbb{N}$ eine Varianz $\tau^2(p)$ existiert, so dass folgende Verteilungskonvergenzaussage gilt

$$\sqrt{n} (\hat{a}_1(p) - a_1(p)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2(p)). \quad (11.55)$$

Da wir an dem asymptotischen Verhalten von $\sqrt{n}(\hat{a}_1(p(n)) - a_1)$ interessiert sind und wir Satz A.29 anwenden wollen, müssen wir zeigen, dass die entsprechende Approximationaussage (A.66) aus Satz A.29 gilt. Dazu werden wir sowohl die linke Seite von (11.55) als auch den uns interessierenden Ausdruck modifizieren. Beginnen wir mit der linken Seite von (11.55). Aufgrund der Definition des Yule-Walker-Schätzers erhalten wir

$$\begin{aligned}
& \sqrt{n}(\widehat{\mathbf{a}}(p) - \mathbf{a}(p)) \\
&= \widehat{\Gamma}(p)^{-1} \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) - \sum_{j=1}^p a_j(p) \widehat{\gamma}(r-j) \\ r = 1, \dots, p \end{pmatrix} \\
&= \widehat{\Gamma}(p)^{-1} \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) + \gamma(r) - \sum_{j=1}^p a_j(p) (\widehat{\gamma}(r-j) - \gamma(r-j)) \\ r = 1, \dots, p \end{pmatrix} \\
&= \Gamma(p)^{-1} \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) + \gamma(r) - \sum_{j=1}^p a_j(p) (\widehat{\gamma}(r-j) - \gamma(r-j)) \\ r = 1, \dots, p \end{pmatrix} + o_P(1),
\end{aligned}$$

wobei für die letzte Identität sowohl Satz 10.15 und Satz 11.14 (Aussage (11.45)) zu beachten sind. Aufgrund von (11.40) ergibt sich die erste Zeile von $\Gamma(p)^{-1}$ zu

$$\frac{1}{\sigma_e^2(p-1)} \left(1, a_1(p-1), \dots, a_{p-1}(p-1) \right),$$

so dass obige Rechnung und (11.55) die Konvergenz

$$\begin{aligned}
Z_{n,p} &:= \\
&\frac{1}{\sigma_e^2(p-1)} \sum_{r=1}^p a_{r-1}(p-1) \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) + \gamma(r) - \sum_{j=1}^p a_j(p) (\widehat{\gamma}(r-j) - \gamma(r-j)) \\ r = 1, \dots, p \end{pmatrix} \\
&\xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2(p))
\end{aligned} \tag{11.56}$$

begründet.

Eine vergleichbare, aber weitergehende Umrechnung für den uns interessierenden Ausdruck liefert

$$\begin{aligned}
& \sqrt{n}(\widehat{\mathbf{a}}(p(n)) - \mathbf{a}(p(n))) \\
&= \widehat{\Gamma}(p(n))^{-1} \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) - \sum_{j=1}^{p(n)} a_j \widehat{\gamma}(r-j) \\ r = 1, \dots, p(n) \end{pmatrix} \\
&= \widehat{\Gamma}(p(n))^{-1} \sqrt{n} \begin{pmatrix} -\widehat{\gamma}(r) + \gamma(r) - \sum_{j=1}^{p(n)} a_j (\widehat{\gamma}(r-j) - \gamma(r-j)) \\ r = 1, \dots, p(n) \end{pmatrix} \\
&\quad - \widehat{\Gamma}(p(n))^{-1} \sqrt{n} \begin{pmatrix} \gamma(r) + \sum_{j=1}^{p(n)} a_j \gamma(r-j) \\ r = 1, \dots, p(n) \end{pmatrix}.
\end{aligned} \tag{11.57}$$

Wegen $\gamma(r) + \sum_{j=1}^{\infty} a_j \gamma(r-j) = 0$, $\forall r = 1, 2, \dots$, folgt die Beziehung $\gamma(r) + \sum_{j=1}^{p(n)} a_j \gamma(r-j) = -\sum_{j=p(n)+1}^{\infty} a_j \gamma(r-j)$.

Das geometrische Abklingen von a_j , die Beschränktheit der Autokovarianzfunktion γ und (11.45) ergeben, dass der zweite Summand nach dem letzten Gleichheitszeichen in (11.57) nach Wahrscheinlichkeit verschwindet.

Die in $h \in \mathbb{N}$ gleichmäßige Beschränktheit von $E(\sqrt{n}(\hat{\gamma}(h) - \gamma(h)))^2$ zusammen mit (11.45) führt dann weiter zu

$$\begin{aligned} & \sqrt{n}(\hat{\mathbf{a}}(p(n)) - \mathbf{a}(p(n))) \\ &= \Gamma(p(n))^{-1} \sqrt{n} \left(-\hat{\gamma}(r) + \gamma(r) - \sum_{\substack{j=1 \\ r=1, \dots, p(n)}}^{p(n)} a_j (\hat{\gamma}(r-j) - \gamma(r-j)) \right) + o_P(1), \end{aligned} \quad (11.58)$$

wobei $o_P(1)$ für einen $p(n)$ -dimensionalen Vektor hier bedeutet, dass die entsprechende Norm in $\mathbb{R}^{p(n)}$ nach Wahrscheinlichkeit gegen Null konvergiert.

Mit der expliziten Darstellung der Inversen der Kovarianzmatrix $\Gamma(p(n))$ aus (11.40) und der Approximation der Koeffizienten $a_j(p)$ durch a_j aus (11.41) erhält man mit etwas längerer Rechnung, dass die erste Zeile aus (11.58) in der folgenden vereinfachten Form geschrieben werden kann.

$$\begin{aligned} & \sqrt{n}(\hat{a}_1(p(n)) - a_1) \\ &= \underbrace{-\frac{1}{\sigma_e^2} \sum_{r=1}^{p(n)} a_{r-1} \sqrt{n} \left(\hat{\gamma}(r) - \gamma(r) + \sum_{j=1}^{p(n)} a_j (\hat{\gamma}(r-j) - \gamma(r-j)) \right)}_{=Z_n} + o_P(1). \end{aligned} \quad (11.59)$$

Nun können wir Satz A.29 auf Z_n und $Z_{n,p}$ (vgl. (11.59) und (11.56)) anwenden. Wir erhalten für $p(n) \geq p$

$$\begin{aligned} & \limsup_{n \rightarrow \infty} E |Z_{n,p} - Z_n| \\ & \leq \limsup_{n \rightarrow \infty} E \left| \sum_{r=1}^p \left(\frac{a_{r-1}(p-1)}{\sigma_e^2(p-1)} - \frac{a_{r-1}}{\sigma_e^2} \right) \right| \\ & \quad \times \sqrt{n} \left| -\hat{\gamma}(r) + \gamma(r) - \sum_{j=1}^p a_j(p) (\hat{\gamma}(r-j) - \gamma(r-j)) \right| \\ & \quad + \limsup_{n \rightarrow \infty} E \left| \sum_{r=1}^p \frac{a_{r-1}}{\sigma_e^2} \sqrt{n} \sum_{j=1}^p (a_j - a_j(p)) (\hat{\gamma}(r-j) - \gamma(r-j)) \right| \\ & \quad + \limsup_{n \rightarrow \infty} E \left| \sum_{r=1}^p \frac{a_{r-1}}{\sigma_e^2} \sqrt{n} \sum_{j=p+1}^{p(n)} a_j (\hat{\gamma}(r-j) - \gamma(r-j)) \right| \\ & \leq \sum_{r=1}^p \left| \frac{a_{r-1}(p-1)}{\sigma_e^2(p-1)} - \frac{a_{r-1}}{\sigma_e^2} \right| \mathcal{O}(1) + \sum_{j=1}^p |a_j - a_j(p)| \mathcal{O}(1) + \sum_{j=p+1}^{\infty} |a_j| \mathcal{O}(1). \end{aligned}$$

Bei der letzten Abschätzung ist erneut die gleichmäßige Beschränktheit der zweiten Momente von $\sqrt{n}(\hat{\gamma}(h) - \gamma(h))$ verwendet worden. (11.41) und das

geometrische Abklingen der a_j ergeben schließlich

$$\lim_{p \rightarrow \infty} \limsup_{n \rightarrow \infty} E|Z_{n,p} - Z_n| = 0,$$

sodass die Voraussetzung (A.66) aus Satz A.29 erfüllt ist.

Der Nachweis, dass $\tau^2(p)$ für $p \rightarrow \infty$ konvergent ist, sei dem Leser als Aufgabe 11.4 überlassen (zur Definition von $\tau^2(p)$ vgl. (11.55) bzw. (11.56)).

Damit haben wir die asymptotische Normalität von $\sqrt{n}(\hat{a}_1(p(n)) - a_1)$ bewiesen.

Die Tatsache, dass der Limes von $\tau^2(p)$ – wenn er denn existiert – tatsächlich wie in (11.52) gefordert 1 ist, wollen wir im Folgenden skizzieren.

Aus (11.59) zusammen mit der gleichmäßigen Beschränktheit der zweiten Momente von $\sqrt{n}(\hat{\gamma}(h) - \gamma(h))$ sowie weiterer Abschätzungen und mit Verwendung von (11.41) erhalten wir nach einiger Rechnung zunächst die folgende Darstellung (wir schreiben zur Abkürzung p statt $p(n)$)

$$\begin{aligned} & \sqrt{n}(\hat{a}_1(p) - a_1) \\ &= \frac{1}{\sigma_e^2} \sum_{r=1}^p a_{r-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t X_{t-r} + o_P(1) \\ &= \frac{1}{\sigma_e^2(p-1)} \sum_{r=1}^p a_{r-1}(p-1) \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t X_{t-r} + o_P(1). \end{aligned}$$

Da wir aufgrund von Aufgabe 11.2 wissen, dass die vierten Momente des ersten Terms auf der rechten Seite beschränkt sind und da wir bereits die Verteilungskonvergenz auch für diesen Ausdruck nachgewiesen haben, können wir mit Hilfe der Aussage (v) in Bemerkung A.26, angewendet auf das Quadrat des obigen Ausdrucks, die Varianz der asymptotischen Normalverteilung auf folgende Weise berechnen

$$\begin{aligned} & E \left(\frac{1}{\sigma_e^2(p-1)} \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \sum_{r=1}^p a_{r-1}(p-1) X_{t-r} \right)^2 \\ &= \frac{1}{\sigma_e^2(p-1)} \sum_{r,s=1}^p a_{r-1}(p-1) \gamma(r-s) a_{s-1}(p-1) \\ &= \sigma_e^2(p-1) \Gamma(p)_{1, \text{Zeile}}^{-1} \Gamma(p) \Gamma(p)_{1, \text{Zeile}}^{-1}, \text{ vgl. (11.40)} \\ &= \sigma_e^2(p-1) \Gamma(p)_{1, \text{Zeile}}^{-1} (1, 0, \dots, 0)^T \\ &= \sigma_e^2(p-1) \overline{\gamma_p(1, 1)}, \text{ vgl. erneut (11.40)} \\ &= a_0^2(p-1) = 1. \end{aligned}$$

Der Beweis zu Satz 11.16 ist abgeschlossen.

□

Bemerkung 11.17 (Asymptot. Konfidenzbereiche in $AR(\infty)$ -Modellen)

Wie schon auf Basis der anderen zentralen Grenzwertaussagen für Parameterschätzer, erlaubt auch Satz 11.16 die Ableitung von asymptotischen Konfidenzbereichen. Aus (11.52) zusammen mit (11.54) erhalten wir

$$\sqrt{n} A^{-1} (\hat{a}_j(p(n)) - a_j : j = 1, \dots, K) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_K), \quad (11.60)$$

wobei I_K die K -dimensionale Einheitsmatrix bezeichnet. Deshalb gilt

$$\|\sqrt{n} A^{-1} (\hat{a}_j(p(n)) - a_j : j = 1, \dots, K)\|^2 \xrightarrow{\mathcal{D}} \chi_K^2$$

und

$$\left\{ \mathbf{a} \in \mathbb{R}^K : \|\sqrt{n} A^{-1} (\hat{a}_j(p(n)) - a_j : j = 1, \dots, K)\|^2 \leq \chi_{K;1-\alpha}^2 \right\} \quad (11.61)$$

stellt einen asymptotischen $(1 - \alpha)$ -Konfidenzbereich für (a_1, \dots, a_K) dar. Genau die gleiche Aussage gilt mit \hat{A} anstelle von A , wobei \hat{A} dadurch entsteht, dass wir in der Matrix A alle auftretenden a_j durch deren Schätzer \hat{a}_j ersetzen. Der entstehende Konfidenzbereich hat dann die Form eines Ellipsoids. Für den Fall $K = 1$ verweisen wir auf Aufgabe 11.7.

□

Bevor wir uns im nächsten Abschnitt den $ARMA$ -Zeitreihen zuwenden, wollen wir noch auf die Analogie der asymptotischen Kovarianzmatrizen in (11.13) und (11.52) jeweils für den Yule-Walker-Schätzer hinweisen. Die Kovarianzmatrix in (11.13) gilt für den Fall eines $AR(p)$ -Modells und besitzt die Darstellung (11.39), die bei formalem Übergang für $p \rightarrow \infty$ in die Darstellung (11.54) der Kovarianzmatrix für den $AR(\infty)$ -Fall übergeht.

11.4 Parameterschätzung für ARMA-Zeitreihen

Im Gegensatz zu der Schätzung der Modellparameter für autoregressive Zeitreihen gestaltet sich die Schätzung für $ARMA$ -Zeitreihen deutlich komplizierter. Grundsätzlich können wir aber die gleichen Ansätze wie in den vorangegangenen Abschnitten verfolgen. So werden wir auch für $ARMA$ -Modelle eine Art Yule-Walker-Gleichung – und somit einen Zusammenhang zwischen der Autokovarianzfunktion und den Modellparametern – herleiten. Auch die Methode der kleinsten Quadrate bzw. der Maximum-Likelihood-Ansatz können zur Schätzung der $ARMA$ -Parameter herangezogen werden. In jedem Fall werden wir aber auf nichtlineare Gleichungen bzw. Optimierungsprobleme stoßen, die typischerweise mit numerischen Methoden gelöst werden müssen. Das Nichtvorhandensein von expliziten Darstellungen für diese Lösungen erschwert dann die Untersuchung der Eigenschaften der entstandenen Schätzer.

Nachdem wir überblicksartig auf diese Ansätze eingegangen sind, werden wir eine ganz andere Idee, die auf invertible ARMA-Zeitreihen (also ARMA-Zeitreihen mit einer $AR(\infty)$ -Darstellung) angewandt werden kann, verfolgen. Diese besteht darin, einen Anfangsteil der $AR(\infty)$ -Koeffizienten zu schätzen und auf dieser Basis auf die Parameter der ARMA-Zeitreihe zurückzuschließen. Für diese Überlegungen werden die Ergebnisse aus Abschnitt 11.3 äußerst hilfreich sein.

Betrachten wir also das $ARMA(p, q)$ -Modell

$$X_t + \sum_{j=1}^p a_j X_{t-j} = e_t + \sum_{k=1}^q b_k e_{t-k}, \quad \forall t \in \mathbb{Z}. \quad (11.62)$$

Dabei gehen wir von einer kausalen Situation aus, d.h. das Polynom $A(z) = 1 + a_1 z + \dots + a_p z^p$ hat ausschließlich Nullstellen mit Betrag größer als eins. X_t besitzt dann eine einseitige $MA(\infty)$ -Darstellung $\sum_{j=0}^{\infty} c_j e_{t-j}$ mit geometrisch abklingenden Koeffizienten ($c_j : j \in \mathbb{N}_0$).

Die folgende Rechnung haben wir schon in Abschnitt 8.3 durchgeführt: Eine Multiplikation von (11.62) mit X_{t-h} und anschließende Erwartungswertbildung liefert wegen der Unkorreliertheit von X_{t-h} mit e_s für alle $s > t - h$ die folgenden Gleichungen vom Yule-Walker-Typ

$$\gamma(h) + \sum_{j=1}^p a_j \gamma(h-j) = 0, \quad h = q+1, q+2, \dots \quad (11.63)$$

Aus der empirischen Autokovarianz $\hat{\gamma}$ für die Lags $h = q+1, \dots, q+p$ können somit unter Verwendung von (11.63) Schätzer für die autoregressiven Parameter a_1, \dots, a_p gewonnen werden. Wegen der linearen Struktur der Gleichungen (11.63) impliziert die Konsistenz von $\hat{\gamma}$ (vgl. etwa Satz 10.15) sofort entsprechende Aussagen für die Parameterschätzer. Wir werden darauf aber nicht genauer eingehen. Für die Lags $h = 0, 1, \dots, q$ erhalten wir aus (11.62)

$$\gamma(h) + \sum_{j=1}^p a_j \gamma(h-j) = \sum_{k=0}^q b_k E e_{t-k} X_{t-h}. \quad (11.64)$$

Mit Hilfe der oben erwähnten $MA(\infty)$ -Darstellung für X_t erhält man $E e_{t-k} X_{t-h} = c_{k-h} \cdot \sigma_e^2$ für $k \geq h$ und 0 sonst, also aus (11.64)

$$\gamma(h) + \sum_{j=1}^p a_j \gamma(h-j) = \sum_{k=h}^q b_k c_{k-h} \cdot \sigma_e^2, \quad h = 0, \dots, q. \quad (11.65)$$

Da nun schließlich die Koeffizienten c_j aus den ARMA-Parametern abgeleitet werden können (vgl. (8.14)), führt (11.65) zu nichtlinearen Gleichungen für b_1, \dots, b_q und σ_e^2 bei gegebenen Schätzern für a_1, \dots, a_p und $\gamma(0), \gamma(1), \dots, \gamma(p+q)$.

Beispiel 11.18 (Yule-Walker-Schätzung im $ARMA(1, 1)$ -Modell)

Die Spezialisierung von (11.63) und (11.65) auf den Fall $p = q = 1$ lautet

$$\begin{aligned}\gamma(2) + a\gamma(1) &= 0 \\ \gamma(1) + a\gamma(0) &= b \cdot \sigma_e^2 \\ \gamma(0) + a\gamma(1) &= (1 + b(b - a)) \cdot \sigma_e^2.\end{aligned}$$

Mit Hilfe der empirischen Autokovarianz kann direkt ein Schätzer für a bestimmt werden. Für b erhalten wir eine quadratische Gleichung.

Konkret ergeben sich aus einer simulierten $ARMA(1, 1)$ -Zeitreihe mit Parametern $a = 0.5$, $b = 0.25$ und $\sigma_e^2 = 1$ die folgenden empirischen Autokovarianzen

$$\hat{\gamma}(0) = 1.1547 \quad \hat{\gamma}(1) = -0.2789 \quad \hat{\gamma}(2) = 0.1476.$$

Hieraus erhalten wir die Schätzwerte $\hat{a} = 0.529$, $\hat{\sigma}_e^2 = 1.066$ sowie die beiden Lösungen 0.311 und 3.253 der quadratischen Gleichung für \hat{b} .

□

Auch die Methode der kleinsten Quadrate lässt sich auf $ARMA$ -Modelle anwenden. Hierfür ist es notwendig, aus Beobachtungen X_{1-p}, \dots, X_n der $ARMA$ -Zeitreihe die zugehörigen Terme des weißen Rauschens e_t so gut wie möglich zu rekonstruieren. Ein Ansatz besteht darin, einen Anfangsabschnitt e_{1-q}, \dots, e_0 des weißen Rauschens festzulegen (etwa $e_{1-q} = \dots = e_0 = 0$) und mit Hilfe der $ARMA$ -Gleichung (11.62) hieraus die Terme e_1, \dots, e_n des weißen Rauschens zumindest approximativ zu rekonstruieren. Dies wird rekursiv, beginnend mit e_1 , durchgeführt und führt für gegebenen Parameter \mathbf{a}' und \mathbf{b}' zu $e_t(\mathbf{a}', \mathbf{b}')$, $t = 1, \dots, n$. Nun werden \mathbf{a}' und \mathbf{b}' so gewählt, dass

$$\sum_{t=1}^n e_t(\mathbf{a}', \mathbf{b}')^2 \quad (11.66)$$

minimal wird. Aufgrund der komplexen Struktur der $e_t(\mathbf{a}', \mathbf{b}')$ können wir dieses Minimierungsproblem ebenfalls nur auf numerischem Weg lösen. Auch eine explizite Darstellung von $e_t(\mathbf{a}', \mathbf{b}')$, die wir im Folgenden ableiten, löst dieses grundsätzliche Dilemma nicht auf.

Falls das $ARMA$ -Modell (11.62) die Invertibilitätsbedingung $1 + \sum_{k=1}^q b_k z^k \neq 0$ für alle $|z| \leq 1$ erfüllt, existieren Koeffizienten $\beta_j = \beta_j(\mathbf{b})$, $j \in \mathbb{N}_0$, mit

$$\left(1 + \sum_{k=1}^q b_k z^k\right)^{-1} = \sum_{j=0}^{\infty} \beta_j z^j, \quad |z| \leq 1.$$

Eine Anwendung des linearen Filters $(\beta_j : j \in \mathbb{N}_0)$ auf den AR -Teil der $ARMA$ -Gleichung ergibt direkt

$$\sum_{j=0}^{\infty} \beta_j \left(X_{t-j} + \sum_{k=1}^p a_k X_{t-j-k} \right) = e_t.$$

Durch Abbrechen der Reihe auf der linken Seite bei $j = t$ erhalten wir

$$e_t(\mathbf{a}', \mathbf{b}') = \sum_{j=0}^t \beta_j(\mathbf{b}') \left(X_{t-j} + \sum_{k=1}^p a'_k X_{t-j-k} \right). \quad (11.67)$$

Auf die numerischen Methoden zur Minimierung von (11.66) wollen wir im Rahmen dieser Einführung nicht näher eingehen. Insbesondere für *ARMA*-Modelle kleinerer Ordnung ist dies aber relativ unproblematisch. Im folgenden Beispiel gehen wir auf das grundlegende *ARMA*(1, 1)-Modell ein.

Beispiel 11.19 (*LS-Schätzung im ARMA(1, 1)-Modell*)

Für $p = q = 1$ lautet (11.67)

$$e_t(a', b') = \sum_{j=0}^t (-b')^j (X_{t-j} + a' X_{t-j-1}), \quad t = 1, \dots, n,$$

Um im konkreten Fall einen *LS*-Schätzer zu bestimmen muss der Ausdruck $\sum_{t=1}^n e_t(a', b')^2$ für a', b' numerisch minimiert werden.

□

Unterstellen wir schließlich eine multivariate Normalverteilung für die zugrunde liegenden *ARMA*-Zeitreihen, so können wir grundsätzlich genauso, wie in Abschnitt 11.2 für autoregressive Zeitreihen beschrieben, vorgehen, um Maximum-Likelihood bzw. Quasi-Maximum-Likelihood-Schätzer zu erhalten. Letztlich benötigen wir eine explizite Darstellung der Autokovarianzmatrix $\Gamma(n) = (\gamma(r-s) : r, s = 1, \dots, n)$ von (X_1, \dots, X_n) in Abhängigkeit von den Parametern $a_1, \dots, a_p, b_1, \dots, b_q$ und σ_e^2 des *ARMA*-Modells, um eine entsprechende Likelihoodfunktion $L_n(a_1, \dots, a_p, b_1, \dots, b_q, \sigma_e^2)$ als gemeinsame Dichte von X_1, \dots, X_n aufstellen zu können. Durch numerische Maximierung dieser Funktion erhalten wir dann Quasi-*ML*-Parameterschätzer.

Etwas übersichtlicher ist die Betrachtung einer geeigneten bedingten Verteilung, nämlich

$$\begin{aligned} & \mathcal{L}(X_1, \dots, X_n \mid X_0, \dots, X_{1-p}, e_0, \dots, e_{1-q}) \\ &= \bigotimes_{s=1}^n \mathcal{L}(X_s \mid X_{s-1}, \dots, X_{1-p}, e_0, \dots, e_{1-q}), \end{aligned} \quad (11.68)$$

denn die bedingten Verteilungen auf der rechten Seite lassen sich unter einer Normalverteilungsannahme an die e_t relativ leicht angeben. Man beachte zunächst, dass wir durch sukzessives Verwenden der *ARMA*-Gleichung

(11.62) für $t = 1, \dots, s$ die Innovationen e_{s-1}, \dots, e_{s-q} explizit als Funktion von $X_{s-1}, \dots, X_{1-p}, e_0, \dots, e_{1-q}$ schreiben können. Damit gilt für $s = 1, \dots, n$

$$\begin{aligned} \mathcal{L}\left(X_s \middle| X_{s-1}, \dots, X_{1-p}, e_0, \dots, e_{1-q}\right) \\ = \mathcal{L}\left(e_s + \tilde{X}_s(\mathbf{a}, \mathbf{b}) \middle| X_{s-1}, \dots, X_{1-p}, e_0, \dots, e_{1-q}\right) \end{aligned} \quad (11.69)$$

$$\begin{aligned} \text{wobei } \tilde{X}_s(\mathbf{a}, \mathbf{b}) := - \sum_{j=1}^p a_j X_{s-j} + \sum_{j=1}^q b_j e_{s-j} \\ = \mathcal{N}\left(\tilde{X}_s(\mathbf{a}, \mathbf{b}), \sigma_e^2\right) \end{aligned} \quad (11.70)$$

Die bedingte Verteilung aus (11.68) ist dann nichts anderes als die Kopplung dieser Normalverteilungen für $s = 1, \dots, n$. Wir erhalten demnach die folgende bedingte Likelihoodfunktion

$$L_n(\mathbf{a}, \mathbf{b}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{n/2}} \exp\left(-\frac{1}{2} \sum_{s=1}^n (X_s - \tilde{X}_s(\mathbf{a}, \mathbf{b}))^2 / \sigma_e^2\right), \quad (11.71)$$

mit $\tilde{X}_s(\mathbf{a}, \mathbf{b})$ gemäß (11.70).

Eine numerische Maximierung der Log-Likelihood-Funktion $\ln L_n(\mathbf{a}, \mathbf{b}, \sigma_e^2)$ in \mathbf{a}, \mathbf{b} und σ_e^2 führt zu sogenannten bedingten Quasi-Maximum-Likelihood-Schätzern im $ARMA(p, q)$ -Modell. Die asymptotischen Eigenschaften dieser bedingten *QML*-Schätzer sind identisch mit denen von unbedingten *QML*-Schätzern in $ARMA$ -Modellen. Wir wollen hierauf im Rahmen dieser Einführung aber nicht weiter eingehen. Den interessierten Leser verweisen wir auf Hannan (1970), Brockwell/Davis (1991) und Straumann (2005) (hier insbesondere Theorem 4.1.1 und Corollary 4.1.4).

Nun wollen wir uns gründlicher mit einer weiteren Methode zur Gewinnung von Parameterschätzern in $ARMA$ -Zeitreihen beschäftigen. Die grundlegende Idee ist, die zum interessierenden $ARMA$ -Modell gehörige $AR(\infty)$ -Reihe aufzustellen, deren Koeffizienten wir mit den in Abschnitt 11.3 zur Verfügung gestellten Methoden schätzen können, und dann die $ARMA$ -Parameterschätzer aus den ersten $p + q$ $AR(\infty)$ -Koeffizientenschätzern zu gewinnen. Konkret können wir auf Basis eines invertiblen $ARMA(p, q)$ -Modells der Form (11.62) die folgende $AR(\infty)$ -Darstellung

$$\sum_{k=0}^{\infty} \alpha_k X_{t-k} = e_t, \quad \forall t \in \mathbb{Z}, \quad (11.72)$$

mit $\alpha_0 = 1$ und

$$a_k = \sum_{j=0}^{\min(k, q)} \alpha_{k-j} b_j, \quad k = 1, \dots, p, \quad (11.73)$$

sowie

$$-\alpha_k = \sum_{j=1}^{\min(k,q)} \alpha_{k-j} b_j, \quad k = p+1, \dots, p+q, \quad (11.74)$$

gewinnen. Es sei ausdrücklich darauf hingewiesen, dass (11.74) für alle $k > p$ gilt.

Verwenden wir nun die in Abschnitt 11.3 abgeleiteten Schätzer $\hat{\alpha}_k(p(n))$, $k = 1, \dots, p+q$, für die ersten $p+q$ Koeffizienten der $AR(\infty)$ –Entwicklung der Zeitreihe, so führt uns das lineare Gleichungssystem (11.74) im Falle der Lösbarkeit zunächst zu Schätzern $\hat{b}_1, \dots, \hat{b}_q$. Diese wiederum ergeben gemäß (11.73) direkt Schätzer $\hat{a}_1, \dots, \hat{a}_p$ für die autoregressiven Koeffizienten.

Es ist nahe liegend zu versuchen, die asymptotischen Eigenschaften der Schätzer $\hat{\alpha}_j(p(n))$, $j = 1, \dots, p+q$, gemäß Satz 11.16 auf die ARMA–Schätzer zu übertragen. Bevor wir uns dieser Fragestellung zuwenden, soll aber noch die grundsätzliche Lösbarkeit des linearen Gleichungssystems (11.74) untersucht werden.

Lemma 11.20

Falls für eine ARMA(p, q)–Zeitreihe X gemäß (11.62) gilt, dass die Polynome $A(z) = 1 + a_1 z + \dots + a_p z^p$ und $B(z) = 1 + b_1 z + \dots + b_q z^q$ keine gemeinsamen Nullstellen und nur solche mit $|z| > 1$ besitzen, dann ist b_1, \dots, b_q die einzige Lösung des linearen Gleichungssystems (11.74).

Beweis: Wir führen den Nachweis durch Widerspruch und nehmen dazu an, dass die $q \times q$ –Matrix

$$A := \begin{pmatrix} & & \alpha_{p+r-s} \\ r = 1, \dots, q, & s = 1, \dots, q \end{pmatrix},$$

wobei $\alpha_j := 0$ für $j < 0$ zu setzen ist, nicht vollen Rang besitzt. Dann existiert ein von Null verschiedener Vektor $\underline{d} := (d_1, \dots, d_q)^T$ mit $A\underline{d} = \underline{0}$. Durch Skalarmultiplikation können wir die Norm und damit die Komponenten von \underline{d} beliebig klein machen. Deshalb kann \underline{d} so gewählt werden, dass das ARMA–Modell mit den Parametern $b'_0 = 1$ und

$$b'_j := b_j + d_j, \quad j = 1, \dots, q, \quad a'_k := \sum_{j=0}^{\min(k,q)} \alpha_{k-j} b'_j, \quad k = 1, \dots, p \quad (11.75)$$

(vgl. mit (11.73)) die Voraussetzungen des Lemmas erfüllt. Dazu beachte man insbesondere, dass die Nullstellen eines Polynoms stetig von den Koeffizienten abhängen.

Wir werden nun im Widerspruch zu Satz 8.10 zeigen, dass die Ausgangszeitreihe X auch das ARMA–Modell mit den Parametern aus (11.75) erfüllt. Dazu

reicht es nachzuweisen, dass die jeweiligen $AR(\infty)$ -Darstellungen für beide Parameterkonstellationen identisch sind. Die Koeffizienten α'_k , $k = 0, 1, \dots$, ergeben sich wie folgt rekursiv ($\alpha'_0 = 0$ und $\alpha'_j := 0$ für $j < 0$)

$$\alpha'_k + b'_1 \alpha'_{k-1} + \dots + b'_q \alpha'_{k-q} = \begin{cases} a'_k, & k = 1, \dots, q \\ 0, & k > p \end{cases} \quad (11.76)$$

Wir zeigen nun, dass die Koeffizienten $(\alpha_k : k = 0, 1, \dots)$ (11.76) lösen. Für $k = 1, \dots, p$ folgt dies direkt aus der Definition der a'_k in (11.75). Für $k = p+1, \dots, p+q$ erhalten wir wegen $d_1 \alpha_{k-1} + \dots + d_q \alpha_{k-q} = 0$ (diese Aussage folgt direkt aus $a\bar{d} = 0$) die Gültigkeit von (11.76). Schließlich erhalten wir diese Aussage auch für alle $k > p+q$ (und damit (11.76) für alle $k \in \mathbb{N}$), denn es gilt (man verwende erneut $\alpha_j = 0$ für $j < 0$)

$$\begin{aligned} \sum_{\nu=1}^q d_\nu \alpha_{k-\nu} &= \sum_{\nu=1}^q d_\nu \left(- \sum_{j=1}^q b_j \alpha_{k-\nu-j} \right) \text{ beachte (11.74) für alle } k > p \\ &= - \sum_{j=1}^q b_j \left(\sum_{\nu=1}^q d_\nu \alpha_{k-\nu-j} \right), \end{aligned}$$

so dass wir die Behauptung per Induktion nach k erhalten. \square

Obwohl die grundsätzliche Struktur des Zusammenhangs der ersten $p+q$ $AR(\infty)$ -Koeffizienten und der $ARMA$ -Parameter gemäß (11.73) und (11.74) relativ einfach ist, gestaltet sich der Nachweis eines zentralen Grenzwertsatzes für das allgemeine $ARMA$ -Modell recht kompliziert. Wir wollen uns im Rahmen dieser Einführung deshalb mit dem Beispiel eines $ARMA(1, 1)$ -Modells und mit $MA(q)$ -Zeitreihen begnügen. Da für $AR(p)$ -Zeitreihen gemäß (11.73) $a_k = \alpha_k$, $k = 1, \dots, p$, gilt, beschreibt Satz 11.16 die Asymptotik für auf dem hier beschriebenen Weg abgeleitete Schätzer direkt. Auf eine gesonderte Darstellung des reinen AR -Falls wird deshalb an dieser Stelle verzichtet. Kommen wir als erstes zum $ARMA(1, 1)$ -Modell.

Beispiel 11.21 ($ARMA(1, 1)$ -Parameterschätzung)

Gegeben sei eine $ARMA(1, 1)$ -Zeitreihe gemäß (11.62) mit $p = q = 1$ und Parametern a und b . Unter der Voraussetzung der Invertibilität, d.h. $|b| < 1$, besitzt die $ARMA(1, 1)$ -Zeitreihe die folgende $AR(\infty)$ -Darstellung

$$X_t + \sum_{j=1}^{\infty} \alpha_j X_{t-j} = e_t, \quad \forall t \in \mathbb{Z},$$

wobei $(1+az)/(1+bz) = 1 + \sum_{j=1}^{\infty} \alpha_j z^j$ für $|z| \leq 1$. Also speziell

$$\alpha_1 = a - b \quad \text{und} \quad \alpha_2 = -b\alpha_1. \quad (11.77)$$

Ausgehend von den Parameterschätzern $\hat{\alpha}_1(p(n))$ und $\hat{\alpha}_2(p(n))$ für α_1 und α_2 im $AR(\infty)$ -Modell gemäß Abschnitt 11.3 erhalten wir die folgenden $ARMA(1, 1)$ -Parameterschätzer

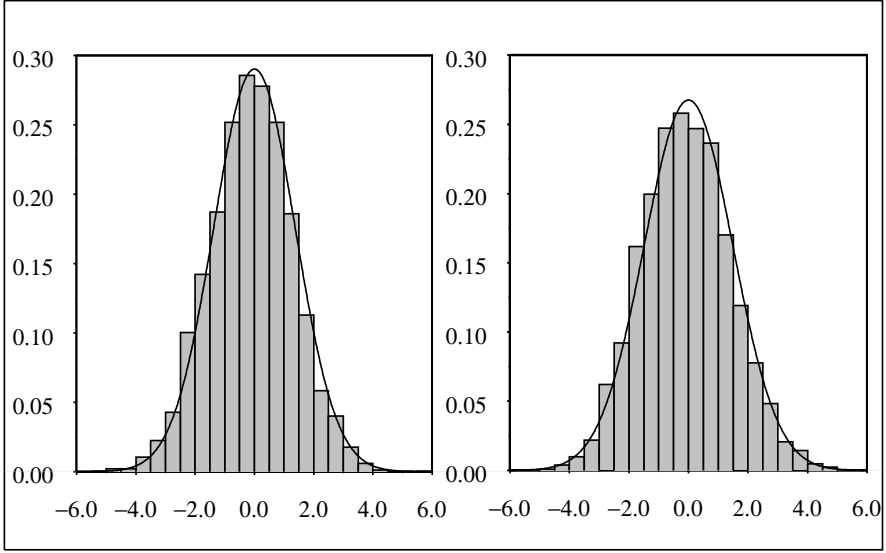


Abb. 11.4 Simulierte Histogramme (5000 Wiederholungen) der zentrierten und standardisierten Schätzer $\sqrt{n}(\hat{a} - a)$ (links) und $\sqrt{n}(\hat{b} - b)$ (rechts) im ARMA(1, 1)-Modell mit den Parametern $a = 0.50, b = -0.25$ sowie Stichprobenumfang $n = 500$ zusammen mit den Dichten der jeweiligen asymptotischen Normalverteilungen

$$\hat{b} = -\frac{\hat{\alpha}_2(p(n))}{\hat{\alpha}_1(p(n))} \quad \text{sowie} \quad \hat{a} = \hat{\alpha}_1(p(n)) + \hat{b}.$$

Unter den Voraussetzungen von Satz 11.16 gilt zunächst

$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_1(p(n)) - \alpha_1 \\ \hat{\alpha}_2(p(n)) - \alpha_2 \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma), \quad (11.78)$$

wobei $\Sigma = \begin{pmatrix} 1 & \alpha_1 \\ \alpha_1 & 1 + \alpha_1^2 \end{pmatrix}$ (vgl. (11.52) und (11.53)).

Da für $g_1(u, v) := u - v/u$ und $g_2(u, v) := -v/u$ folgende Beziehung gilt

$$\hat{a} = g_1(\hat{\alpha}_1(p(n)), \hat{\alpha}_2(p(n))), \quad \hat{b} = g_2(\hat{\alpha}_1(p(n)), \hat{\alpha}_2(p(n))),$$

erhalten wir mit Hilfe der Delta-Methode aus (11.78) (vgl. Satz A.31) die folgende asymptotische Normalität von (\hat{a}, \hat{b}) :

$$\sqrt{n} \begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{\alpha_1^4} \begin{pmatrix} \alpha_1^2 + \alpha_2^2 & \alpha_1^2 - \alpha_1^2 \alpha_2 + \alpha_2^2 \\ \alpha_1^2 - \alpha_1^2 \alpha_2 + \alpha_2^2 & \alpha_1^4 + \alpha_1^2 - 2\alpha_1^2 \alpha_2 + \alpha_2^2 \end{pmatrix} \right)$$

bzw. durch Einsetzen von (11.77)

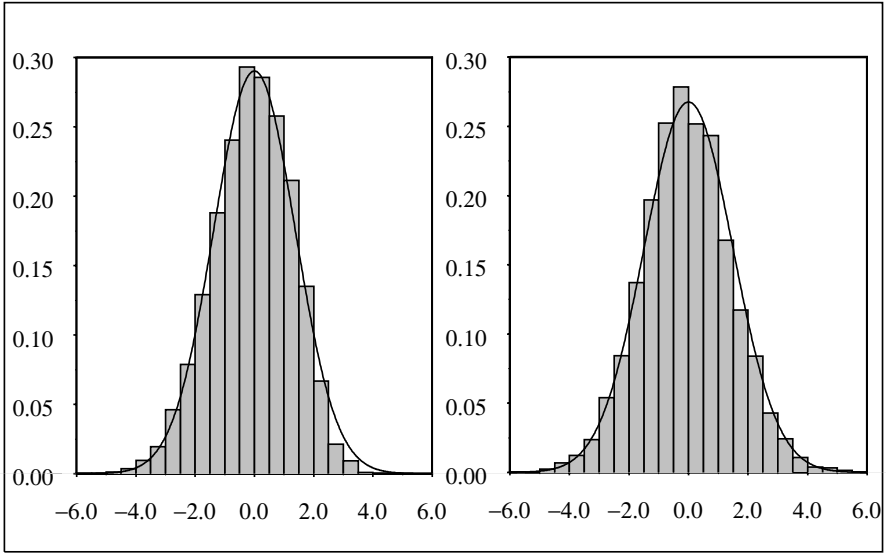


Abb. 11.5 Simulierte Histogramme (5000 Wiederholungen) der zentrierten und standardisierten *QML*-Schätzer $\sqrt{n}(\hat{a}_{ML} - a)$ (links) und $\sqrt{n}(\hat{b}_{ML} - b)$ (rechts) im $ARMA(1,1)$ -Modell mit den Parametern $a = 0.50, b = -0.25$ sowie Stichprobenumfang $n = 500$ zusammen mit den Dichten der jeweiligen asymptotischen Normalverteilungen

$$\sqrt{n} \begin{pmatrix} \hat{a} - a \\ \hat{b} - b \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{(a-b)^2} \begin{pmatrix} 1+b^2 & 1+ab \\ 1+ab & 1+a^2 \end{pmatrix} \right). \quad (11.79)$$

An der expliziten Darstellung der asymptotischen Kovarianzmatrix erkennt man, dass der Fall $a \approx b$ zu sehr hohen Varianzen sowohl für den Schätzer für a wie für den Schätzer für b führt. Außerdem ist die asymptotische Kovarianzmatrix in (11.79) dann nahe an der Singularität. Es sei darauf hingewiesen, dass der Fall $a = b \neq 0$ zur Nichtidentifizierbarkeit des $ARMA(1,1)$ -Modells führt (X_t ist in diesem Fall nämlich nicht von einem weißen Rauschen zu unterscheiden) und deshalb im Rahmen der Parameterschätzung ausgeschlossen werden muss (vgl. mit Satz 8.10).

Selbstverständlich kann man auf Basis der zentralen Grenzwertaussage (11.79) asymptotische $(1 - \alpha)$ -Konfidenzbereiche für (a, b) ableiten. Konkret erhält man etwa für den *AR*-Parameter a das folgende asymptotische Konfidenzintervall

$$\hat{a} \pm \frac{(1 + \hat{b}^2)^{1/2}}{\sqrt{n} |\hat{a} - \hat{b}|} \cdot z_{1-\alpha/2}. \quad (11.80)$$

Um die praktische Relevanz des vorgeschlagenen Vorgehens beurteilen zu können, wollen wir eine kleine Simulationsstudie betrachten. Aus 5000 simu-

lierten $ARMA(1, 1)$ -Zeitreihen jeweils mit der Länge $n = 500$ und Parametern $a = 0.5$ und $b = -0.25$ haben wir auf Basis einer Anpassung eines $AR(5)$ -Modells (also $p(n) = 5$) Schätzer \hat{a} und \hat{b} berechnet. In Abbildung 11.4 sind die resultierenden Histogramme der mit dem tatsächlichen Wert zentrierten und mit \sqrt{n} standardisierten Schätzer zusammen mit der entsprechenden Dichte der asymptotischen Normalverteilung (vgl. (11.79)) angegeben. Man erkennt eine sehr gute Übereinstimmung der simulierten Histogramme mit der asymptotischen Verteilung.

Um einen Eindruck von der Güte dieser Schätzer zu erhalten, vergleichen wir sie mit den aus der S-PLUS[®]-Routine `arima.mle` bestimmten *QML*-Schätzern für das $ARMA(1, 1)$ -Modell. Als Startwerte für die numerische Prozedur wurden die tatsächlichen Parameterwerte benutzt. Das zu Abbildung 11.4 korrespondierende Simulationsergebnis für die *QML*-Schätzer ist in Abbildung 11.5 angegeben. Es scheint (jedenfalls für den hier betrachteten Stichprobenumfang $n = 500$) keine wesentlichen Unterschiede der beiden Prozeduren zu geben.

Zum Abschluss der Simulationsstudie wollen wir uns noch das Verhalten des *QML*-Schätzers im $ARMA(1, 1)$ -Modell für den Fall $a = 0.50$ und $b = 0.48$ (ebenfalls $n = 500$) ansehen. Die genannten Parameter sind nahe beieinander und dies bedeutet, dass das zugehörige $ARMA(1, 1)$ -Modell nahe bei der Situation $a = b$ der Nichtidentifizierbarkeit liegt (vgl. mit Satz 8.10). Für 100 simulierte Zeitreihen haben wir die *QML*-Schätzer berechnet und die Ergebnisse in einem $\hat{a}_{ML} - \hat{b}_{ML}$ -Plot in Abbildung 11.6 dargestellt. Man erkennt deutlich, dass die jeweiligen Schätzer zwar recht gut die Aussage $\hat{a}_{ML} \approx \hat{b}_{ML}$ erfüllen, aber recht wahllos zwischen -1 und $+1$ hin und her springen. \square

Nun wollen wir uns den reinen $MA(q)$ -Zeitreihen zuwenden.

Beispiel 11.22 (Parameterschätzung in $MA(q)$ -Zeitreihen)

$(X_t : t \in \mathbb{Z})$ sei eine invertible $MA(q)$ -Zeitreihe mit Parametern b_1, \dots, b_q . Die $AR(\infty)$ -Darstellung dieser Zeitreihe lautet

$$X_t + \sum_{j=1}^{\infty} \alpha_j X_{t-j} = e_t,$$

wobei die Koeffizienten α_j wie folgt bestimmt sind

$$\alpha_k + b_1 \alpha_{k-1} + \dots + b_q \alpha_{k-q} = 0, \quad k = 1, 2, \dots \quad (11.81)$$

Hierbei ist natürlich $\alpha_r = 0$ für $r < 0$ und $\alpha_0 = 1$ zu beachten. Aus dem Gleichungssystem (11.81) für $k = 1, \dots, q$ können offensichtlich die Werte b_1, \dots, b_q bei gegebenen $\alpha_1, \dots, \alpha_k$ rekursiv bestimmt werden. Konkret erhalten wir für geeignete stetig differenzierbare Funktionen g_1, \dots, g_q die Darstellung

$$b_k = g_k(\alpha_1, \dots, \alpha_k), \quad k = 1, \dots, q. \quad (11.82)$$

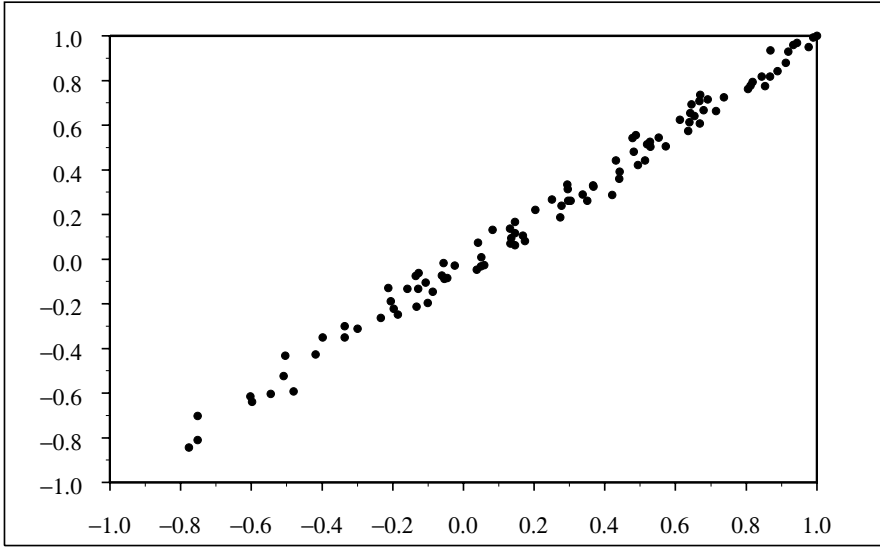


Abb. 11.6 Simulierter $\hat{a}_{ML} - \hat{b}_{ML}$ -Plot (100 Wiederholungen) der QML-Schätzer \hat{a}_{ML} (x -Achse) und \hat{b}_{ML} (y -Achse) im $ARMA(1, 1)$ -Modell mit den Parametern $a = 0.50, b = 0.48$ sowie Stichprobenumfang $n = 500$

Wir erhalten etwa für $k = 2$ aus (11.81)

$$b_2 = -\alpha_1 b_1 - \alpha_2 = \alpha_1^2 - \alpha_2.$$

Damit also $g_2(u, v) = u^2 - v$. Mit Hilfe der Yule-Walker-Schätzer $\hat{\alpha}_k(p(n))$ (kurz: $\hat{\alpha}_k$), $k = 1, \dots, q$, für $\alpha_1, \dots, \alpha_q$ erhalten wir durch direktes Einsetzen in (11.81) bzw. (11.82) Schätzer $\hat{b}_1, \dots, \hat{b}_q$ für die gesuchten MA-Koeffizienten. Der zentrale Grenzwertsatz 11.16 ergibt zusammen mit der Delta-Methode die asymptotische Normalität von $(\hat{b}_1, \dots, \hat{b}_q)^T$. Die asymptotische Kovarianzmatrix lautet (vgl. (11.52) und Satz A.31)

$$\begin{pmatrix} \frac{\partial g_k}{\partial \alpha_j}(\alpha_1, \dots, \alpha_k) \\ k = 1, \dots, q; j = 1, \dots, q \end{pmatrix} \begin{pmatrix} \Sigma(r, s) \\ r, s = 1, \dots, q \end{pmatrix} \begin{pmatrix} \frac{\partial g_j}{\partial \alpha_k}(\alpha_1, \dots, \alpha_j) \\ k = 1, \dots, q; j = 1, \dots, q \end{pmatrix}. \quad (11.83)$$

Zur Berechnung dieser Kovarianzmatrix zeigen wir mit Hilfe von (11.53) und (11.54), dass

$$\begin{pmatrix} \frac{\partial g_k}{\partial \alpha_j}(\alpha_1, \dots, \alpha_k) \\ k = 1, \dots, q; j = 1, \dots, q \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ \alpha_1 & \ddots & \ddots & \dots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{q-1} & \dots & \alpha_1 & 1 \end{pmatrix} = - \begin{pmatrix} 1 & 0 & \dots & 0 \\ b_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ b_{q-1} & \dots & b_1 & 1 \end{pmatrix} \quad (11.84)$$

gilt. Zusammen mit (11.83) folgt hieraus, dass die asymptotische Kovarianzmatrix von $(\hat{b}_1, \dots, \hat{b}_q)^T$ gleich

$$\begin{pmatrix} \sum_{j=1}^{\min(r,s)} b_{r-j} b_{s-j} \\ r, s = 1, \dots, q \end{pmatrix} \quad (11.85)$$

ist. Diese Darstellung korrespondiert zu der Darstellung der Kovarianzmatrix im $AR(\infty)$ -Modell (vgl. Satz 11.16 und insbesondere (11.52)), wobei lediglich die autoregressiven Koeffizienten durch diejenigen des MA -Prozesses ersetzt werden. Ein vergleichbares Resultat für $MA(\infty)$ -Zeitreihen findet man bei Brockwell/Davis (1991), Theorem 8.3.1.

Speziell für den $MA(1)$ -Prozess erhalten wir

$$\sqrt{n}(\hat{b}_1 - b_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \quad (11.86)$$

Abschließend ist nun noch (11.84) zu beweisen. Hierfür genügt es zu zeigen, dass für alle k die Gleichungen

$$\frac{\partial g_k}{\partial \alpha_j} + \alpha_1 \frac{\partial g_k}{\partial \alpha_{j+1}} + \dots + \alpha_{k-j} \frac{\partial g_k}{\partial \alpha_k} = -b_{k-j}, \quad j = 1, \dots, k, \quad (11.87)$$

erfüllt sind. Die Gültigkeit dieser Aussage erhalten wir per Induktion. Zunächst gilt für $k = 1$

$$\frac{\partial g_1}{\partial \alpha_1} = \frac{\partial}{\partial \alpha_1}(-\alpha_1) = -1 = -b_0.$$

Für den Induktionsschritt werde nun die Gültigkeit von (11.87) bis zu einem festen k angenommen. Da stets

$$\frac{\partial g_{k+1}}{\partial \alpha_{k+1}} = -1 = -b_0$$

gilt, reicht es die Gültigkeit von (11.87) mit $k + 1$ an Stelle von k für alle $j = 1, \dots, k$ nachzuweisen.

Auf Konfidenzbereiche für die Parameter des $MA(q)$ -Modells gehen wir nicht ein.

□

Aufgaben

Aufgabe 11.1 Seien $(X_{1n}, \dots, X_{kn})^T \sim \mathcal{N}(0, S_n)$, $n \geq 1$, und es gelte $X_{jn} \rightarrow X_j \forall j = 1, \dots, k$ im quadratischen Mittel für $n \rightarrow \infty$. Man zeige, dass dann auch $(X_1, \dots, X_k)^T$ normalverteilt ist.

Aufgabe 11.2 $(X_t : t \in \mathbb{Z})$ sei eine kausale $AR(p)$ -Zeitreihe mit i.i.d. weißem Rauschen (e_t) , für welches $Ee_1^4 < \infty$ gilt. Man zeige, dass für jedes $p_0 \geq p$ gilt

$$E \left\| \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t (X_{t-1}, \dots, X_{t-p_0})^T \right\|^4 < \infty.$$

Hieraus folgere man, dass die asymptotische Kovarianzmatrix im Beweis von Satz 11.4 gemäß (11.17) bestimmt werden kann.

Hinweis: Man überlege sich, dass man sich ohne größere Einschränkung auf den Fall $p_0 = p = 1$ beschränken kann und verwende die $MA(\infty)$ -Darstellung. Dann zeige man zunächst, dass $E(\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t e_{t-k})^4 \leq C < \infty \quad \forall k \geq 1$ gilt. Dies geschieht durch Ausmultiplikation der vierten Potenz. Es bleiben Erwartungswerte der Form $Ee_s^2 e_{s-k}^2 e_t^2 e_{t-k}^2$ durch eine Konstante abzuschätzen. Durch zweimalige Anwendung der Cauchy-Schwarz-Ungleichung erkennt man, dass $(Ee_1^4)^2$ eine solche Konstante ist.

Aufgabe 11.3 Man zeige per Induktion die Gültigkeit von (11.39).

Aufgabe 11.4 Man zeige $\tau^2(p) \rightarrow 1$ für $p \rightarrow \infty$ im Rahmen des Beweises zu Satz 11.16.

Aufgabe 11.5 Man verifiziere (11.73) und (11.74).

Aufgabe 11.6 Man zeige, dass die bedingte Likelihood-Funktion im normalverteilten $ARMA(p, q)$ -Modell unter der Annahme $X_0 = \dots = X_{1-p} = 0$, $e_0 = \dots = e_{1-q} = 0$ in der folgenden Form geschrieben werden kann

$$L_n(\mathbf{a}, \mathbf{b}, \sigma_e^2) = \frac{1}{(2\pi\sigma_e^2)^{n/2}} \exp \left(-\frac{1}{2} \sum_{s=1}^n (X_s - \tilde{X}_s(\mathbf{a}, \mathbf{b}))^2 / \sigma_e^2 \right),$$

wobei für $s = 1, \dots, n$

$$\begin{aligned} \tilde{X}_s(\mathbf{a}, \mathbf{b}) &= - \sum_{j=1}^{\min(p, s-1)} a_j X_{s-j} + \sum_{j=1}^{\min(q, s-1)} b_j \tilde{e}_{s-j}(\mathbf{a}, \mathbf{b}) \\ \tilde{e}_s(\mathbf{a}, \mathbf{b}) &= X_s + \sum_{j=1}^{\min(p, s-1)} a_j X_{s-j} - \sum_{j=1}^{\min(q, s-1)} b_j \tilde{e}_{s-j}(\mathbf{a}, \mathbf{b}). \end{aligned}$$

Aufgabe 11.7 Man gebe für den Parameter a_1 einer kausalen $AR(p)$ -Reihe mit i.i.d. reellem, weißem Rauschen (e_t) , welches $Ee_1^4 < \infty$ erfüllt, ein asymptotisches $(1 - \alpha)$ -Konfidenzintervall an. Was ergibt sich in den Spezialfällen $p = 1$ bzw. $p > 1$ und $a_2, \dots, a_p = 0$?

Hinweis: Lemma 11.13 und Satz 11.4.

Aufgabe 11.8 Man leite aus (11.28) unter einer Normalverteilungsannahme her, dass der ML -Schätzer $\hat{\sigma}_e^2$ für σ_e^2 im kausalen $AR(p)$ -Modell die Gleichung

$$\hat{\sigma}_e^2 = \frac{1}{n} \left[\underline{X}_0^T \hat{G}(p)^{-1} \underline{X}_0 + \sum_{t=1}^n \left(X_t + \sum_{j=1}^p \hat{a}_j X_{t-j} \right)^2 \right]$$

erfüllt. Dabei bezeichnen $\hat{a}_1, \dots, \hat{a}_p$ die ML -Schätzer für a_1, \dots, a_p und mit $G(p) = \sigma_e^{-2} \Gamma(p)$ sei $\hat{G}(p)$ diejenige Matrix, die entsteht, wenn die Kovarianzmatrix $\Gamma(p)$ explizit als Funktion von a_1, \dots, a_p geschrieben und dann die a_j durch die \hat{a}_j ersetzt werden.

Aufgabe 11.9 Man betrachte das bei $t = 0$ startende $AR(1)$ -Modell

$$X_0 := e_0 \text{ und } X_t + aX_{t-1} = e_t, \quad t = 1, 2, \dots,$$

für $|a| > 1$. Die e_t seien i.i.d. mit $e_t \sim \mathcal{N}(0, 1)$. Man zeige, dass der nur von zwei Beobachtungen X_n und X_{n+1} abhängende Schätzer $\hat{a}_n := -X_{n+1}/X_n$ die Eigenschaft

$$|a|^n \cdot |\hat{a}_n - a| = \mathcal{O}_P(1)$$

besitzt.

Hinweis: Für zwei stochastisch unabhängige, standard normalverteilte Zufallsvariablen U und V besitzt U/V eine Cauchy-Verteilung.

Schätzen im Spektralbereich

In Kapitel 3 haben wir gesehen, dass sich eine Reihe von charakteristischen Eigenschaften einer stationären Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit Hilfe der Spektraldarstellung ausdrücken, gut veranschaulichen und interpretieren lassen. Ganz konkret bestimmt etwa das Spektralmaß einer stationären Zeitreihe eindeutig die zugehörige Autokovarianzfunktion (vgl. Satz 3.4, Herglotz-Lemma). D.h. alle in der Autokovarianzfunktion enthaltenen Informationen spiegeln sich vollständig im Spektralmaß wider. Als besonders vorteilhaft hat sich das Spektralmaß erwiesen, wenn es um die Darstellung der Zeitreihe als Überlagerung von Schwingungen in Form eines stochastischen Integrals geht (vgl. Satz 5.7, Spektralsatz). Hier gibt das Spektralmaß gerade die Verteilung über die verschiedenen Frequenzen an. Ganz besonders wichtig für die Anwendung ist natürlich der Fall, in dem wir eine Dichte f_X des Spektralmaßes vorliegen haben. Die Inversionsformel (Satz 3.8) sichert uns stets eine solche (dann sogar stetige) Spektraldichte, wenn nur die zugehörige Autokovarianzfunktion γ_X absolut summierbar ist. In diesem Fall gilt die Darstellung

$$f_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \gamma_X(h) e^{-ih\omega}, \quad \forall \omega \in [-\pi, \pi]. \quad (12.1)$$

Für reellwertige stationäre Zeitreihen X kann man sich wegen der Symmetrie von γ_X grundsätzlich auf Frequenzen $0 \leq \omega \leq \pi$ beschränken. Wegen $\int_{-\pi}^{+\pi} f_X(\omega) d\omega = \gamma_X(0)$ führt eine Ersetzung von γ_X in (12.1) durch die Autokorrelationsfunktion ϱ_X zu einer normierten Form der Spektraldichte als Wahrscheinlichkeitsdichte, was häufig aus Darstellungs- und Interpretationsgründen von Vorteil ist.

Gehen wir im Folgenden von Realisierungen X_1, \dots, X_n einer reellwertigen Zeitreihe aus, so erhalten wir den schon wiederholt verwendeten Schätzer

$$\hat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \bar{X}_n)(X_{t+|h|} - \bar{X}_n), \quad |h| < n, \quad (12.2)$$

für die Autokovarianzen $\gamma(h)$ mit $|h| < n$. Setzt man nun noch $\hat{\gamma}_n(h) = 0$ für $|h| \geq n$, so kann man leicht zeigen, dass $\hat{\gamma}_n$ positiv semidefinit ist (siehe Aufgaben 8.3, 12.1, 12.2), also selbst die Kovarianzfunktion einer stationären Zeitreihe ist, deren Spektraldichte als Schätzung für die theoretische Spektraldichte f_X ganz natürlich erscheint. Die so zu $\hat{\gamma}_n$ gehörende Spektraldichte ist nach der Inversionsformel Satz 3.8 gerade

$$I_{n, \bar{X}_n}(\omega) := \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_n(h) e^{-ih\omega}, \quad -\pi \leq \omega \leq \pi, \quad (12.3)$$

wobei der Index \bar{X}_n darauf hindeutet, dass zentrierte Beobachtungen $X_t - \bar{X}_n$

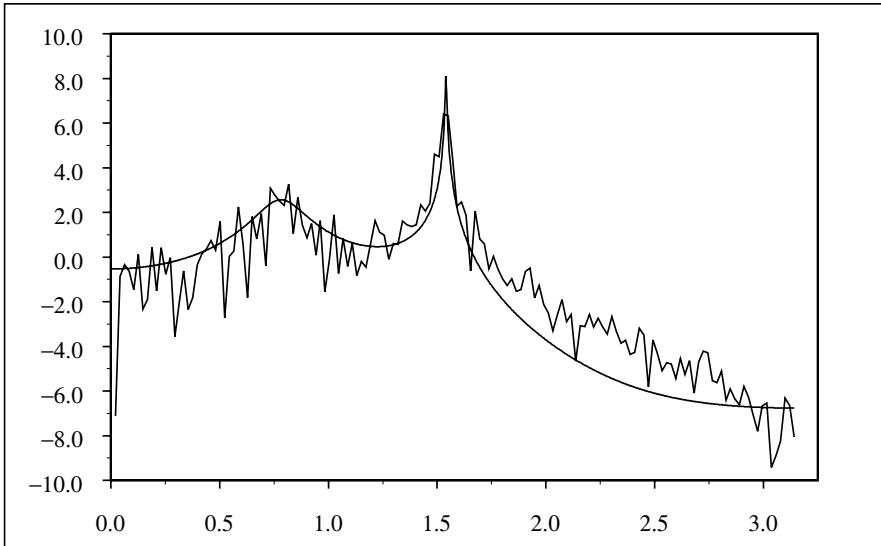


Abb. 12.1 Logarithmiertes Periodogramm einer simulierten $ARMA(4, 2)$ -Zeitreihe der Länge $n = 300$ und tatsächliche logarithmierte Spektraldichte (glatte Kurve)

zur Berechnung verwendet werden. Da $\hat{\gamma}_n$ für $|h| \geq n$ verschwindet, ist die zu $\hat{\gamma}_n$ gehörige Zeitreihe nach Lemma 9.10 eine $MA(q)$ -Reihe mit $q = n - 1$. Genauer gilt mit einem beliebigen weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t^2 = 1/n$ und Koeffizienten $b_j = (X_{j+1} - \bar{X}_n)$, dass $Y_t = \sum_{j=0}^{n-1} b_j e_{t-j}$ eine Zeitreihe mit Autokovarianzfunktion $\hat{\gamma}_n$ und Spektraldichte (12.3) ist.

Nach dem Spektralsatz gilt wegen $Y_t = \int (\sum_{j=0}^{n-1} b_j e^{-ij\cdot}) e^{it\cdot} dZ_e$, dass die Spektraldichte von Y_t auch in der folgenden Form geschrieben werden kann

$$I_{n,\bar{X}}(\omega) = \frac{1}{2\pi n} \left| \sum_{j=0}^{n-1} b_j e^{-ij\omega} \right|^2 = \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \bar{X}_n) e^{-it\omega} \right|^2. \quad (12.4)$$

Bemerkung 12.1

Ersetzt man in (12.2) das Stichprobenmittel \bar{X}_n durch eine beliebige reelle Zahl μ gemäß

$$\hat{\gamma}_{n,\mu}(h) := \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \mu)(X_{t+|h|} - \mu), \quad |h| < n, \quad (12.5)$$

und definiert wieder $\hat{\gamma}_{n,\mu}(h) := 0$ für $|h| \geq n$, so ist auch $\hat{\gamma}_{n,\mu}$ positiv semidefinit und es gilt mit denselben Überlegungen wie oben für $-\pi \leq \omega \leq \pi$ die Gleichheit

$$I_{n,\mu}(\omega) := \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_{n,\mu}(h) e^{-ih\omega} = \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \mu) e^{-it\omega} \right|^2. \quad (12.6)$$

Traditionell wird für den Beweis von (12.6) in der Literatur die folgende direkte Rechnung durchgeführt, bei der Formel (A.6) aus Anhang A.1 verwendet wird.

$$\begin{aligned} \frac{1}{2\pi n} \left| \sum_{t=1}^n (X_t - \mu) e^{-it\omega} \right|^2 &= \frac{1}{2\pi n} \sum_{s,t=1}^n (X_t - \mu)(X_s - \mu) e^{-i(t-s)\omega} \\ &= \frac{1}{2\pi n} \sum_{h=0}^{n-1} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \mu)(X_t - \mu) e^{-ih\omega} \\ &\quad + \frac{1}{2\pi n} \sum_{h=-1}^{-(n-1)} \sum_{t=1}^{n-|h|} (X_t - \mu)(X_{t+|h|} - \mu) e^{-ih\omega} \\ &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \mu)(X_{t+|h|} - \mu) e^{-ih\omega} \\ &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \hat{\gamma}_{n,\mu}(h) e^{-ih\omega}. \end{aligned} \quad (12.7)$$

Man nennt I_{n,\bar{X}_n} das (*zentrierte*) *Periodogramm* und $I_n := I_{n,0}$ einfach *Periodogramm* der Beobachtungen X_1, \dots, X_n .

In Abbildung 12.1 ist für eine simulierte $ARMA(4,2)$ -Zeitreihe der Länge $n = 300$ mit ausgeprägtem peak der Spektraldichte in der Frequenz $\omega = \pi/2$ und moderatem peak in der Frequenz $\omega = \pi/4$ die tatsächliche Spektraldichte und das Periodogramm der simulierten Daten dargestellt. Um die Form der Kurven besser deutlich zu machen wurden jeweils logarithmierte Werte aufgetragen. \square

Im Abschnitt 12.2 werden wir uns ausführlich mit dem Periodogramm I_n und seiner Interpretation auseinander setzen, bevor wir im Abschnitt 12.3 (asymptotische) Eigenschaften des Periodogramms, zunächst für den einfachen Fall eines weißen Rauschens, danach aber auch für allgemeine lineare Prozesse, herleiten. Dort werden wir sehen, dass das Periodogramm zwar einen asymptotisch erwartungstreuen aber *nicht* konsistenten Schätzer für die Spektraldichte darstellt. Deshalb stellen wir in den Abschnitten 12.4 und 12.5 Modifikationen in Form von Glättungen des Periodogramms vor, die zu konsistenten Schätzern der Spektraldichte führen.

Die erzielten Grenzwertaussagen erlauben uns die Konstruktion von Konfidenzintervallen für die Spektraldichte zum vorgegebenen asymptotischen Niveau. Hierauf gehen wir in Abschnitt 12.6 ein, bis wir schließlich in Abschnitt 12.7 sogenannte integrierte Periodogrammschätzer und ihre Eigenschaften vorstellen werden.

Bevor wir aber mit den nichtparametrischen Spektraldichteschätzern, die über entsprechende Glättungen des Periodogramms erklärt sind, beginnen, wollen wir noch einen ebenfalls naheliegenden modellbasierten sogenannten parametrischen Ansatz über anzupassende AR - oder $ARMA$ -Modelle im ersten Abschnitt 12.1 vorstellen.

12.1 Parametrische Spektraldichteschätzung

Sehr ausführlich haben wir uns in Kapitel 11 mit der Schätzung der Modellparameter in $ARMA$ -Modellen auseinander gesetzt. Selbstverständlich können wir die dort untersuchten Schätzer auch im Rahmen der Spektraldichteschätzung für $ARMA$ -Zeitreihen verwenden. Die Idee ist einfach, die entsprechenden Parameterschätzer in den in Satz 7.6 hergeleiteten Ausdruck für die Spektraldichte einzusetzen, um auf diese Weise einen sogenannten parametrischen Spektraldichteschätzer zu gewinnen. Bezeichnen wir die zugehörigen Parameterschätzer für ein $ARMA(p, q)$ -Modell mit $\hat{a}_1, \dots, \hat{a}_p$ beziehungsweise mit $\hat{b}_1, \dots, \hat{b}_q$ und sei $\hat{\sigma}_e^2$ ein Schätzer für die Varianz des zugehörigen weißen Rauschens, so lauten die zugehörigen parametrischen Spektraldichteschätzer im reinen $MA(q)$ -Fall, also für $p = 0$, wie folgt:

$$\hat{f}_{MA}(\omega) = \frac{\hat{\sigma}_e^2}{2\pi} \left| 1 + \sum_{j=1}^q \hat{b}_j e^{-ij\omega} \right|^2, \quad \omega \in [-\pi, \pi]. \quad (12.8)$$

Für den Fall, dass das geschätzte Polynom $1 + \sum_{j=1}^p \hat{a}_j z^j$ in der komplexen Variablen z keine Nullstelle vom Betrag 1 besitzt (und dies wird ja bei Verwendung der Yule-Walker-Parameterschätzer im autoregressiven Modell stets garantiert (vgl. Satz 8.7)), so können wir auch für reine AR - und $ARMA$ -Modelle entsprechende parametrische Spektraldichteschätzer in der folgenden Form angeben.

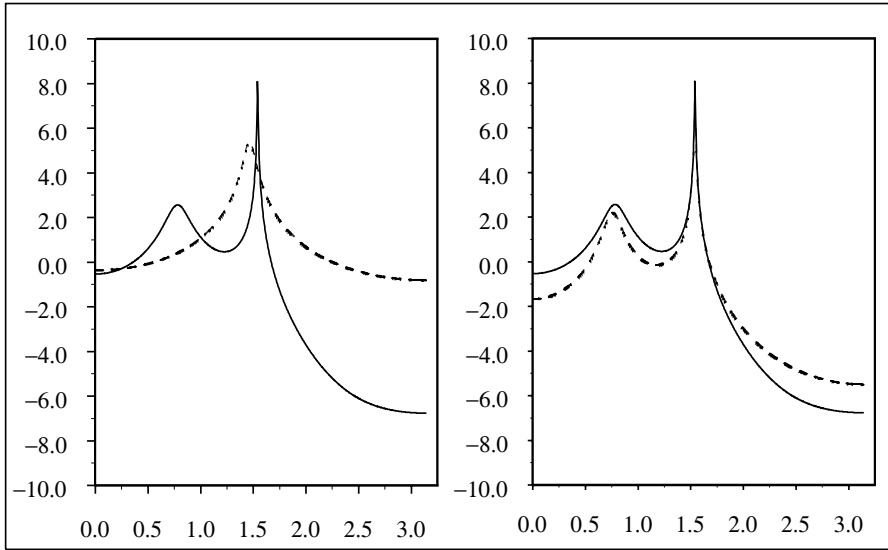


Abb. 12.2 Tatsächliche logarithmierte Spektraldichte des $ARMA(4,2)$ -Modells (12.11) (durchgezogen) und basierend auf einer simulierten Zeitreihe der Länge $n = 300$ gemäß (12.9) parametrische autoregressive Spektraldichteschätzer der Ordnung $p = 2$ (links) und $p = 4$ (rechts) (ebenfalls logarithmierte Versionen, gestrichelt)

$$\hat{f}_{AR}(\omega) = \frac{\hat{\sigma}_e^2}{2\pi} \left| 1 + \sum_{j=1}^q \hat{a}_j e^{-ij\omega} \right|^{-2}, \quad \omega \in [-\pi, \pi], \quad (12.9)$$

und

$$\hat{f}_{ARMA}(\omega) = \frac{\hat{\sigma}_e^2}{2\pi} \frac{\left| 1 + \sum_{j=1}^q \hat{b}_j e^{-ij\omega} \right|^2}{\left| 1 + \sum_{j=1}^q \hat{a}_j e^{-ij\omega} \right|^2}, \quad \omega \in [-\pi, \pi]. \quad (12.10)$$

Grundsätzlich übertragen sich die Eigenschaften der Parameterschätzer (wie Konsistenz und asymptotische Normalität) auf die obigen Spektraldichteschätzer (jedenfalls für festes ω), denn die Ausdrücke (12.8)–(12.10) hängen jeweils stetig von den Parameterschätzern ab, so dass das continuous mapping theorem (Satz A.27) angewendet werden kann. Auf jeden Fall ist also klar, dass bei tatsächlich zugrunde liegendem $ARMA$ -Modell von der Ordnung p, q , der gemäß (12.10) auf Basis von \sqrt{n} -konsistenten Parameterschätzern \hat{a}_j und \hat{b}_k berechnete Spektraldichteschätzer ebenfalls punktweise (d.h. für jedes feste ω) \sqrt{n} -konsistent für die tatsächliche Spektraldichte ist.

Auch für den Fall, dass die vorliegenden Beobachtungen keine Realisierung einer bestimmten $ARMA$ -Zeitreihe darstellen, macht es Sinn aufgrund eines lediglich angepassten $ARMA$ -Modells den zugehörigen parametrischen

Spektraldichteschätzer zu verwenden. Aus Abschnitt 9.2 wissen wir nämlich, dass stetige Spektraldichten beliebig genau durch Spektraldichten von AR -, MA - und damit natürlich erst recht von $ARMA$ -Modellen geeigneter Ordnungen approximiert werden können. Zur Konkretisierung betrachten wir das $ARMA(4, 2)$ -Modell

$$\begin{aligned} X_t - 1.34X_{t-1} + 1.88X_{t-2} - 1.32X_{t-3} + 0.80X_{t-4} \\ = e_t + 0.71e_{t-1} + 0.25e_{t-2}, \quad \forall t \in \mathbb{Z}, \end{aligned} \quad (12.11)$$

mit standard normalverteiltem weißen Rauschen (e_t) . Für eine simulierte Zeitreihe der Länge $n = 300$ sind in den Abbildungen 12.2 und 12.3 die Spektraldichten von angepassten autoregressiven Modellen der Ordnungen 2, 4 und 6 (autoregressive Spektraldichteschätzer) sowie die Spektraldichte eines geschätzten $ARMA(4, 2)$ -Modells zu sehen. Man erkennt, dass bereits ab $AR(4)$ die parametrischen Spektraldichteschätzer eine recht gute Approximation an die tatsächliche Spektraldichte abgeben. Aus Gründen der besseren Sichtbarkeit sind wiederum logarithmierte Versionen dargestellt.

Wir wollen auf parametrische Spektraldichteschätzer nicht weiter im Detail eingehen, sondern unser Augenmerk nunmehr auf die periodogrammbasierten nichtparametrischen Spektraldichteschätzer richten.

12.2 Das Periodogramm

Wir rekapitulieren die Definition des Periodogramms.

Definition 12.2 (*Periodogramm*)

Für gegebene Realisierungen X_1, \dots, X_n einer reellen, stationären Zeitreihe definieren wir für $-\pi \leq \omega \leq \pi$ das Periodogramm $I_n(\omega)$ gemäß

$$I_n(\omega) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-it\omega} \right|^2. \quad (12.12)$$

Für reelle X_1, \dots, X_n gilt stets $I_n(\omega) = I_n(-\omega)$, so dass man in diesem Fall in der Regel das Periodogramm nur auf $[0, \pi]$ betrachtet.

Bemerkung 12.3

Natürlich können wir gemäß (12.12) nicht nur für Realisierungen einer Zeitreihe, sondern für jeden Vektor $(X_1, \dots, X_n) \in \mathbb{R}^n$ (sogar für $(X_1, \dots, X_n) \in \mathbb{C}^n$) ein Periodogramm berechnen. Die Eigenschaften des Periodogramms werden wir aber ausschließlich im Zusammenhang mit Zeitreihen untersuchen. \square

Offensichtlich weichen die Definitionen des zentrierten Periodogramms $I_{n,\bar{X}}$ und des Periodogramms I_n voneinander ab. Wir werden jedoch in dem folgenden Lemma sehen, dass für die sogenannten *Fourier-Frequenzen* $\omega_j := \frac{2\pi j}{n}$ mit $j = -[\frac{n-1}{2}], \dots, [\frac{n}{2}]$, $j \neq 0$, tatsächliche Gleichheit gilt.

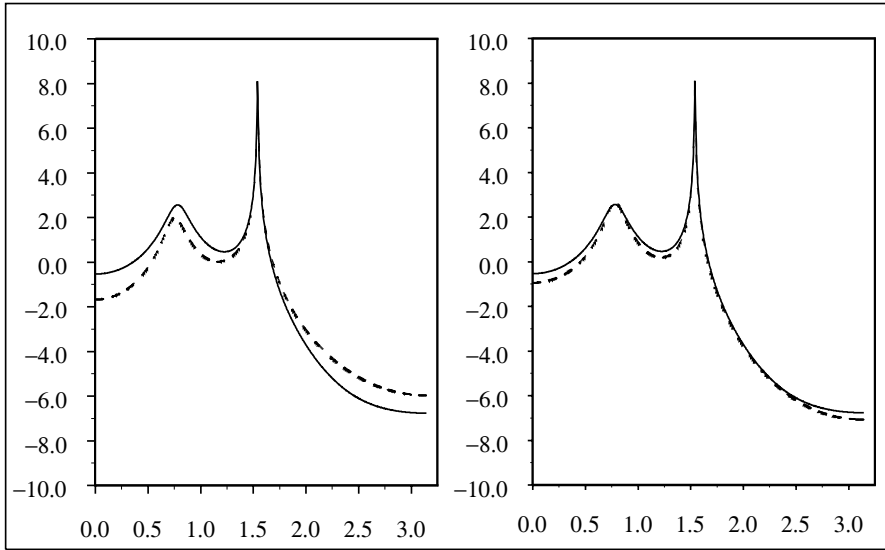


Abb. 12.3 Tatsächliche logarithmierte Spektraldichte des $ARMA(4,2)$ -Modells (12.11) (durchgezogen) und basierend auf einer simulierten Zeitreihe der Länge $n = 300$ gemäß (12.9) der parametrische autoregressive Spektraldichteschätzer der Ordnung $p = 6$ (links) und der Spektraldichteschätzer (12.10) des geschätzten $ARMA$ -Modells (rechts) (ebenfalls logarithmierte Versionen, gestrichelt)

Lemma 12.4

Für reelle X_1, \dots, X_n , $\omega_j = 2\pi j/n$, $j = -[(n-1)/2], \dots, [n/2]$, gilt

$$I_n(\omega_j) = I_{n,\bar{X}}(\omega_j) = I_{n,\mu}(\omega_j), \quad \forall \mu \in \mathbb{R}, \quad (12.13)$$

falls $\omega_j \neq 0$ (d.h. $j \neq 0$).

Beweis: Die Gleichungen (12.13) folgen unmittelbar aus (12.6) und der Tatsache, dass für $\omega_j \neq 0$ stets $\sum_{t=1}^n e^{-it\omega_j} = 0$ gilt. \square

Bezeichnungen:

Für spätere Betrachtungen wollen wir an dieser Stelle die folgenden empirischen Autokovarianzen zusammenstellen:

$$\widehat{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_t - \overline{X}_n)(X_{t+|h|} - \overline{X}_n), \quad |h| < n, \quad (12.14)$$

$$\overline{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} X_t X_{t+|h|}, \quad |h| < n, \quad (12.15)$$

$$\widetilde{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^n X_t X_{t+|h|}, \quad |h| \geq 0. \quad (12.16)$$

Wie bereits erwähnt, können wir grundsätzlich für jeden n -dimensionalen komplexen Vektor ein Periodogramm berechnen. Wir werden im Folgenden sehen, dass das Periodogramm sehr eng mit der Darstellung des Vektors (X_1, \dots, X_n) ($\in \mathbb{C}^n$) in einer anderen als der kanonischen Basis zusammenhängt. Dazu beachte man zunächst, dass folgendes Resultat gilt.

Lemma 12.5

Die n Vektoren $\mathbf{b}_j := n^{-1/2}(e^{i\omega_j}, e^{i2\omega_j}, \dots, e^{in\omega_j})^T$, mit den Fourier-Frequenzen $\omega_j = 2\pi j/n$, $j = -[(n-1)/2], \dots, [n/2]$, bilden eine Orthonormalbasis des \mathbb{C}^n .

Beweis: Siehe Aufgabe 12.4. □

Wegen Lemma 12.5 läßt sich also der Vektor (X_1, \dots, X_n) gemäß

$$(X_1, \dots, X_n)^T = \sum_{j=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} C_j \mathbf{b}_j \quad (12.17)$$

bzw.

$$X_t = \frac{1}{\sqrt{n}} \sum_{j=-[\frac{n-1}{2}]}^{[\frac{n}{2}]} C_j e^{it\omega_j}, \quad t = 1, \dots, n, \quad (12.18)$$

mit $C_j := n^{-1/2} \sum_{t=1}^n X_t e^{-it\omega_j}$ als Linearkombination von Schwingungen mit Frequenzen ω_j , $j = -[(n-1)/2], \dots, [n/2]$, darstellen. Die Koeffizienten C_j , $j = -[(n-1)/2], \dots, [n/2]$, heißen auch *finite Fourier Transformation* von X_1, \dots, X_n .

Das zu den Werten X_1, \dots, X_n gehörende Periodogramm I_n erfüllt nun für $j = -[(n-1)/2], \dots, [n/2]$ die Gleichungen

$$I_n(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-it\omega_j} \right|^2 = \frac{1}{2\pi} |C_j|^2. \quad (12.19)$$

Für reelle Daten X_1, \dots, X_n gilt $C_{-j} = \overline{C_j}$.

Die Darstellung (12.18) erinnert stark an das Modell einer zyklischen Zeitreihe

(vgl. Beispiel 5.9). Dort ergab das Quadrat der L_2 -Norm der stochastischen Amplitude zu einer Frequenz gerade das Gewicht dieser Frequenz im (in diesem Fall diskreten) Spektralmaß an. Hier greift nun die Analogie zum Periodogramm (ausgewertet an einer Fourier-Frequenz ω_j). Der quadrierte Betrag des Koeffizienten C_j/\sqrt{n} des Schwingungsanteils zur Frequenz ω_j in der Darstellung (12.18) stimmt wegen (12.19) gerade mit $(2\pi/n)I_n(\omega_j)$ überein. Stellt nun $X = (X_t : t \in \mathbb{Z})$ eine stationäre Zeitreihe mit Spektraldichte f dar, so liefert eine Diskretisierung von f auf die Fourier-Frequenzen auf der Frequenz $\omega_j = 2\pi j/n$ gerade das Gewicht $2\pi/n \cdot f(\omega_j)$. Die obigen heuristischen Überlegungen legen über die Argumentation in der Einleitung zu diesem Kapitel hinaus nahe, dass $I_n(\omega_j)$, basierend auf Realisierungen X_1, \dots, X_n dieser Zeitreihe, eine plausible Schätzung für $f(\omega_j)$ darstellen könnte.

Es sollte noch erwähnt werden, dass gemäß (12.18) eine bijektive Zuordnung zwischen X_1, \dots, X_n und $C_j, j = -[(n-1)/2], \dots, [n/2]$ gegeben ist. Das stochastische Verhalten von X_1, \dots, X_n kann also vollständig durch die Zufallsvariablen C_j beschrieben werden.

Beim Übergang zum Periodogramm, d.h. im Wesentlichen beim Übergang von C_j zu $|C_j|^2$, geht diese bijektive Zuordnung allerdings verloren. Allein mit Kenntnis des Periodogramms ist somit kein Rückschluß auf die zugrunde liegenden Daten mehr möglich. Allerdings können wir über das Herglotz-Lemma aus $I_{n,\mu}$ die empirischen Autokovarianzen $\hat{\gamma}_{n,\mu}$ der Daten rekonstruieren gemäß $\hat{\gamma}_{n,\mu}(h) = \int_{-\pi}^{\pi} I_{n,\mu}(\omega) e^{ih\omega} d\omega$.

12.3 Eigenschaften des Periodogramms

Im vorangegangenen Abschnitt haben wir argumentiert, dass das Periodogramm ein plausibler Schätzansatz für die Spektraldichte einer Zeitreihe darstellt. Hier sollen nun die Eigenschaften des Periodogramms genau untersucht werden. Es wird sich zeigen, dass das Periodogramm *keine* konsistente Schätzung für die Spektraldichte einer Zeitreihe darstellt. Aber zunächst gilt

Satz 12.6 (Erwartungstreue des Periodogramms)

X_1, \dots, X_n seien Elemente einer stationären und reellwertigen Zeitreihe mit absolut summierbarer Autokovarianzfunktion γ und deshalb existierender Spektraldichte f . Konvergenzen im Folgenden sind für $n \rightarrow \infty$.

Ist $EX_t = 0$, so gilt für das Periodogramm

$$EI_n(\omega) \rightarrow f(\omega), \quad \forall \omega \in [0, \pi]. \quad (12.20)$$

Diese Konvergenz ist gleichmäßig in $\omega \in [0, \pi]$. Falls $EX_t = \mu$ (evtl. $\mu \neq 0$) erhalten wir

$$EI_n(0) - \frac{n}{2\pi} \mu^2 \rightarrow f(0) \quad (12.21)$$

und mit den Fourier-Frequenzen $\omega_j = 2\pi j/n$

$$\sup_{j=1, \dots, [n/2]} |EI_n(\omega_j) - f(\omega_j)| \rightarrow 0. \quad (12.22)$$

Beweis: Wir betrachten nur den Fall $EX_t = \mu$. Zunächst gilt mit Hilfe von Gleichung (A.11)

$$\begin{aligned} EI_n(0) - \frac{n}{2\pi}\mu^2 &= \frac{1}{2\pi n} \left(E \left| \sum_{t=1}^n X_t \right|^2 - n^2\mu^2 \right) \\ &= \frac{1}{2\pi n} \sum_{s,t=1}^n E(X_s - \mu)(X_t - \mu) = \frac{1}{2\pi n} \sum_{s,t=1}^n \gamma(s-t) \\ &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n} \right) \gamma(h). \end{aligned}$$

Da $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, ergibt der Satz von Lebesgue $\sum_{h=-(n-1)}^{n-1} \frac{|h|}{n} \gamma(h) \rightarrow 0$, also mit obiger Rechnung $(EI_n(0) - \frac{n}{2\pi}\mu^2) \rightarrow \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) = f(0)$, und damit (12.21).

Zum Nachweis von (12.22) verwenden wir die Darstellung (12.6) des Periodogramms mit $\mu = EX_t$. Es ergibt sich für $\omega_j \neq 0$

$$EI_n(\omega_j) = \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n} \right) \gamma(h) e^{-ih\omega_j}.$$

Da sowohl $\sum_{|h| \geq n} |\gamma(h)|$ wie $\sum_{|h| < n} \frac{|h|}{n} |\gamma(h)|$ für $n \rightarrow \infty$ gegen Null konvergieren, folgt die behauptete gleichmäßige Konvergenz in (12.22). (12.20) kann vollständig analog nachgerechnet werden. Wegen $\mu = 0$ ist die Konvergenz hier gleichmäßig über ganz $[0, \pi]$. \square

Die Darstellung $I_n(0) = \frac{1}{2\pi} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t \right|^2$ zeigt wegen der typischerweise geltenden asymptotischen Normalität von $n^{-1/2} \sum_{t=1}^n X_t$ für zentrierte Zeitreihen (vgl. Theorem 10.12), dass zumindest $I_n(0)$ – selbst für zentrierte stationäre Zeitreihen – keine konsistente Schätzung für die Spektraldichte f an der Stelle Null darstellt. $I_n(0)$ konvergiert für wachsendes n vielmehr gegen das Quadrat einer $\mathcal{N}(0, f(0))$ -Verteilung, hat damit zwar asymptotisch den Erwartungswert $f(0)$, ist aber nicht konsistent. Wir werden im Folgenden sehen, dass analoge Aussagen für alle Frequenzen ω des Periodogramms gelten.

Zunächst erhalten wir die Nicht-Konsistenz des Periodogramms für ein stochastisch unabhängiges weißes Rauschen.

Satz 12.7 (Nicht-Konsistenz des Periodogramms)

$e = (e_t : t \in \mathbb{Z})$ sei ein aus i.i.d. Zufallsvariablen bestehendes reelles weißes

Rauschen mit $Ee_1^2 = \sigma_e^2$ und $Ee_1^4 =: \eta\sigma_e^4 < \infty$. I_n bezeichne das zugehörige Periodogramm basierend auf e_1, \dots, e_n .

(i) Für $\omega, \omega_1, \omega_2 \in [0, \pi]$ mit $\omega_1 \neq \omega_2$ gilt

$$\text{Cov}(I_n(\omega_1), I_n(\omega_2)) = \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} + \mathcal{O}(n^{-2}) \quad (12.23)$$

und

$$\text{Var} I_n(\omega) = \begin{cases} \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} + \frac{\sigma_e^4}{2\pi^2} & , \omega = 0, \pi, \\ \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} + \frac{\sigma_e^4}{4\pi^2} + \mathcal{O}(n^{-2}) & , 0 < \omega < \pi. \end{cases} \quad (12.24)$$

(ii) Speziell für Fourier-Frequenzen $\omega_j = 2\pi j/n$ und $\omega_k = 2\pi k/n$ mit $j, k = 0, \dots, [n/2], j \neq k$, ergibt sich

$$\text{Cov}(I_n(\omega_j), I_n(\omega_k)) = \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} \quad (12.25)$$

und

$$\text{Var} I_n(\omega_j) = \begin{cases} \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} + \frac{\sigma_e^4}{2\pi^2} & , \omega_j = 0, \pi, \\ \frac{(\eta-3)\sigma_e^4}{4\pi^2 n} + \frac{\sigma_e^4}{4\pi^2} & , 0 < \omega_j < \pi. \end{cases} \quad (12.26)$$

Bemerkung 12.8

Zunächst erkennt man aus (12.24) bzw. (12.26), dass die Varianz des Periodogramms für $n \rightarrow \infty$ *nicht* gegen Null konvergiert. Demnach stellt das Periodogramm zwar eine asymptotisch erwartungstreue aber *keine* konsistente Schätzung der Spektraldichte dar! Vgl. zur Veranschaulichung die Abbildungen 12.1 und 12.4.

Ferner sieht man aus (12.23) bzw. (12.25), dass die Werte des Periodogramms für verschiedene Frequenzen ω_1 und ω_2 asymptotisch unkorreliert sind. Für ein normalverteiltes Rauschen (e_t) gilt wegen $\eta = 3$ für Fourier-Frequenzen $\omega_j \neq \omega_k$ sogar die finite Unkorreliertheit. Es gilt in diesem Fall sogar die Unabhängigkeit des Periodogramms an verschiedenen Fourier-Frequenzen, obwohl ihre Verteilungen nicht normal, sondern vom χ^2 -Typ sind, vgl. Aufgabe 12.3 und Satz 12.11. \square

Beweis: (von Satz 12.7)

Zunächst gilt für alle $\omega \in [0, \pi]$

$$EI_n(\omega) = \frac{1}{2\pi n} E \left| \sum_{t=1}^n e_t e^{-it\omega} \right|^2 = \frac{1}{2\pi n} \sum_{t=1}^n Ee_t^2 = \frac{\sigma_e^2}{2\pi}.$$

Für $\omega_1, \omega_2 \in [0, \pi]$ erhalten wir weiter

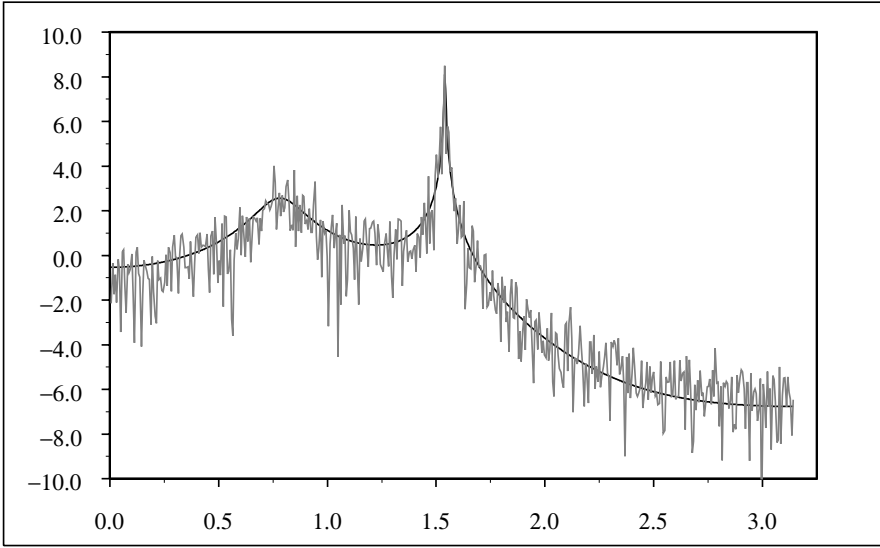


Abb. 12.4 Tatsächliche logarithmierte Spektraldichte des $ARMA(4,2)$ -Modells (12.11) (gestrichelt) und das Periodogramm basierend auf einer simulierten Zeitreihe der Länge $n = 1000$

$$\begin{aligned}
 EI_n(\omega_1)I_n(\omega_2) &= \frac{1}{4\pi^2 n^2} E \left| \sum_{t=1}^n e_t e^{-it\omega_1} \right|^2 \left| \sum_{v=1}^n e_v e^{-iv\omega_2} \right|^2 \\
 &= \frac{1}{4\pi^2 n^2} \sum_{s,t=1}^n \sum_{u,v=1}^n E(e_s e_t e_u e_v) e^{-i(t-s)\omega_1} e^{-i(u-v)\omega_2}.
 \end{aligned}$$

Beachtet man, dass $E(e_s e_t e_u e_v)$ nur dann nicht Null ist, wenn alle Indizes gleich sind oder wenn zwei verschiedene Indexpaare auftreten, so erhalten wir

$$\begin{aligned}
 EI_n(\omega_1)I_n(\omega_2) &= \frac{(\eta - 3)\sigma_e^4}{4\pi^2 n} + \frac{\sigma_e^4}{4\pi^2} + \frac{\sigma_e^4}{4\pi^2 n^2} \left| \sum_{t=1}^n e^{-it(\omega_1 - \omega_2)} \right|^2 \\
 &\quad + \frac{\sigma_e^4}{4\pi^2 n^2} \left| \sum_{t=1}^n e^{-it(\omega_1 + \omega_2)} \right|^2. \tag{12.27}
 \end{aligned}$$

Für $\omega_1 \neq \omega_2$ folgt wegen $\left| \sum_{t=1}^n e^{-it(\omega_1 \pm \omega_2)} \right|^2 \leq 1/\sin^2((\omega_1 \pm \omega_2)/2)$ unter Beachtung von (A.6) schließlich (12.23). Auf dem gleichen Wege verifiziert man (12.24). Für Fourier-Frequenzen ω_j und ω_k gilt schließlich $\sum_{t=1}^n e^{-it(\omega_j \pm \omega_k)} = 0$, falls $\omega_j \neq \omega_k$, und $\sum_{t=1}^n e^{-it\omega_j} = 0$, falls $\omega_j \neq 0$, bzw. $= n$, falls $\omega_j = 0$, sodass sich auch (12.25) und (12.26) ergeben.

□

Mit Hilfe des Zentralen Grenzwertsatzes von Lindeberg können wir die asymptotische Verteilung des Periodogramms ausrechnen. Wie bereits eingangs an-

gedeutet, erhalten wir im Grenzfall $n \rightarrow \infty$ Exponentialverteilungen. Das genaue Resultat lautet

Satz 12.9 (*Asymptotische Verteilung des Periodogramms*)

Unter genau den gleichen Voraussetzungen wie in Satz 12.7 erhalten wir für $0 = \omega_1 < \omega_2 < \dots < \omega_k = \pi$ (wobei die ω_j hier nicht notwendig Fourier-Frequenzen sind) und $n \rightarrow \infty$

$$(I_n(\omega_1), \dots, I_n(\omega_k)) \xrightarrow{\mathcal{D}} \frac{\sigma_e^2}{2\pi} (Z_1, \dots, Z_k), \quad (12.28)$$

wobei die Zufallsvariablen Z_1, \dots, Z_k stochastisch unabhängig sind mit $E Z_j = 1$ für alle $j = 1, \dots, k$. Ferner sind Z_2, \dots, Z_{k-1} exponentialverteilt und Z_1, Z_k besitzen eine χ_1^2 -Verteilung.

Beweis: Da

$$\begin{aligned} I_n(\omega) &= \frac{1}{2\pi n} \left| \sum_{t=1}^n e_t e^{-it\omega} \right|^2 \\ &= \frac{1}{2\pi} \left[\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \cos(t\omega) \right)^2 + \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \sin(t\omega) \right)^2 \right], \end{aligned}$$

reicht es zu zeigen, dass der Vektor

$Y_n := (A_n(\omega_1), A_n(\omega_2), B_n(\omega_2), \dots, A_n(\omega_{k-1}), B_n(\omega_{k-1}), A_n(\omega_k))^T$ mit

$$A_n(\omega) = \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \cos(t\omega) \quad (12.29)$$

und

$$B_n(\omega) = \frac{1}{\sqrt{n}} \sum_{t=1}^n e_t \sin(t\omega) \quad (12.30)$$

eine asymptotische $\mathcal{N}(\mathbf{0}, \sigma_e^2/2 \cdot \text{diag}(2, 1, \dots, 1, 2))$ -Verteilung besitzt. Hierzu wendet man die Cramér–Wold–Technik und den Zentralen Grenzwertsatz von Lindeberg an. Dabei ist zu beachten, dass gelten

$$\text{Var} A_n(\omega_1) = \text{Var} A_n(\omega_k) = \sigma_e^2 \quad \text{und} \quad (12.31)$$

$$\lim_{n \rightarrow \infty} \text{Var} A_n(\omega_j) = \lim_{n \rightarrow \infty} \text{Var} B_n(\omega_j) = \sigma_e^2/2, \quad 1 < j < k, \quad (12.32)$$

vgl. Aufgabe 12.5.

Außerdem gilt, dass alle Kovarianzen für $n \rightarrow \infty$ verschwinden. Dies sei exemplarisch für die Kovarianz von $A_n(\omega)$ und $B_n(\omega)$ gezeigt.

$$\begin{aligned}
& |Cov(A_n(\omega), B_n(\omega))| = |EA_n(\omega)B_n(\omega)| \\
&= \frac{\sigma_e^2}{n} \left| \sum_{t=1}^n \sin(t\omega) \cos(t\omega) \right| = \frac{\sigma_e^2}{4n} \left| \sum_{t=1}^n (e^{it\omega} - e^{-it\omega})(e^{it\omega} + e^{-it\omega}) \right| \\
&= \mathcal{O}(n^{-1}),
\end{aligned}$$

wobei die letzte Gleichung mit Hilfe von Formel (A.6) für alle $0 \leq \omega \leq \pi$ folgt. Damit ist die asymptotische Kovarianzmatrix von Y_n tatsächlich gleich $(\sigma_e^2/2) \cdot \text{diag}(2, 1, \dots, 1, 2)$. Anschließend bleibt die Lindeberg-Bedingung zu zeigen. Auch diese soll wiederum nur exemplarisch für $A_n(\omega)$ geschehen. Für jedes $\varepsilon > 0$ gilt

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n E e_t^2 \cos^2(t\omega) 1 \{ |e_t \cos(t\omega)| > \varepsilon \sqrt{n} \sigma_e \} \\
& \leq E e_1^2 1 \{ |e_1| > \varepsilon \sqrt{n} \sigma_e \} \rightarrow 0 \text{ für } n \rightarrow \infty.
\end{aligned}$$

Damit sind alle Aussagen des Satzes 12.9 bewiesen. \square

Die Übertragung der Sätze 12.7 und 12.9 auf allgemeine Klassen von Zeitreihen gelingt mit der folgenden Aussage, die in gewisser Weise die Transformation des Periodogramms unter linearen Filtern angibt. Es sei daran erinnert, dass für die Spektraldichte f_X eines zweiseitigen $MA(\infty)$ -Prozesses

$$X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (12.33)$$

mit einem weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ gilt

$$f_X(\omega) = \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2 f_e(\omega), \quad \forall \omega \in (-\pi, \pi].$$

f_e bezeichnet dabei die Spektraldichte des weißen Rauschens. Für das Periodogramm gilt

Satz 12.10

$X = (X_t : t \in \mathbb{Z})$ sei ein linearer Prozess (12.33) mit $\sum_j \sqrt{|j|} |b_j| < \infty$ und einem i.i.d. weißen Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $E e_1^4 < \infty$. I_n^X bezeichne das Periodogramm basierend auf X_1, \dots, X_n und I_n^e das Periodogramm basierend auf e_1, \dots, e_n . Dann gilt

$$E \left[I_n^X(\omega) - \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2 I_n^e(\omega) \right]^2 = \mathcal{O}(n^{-1}). \quad (12.34)$$

Insbesondere folgt hieraus natürlich

$$I_n^X(\omega) = \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2 I_n^e(\omega) + \mathcal{O}_P(n^{-1/2}). \quad (12.35)$$

Bevor wir diese Aussage beweisen, wollen wir mit ihrer Hilfe die Sätze 12.7 und 12.9 auf lineare Prozesse übertragen.

Satz 12.11 (*Asymptotisches Verhalten des Periodogramms*)

$X = (X_t : t \in \mathbb{Z})$ sei ein linearer Prozeß mit Darstellung (12.33) für ein i.i.d. weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t = 0$, $Ee_t^2 = \sigma_e^2$, $Ee_1^4 < \infty$, sowie $\sum_{j=-\infty}^{\infty} \sqrt{|j|} |b_j| < \infty$. $f(\omega) = (\sigma_e^2/2\pi) \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2$, $\omega \in (-\pi, \pi]$, bezeichne die Spektraldichte und I_n^X das Periodogramm basierend auf Realisierungen X_1, \dots, X_n .

(i) Es gilt für $\omega, \omega_1, \omega_2 \in [0, \pi]$ mit $\omega_1 \neq \omega_2$ einerseits

$$\text{Cov}(I_n^X(\omega_1), I_n^X(\omega_2)) = \mathcal{O}(n^{-1/2}) \quad (12.36)$$

und andererseits

$$\text{Var} I_n^X(\omega) = \begin{cases} 2f^2(\omega) + \mathcal{O}(n^{-1/2}), & \omega = 0, \pi, \\ f^2(\omega) + \mathcal{O}(n^{-1/2}), & 0 < \omega < \pi. \end{cases} \quad (12.37)$$

(ii) Für $0 = \omega_1 < \omega_2 < \dots < \omega_k = \pi$ (wobei die ω_j keine Fourier-Frequenzen sein müssen) und $n \rightarrow \infty$ gilt

$$(I_n^X(\omega_1), \dots, I_n^X(\omega_k)) \xrightarrow{\mathcal{D}} (f(\omega_1)Z_1, \dots, f(\omega_k)Z_k), \quad (12.38)$$

wobei die Zufallsvariablen Z_1, \dots, Z_k stochastisch unabhängig sind mit $EZ_j = 1$, $j = 1, \dots, k$, und Z_2, \dots, Z_{k-1} eine Exponentialverteilung mit Parameter 1 besitzen, während Z_1 und Z_k jeweils χ_1^2 -verteilt sind.

Beweis: Zunächst folgt (12.38) direkt aus Satz 12.9 und Satz 12.10, 12.35 mit Hilfe des Lemmas von Slutsky. (12.36) sieht man wie folgt ein. Setze dabei zur Abkürzung (beachte (12.34))

$$R_n(\omega) := I_n^X(\omega) - \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2 I_n^e(\omega).$$

Da

$$\begin{aligned} & \text{Cov}(I_n^X(\omega_1), I_n^X(\omega_2)) \\ &= \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega_1} \right|^2 \left| \sum_{k=-\infty}^{\infty} b_k e^{-ik\omega_2} \right|^2 \text{Cov}(I_n^e(\omega_1), I_n^e(\omega_2)) \\ &+ \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega_1} \right|^2 \text{Cov}(I_n^e(\omega_1), R_n(\omega_2)) \\ &+ \left| \sum_{k=-\infty}^{\infty} b_k e^{-ik\omega_2} \right|^2 \text{Cov}(R_n(\omega_1), I_n^e(\omega_2)) \\ &+ \text{Cov}(R_n(\omega_1), R_n(\omega_2)), \end{aligned}$$

folgt die Behauptung aus $\text{Var} R_n(\omega) \leq ER_n^2(\omega) = \mathcal{O}(n^{-1})$, vgl. (12.34), denn $\text{Var} I_n^e(\omega)$ ist beschränkt (vgl. Satz 12.7).

Auf die gleiche Weise kann (12.37) verifiziert werden (vgl. Aufgabe 12.6). \square

Bemerkung 12.12

Statt (12.36) gilt unter den gleichen Voraussetzungen wie in Satz 12.11 sogar

$$\text{Cov}(I_n^X(\omega_1), I_n^X(\omega_2)) = \mathcal{O}(n^{-1}) \quad (12.39)$$

für $\omega_1 \neq \omega_2 \in [0, \pi]$. Siehe Brockwell/Davis (1991), Theorem 10.3.2. \square

Es bleibt noch der Beweis von Satz 12.10 nachzutragen.

Beweis: (Zu Satz 12.10) Zunächst gilt $I_n^X(\omega) = Z_n^X(\omega) \bar{Z}_n^X(\omega)$ mit der Definition

$$\begin{aligned} Z_n^X(\omega) &:= \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{-i\omega t} \\ &= \frac{1}{\sqrt{2\pi n}} \sum_{\nu=-\infty}^{\infty} b_\nu e^{-i\omega\nu} \sum_{t=1-\nu}^{n-\nu} e^{-i\omega t} \\ &= \left(\sum_{\nu=-\infty}^{\infty} b_\nu e^{-i\omega\nu} \right) \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n e_t e^{-i\omega t} \\ &\quad + \frac{1}{\sqrt{2\pi n}} \left(\sum_{t=1-\nu}^{n-\nu} e_t e^{-i\omega t} - \sum_{t=1}^n e_t e^{-i\omega t} \right) \\ &=: B(e^{-i\omega}) Z_n^e(\omega) + D_n(\omega), \end{aligned} \quad (12.40)$$

mit $B(z) := \sum_{\nu} b_\nu z^\nu$ und Z_n^e analog zu Z_n^X definiert, sowie dem Restterm D_n . Unter Verwendung von $I_n^X(\omega) = Z_n^X(\omega) \bar{Z}_n^X(\omega)$ ergibt sich somit

$$\begin{aligned} \Delta_n &:= I_n^X(\omega) - |B(e^{-i\omega})|^2 I_n^e(\omega) \\ &= |B(e^{-i\omega}) Z_n^e(\omega) \bar{D}_n(\omega) + \bar{B}(e^{-i\omega}) \bar{Z}_n^e(\omega) D_n(\omega) + |D_n(\omega)|^2| \\ &\leq 2|B(e^{-i\omega})| |Z_n^e(\omega)| |D_n(\omega)| + |D_n(\omega)|^2. \end{aligned} \quad (12.41)$$

Mit $|B(e^{-i\omega})| \leq \sum_{\nu} |b_\nu| =: c_0 < \infty$ ergibt sich zunächst

$$\Delta_n^2 \leq 8c_0^2 I_n^e(\omega) |D_n(\omega)|^2 + 2|D_n(\omega)|^4, \quad (12.42)$$

und daraus mit $E|Z_n^e(\omega)|^4 = E(I_n^e)^2 \leq c_1 < \infty$ (unabhängig von ω , siehe (12.30)) unter Verwendung der Cauchy-Schwarz-Ungleichung

$$E\Delta_n^2 \leq 8c_0^2 \sqrt{c_1} \left(E|D_n(\omega)|^4 \right)^{1/2} + 2E|D_n(\omega)|^4. \quad (12.43)$$

Nun ist mit $D_{n\nu}(\omega) := \sum_{t=1-\nu}^{n-\nu} e_t e^{-i\omega t} - \sum_{t=1}^n e_t e^{-i\omega t}$

$$\begin{aligned} E|D_n(\omega)|^4 &= \frac{1}{(2\pi n)^2} E \left| \sum_{\nu=-\infty}^{\infty} b_\nu e^{-i\omega\nu} D_{n\nu}(\omega) \right|^4 \\ &\leq \frac{1}{(2\pi n)^2} E \left(\sum_{\nu=-\infty}^{\infty} |b_\nu| |D_{n\nu}(\omega)| \right)^4. \end{aligned} \quad (12.44)$$

Die grobe Abschätzung

$$E \left| \sum_t Y_t \right|^4 \leq \sum_{s,t,u,v} (EY_s^4 EY_t^4 EY_u^4 EY_v^4)^{1/4} = \left(\sum_t (EY_t^4)^{1/4} \right)^4,$$

die man durch zweifache Anwendung der Cauchy-Schwarz-Ungleichung erhält, führt nun bereits zum Ziel. Beachtet man

$$E|D_{n\nu}(\omega)|^4 \leq 2|\nu| Ee_1^4 + 12\nu^2 \sigma_e^4 \leq c\nu^2, \quad (12.45)$$

mit $c := 2Ee_1^4 + 12\sigma_e^4$, siehe Aufgabe 12.18, so erhält man schließlich

$$E|D_n(\omega)|^4 \leq \frac{c}{(2\pi n)^2} \left(\sum_\nu |b_\nu| |\nu|^{1/2} \right)^4.$$

Einsetzen in (12.43) ergibt mit einer von ω unabhängigen Konstanten K

$$E\Delta_n^2 \leq K \frac{1}{n},$$

also (12.34). □

Bemerkung 12.13

In Ergänzung zu Satz 12.10 sei erwähnt, dass die Aussage

$$I_n^X(\omega) = \left| \sum_{j=-\infty}^{\infty} b_j e^{-ij\omega} \right|^2 I_n^e(\omega) + o_P(1)$$

auch ohne die Annahme $\sum_j \sqrt{|j|} |b_j| < \infty$ nachgewiesen werden kann. Siehe Brockwell/Davis (1991), Theorem 10.3.1. □

12.4 Lag-Window-Schätzer der Spektraldichte

Im vorangegangenen Abschnitt haben wir eine ganze Reihe von zentralen asymptotischen Aussagen für das Periodogramm hergeleitet. Dabei hat sich gezeigt, dass das Periodogramm *kein* konsistenter Schätzer für die Spektraldichte (etwa eines linearen Prozesses) ist, da die Varianz für gegen unendlich

wachsenden Stichprobenumfang *nicht* gegen Null konvergiert.

Jetzt sollen konsistente Schätzer für die Spektraldichte einer stationären Zeitreihe vorgestellt werden. Die grundlegende Idee ist eine *Glättung* des Periodogramms, die auf vielfältige Weise erreicht werden kann. Z.B. führen bekanntlich sogenannte Moving Average Filter

$$\sum_{|k| \leq m} \frac{1}{2m+1} I_n(\omega_{j+k})$$

(oder auch mit allgemeineren Gewichten $\sum_{|k| \leq m} w_{k,n} I_n(\omega_{j+k})$) zu einer Glättung des Periodogramms an der Fourier-Frequenz $\omega_j = 2\pi j/n$.

Ganz ähnlich könnte man mit einer mehr oder weniger um eine bestimmte Frequenz ω_0 konzentrierten Funktion $W : [-\pi, \pi] \rightarrow \mathbb{R}$ über

$$\int_{-\pi}^{\pi} W(\omega) I_n(\omega) d\omega$$

zu einer Glättung von I_n an der Stelle ω_0 gelangen. Wir werden hierauf in Abschnitt 12.7 zurückkommen.

Zunächst wollen wir jedoch einen etwas anderen Weg einschlagen und aufbauend auf der Kovarianzform des Periodogramms, siehe (12.7), zu einer Glättung gelangen. Konkret werden wir für $M(n)$, mit $M(n) < n$ für alle $n \in \mathbb{N}$, den folgenden Spektraldichteschätzer \hat{f}_n gemäß

$$\hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{|h| \leq M(n)} k\left(\frac{h}{M(n)}\right) \bar{\gamma}_n(h) e^{-ih\omega}, \quad 0 \leq \omega \leq \pi, \quad (12.46)$$

mit $\bar{\gamma}_n(h) = 1/n \sum_{t=1}^{n-|h|} X_t X_{t+|h|}$ (siehe (12.15)) und einer Gewichtsfunktion $k : [-1, 1] \rightarrow \mathbb{R}$ betrachten. \hat{f}_n heißt *Lag-Window-Schätzer* und ist in der Tat eine Glättung des Periodogramms, wie wir später sehen werden (zur Veranschaulichung vgl. die Abbildungen 12.5 und 12.6). Die Gewichtsfunktion k wichtet in der Inversionsformel die geschätzten Autokovarianzen in bestimmter Weise. Sie wird mit den folgenden Eigenschaften gewählt

$$\begin{aligned} k : [-1, +1] &\rightarrow \mathbb{R} \text{ ist symmetrisch und stetig sowie} \\ k(0) &= 1 \quad \text{und} \quad |k(u)| \leq 1 \text{ für alle } |u| \leq 1. \end{aligned} \quad (12.47)$$

k heißt auch *Lag Window*. Einige in der Literatur vorgeschlagene Beispiele für k sind in Tabelle 12.1 angegeben.

Bei der Wahl des Abschneidepunktes $M(n)$, auch *Lag Number* oder *Truncation Point* genannt, wird darauf zu achten sein, dass dieser für $n \rightarrow \infty$ hinreichend schnell gegen unendlich konvergiert, um die asymptotische Erwartungstreue auch für \hat{f}_n sicher zu stellen. Man beachte dabei, dass \hat{f}_n mit $M(n) = n - 1$ exakt dem Periodogramm entspricht. Andererseits darf $M(n)$ nicht zu schnell

gegen unendlich konvergieren, damit wir im Gegensatz zum Periodogramm für \hat{f}_n die Konsistenz erreichen.

Genau wie bei der Untersuchung der asymptotischen Eigenschaften des Periodogramms werden wir auch in diesem Abschnitt wieder einen linearen Prozess der Form (12.33) mit i.i.d. weißem Rauschen zugrunde legen. Zunächst erhalten wir das folgende Resultat.

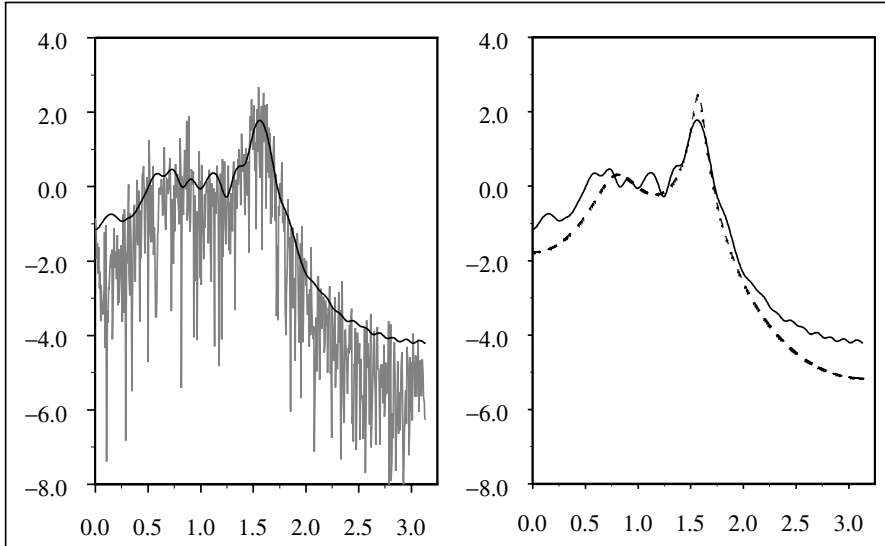


Abb. 12.5 Lag-Window-Schätzer $\hat{f}(\omega) = \frac{1}{2\pi} \sum_{h=-M}^M (1 - \frac{|h|}{M}) \hat{\gamma}(h) e^{-ih\omega}$ mit $M = 50$ für ein simuliertes AR(4)-Modell $X_t - 1.13X_{t-2} + 1.54X_{t-2} - 1.02X_{t-3} + 0.58X_{t-4} = e_t$ der Länge $n = 1000$ mit standard normalverteiltem weißem Rauschen. Links: $\ln \hat{f}$ und logarithmiertes Periodogramm. Rechts: $\ln \hat{f}$ und logarithmierte Spektraldichte (gestrichelt)

Satz 12.14 (*Asymptot. Varianz und Kovarianz des Lag-Window-Schätzers*)
 $X = (X_t : t \in \mathbb{Z})$ sei ein linearer Prozess mit Darstellung (12.33) für ein i.i.d. weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t^2 = \sigma_e^2$, $Ee_t^4 < \infty$ und absolut summierbaren Koeffizienten $\{b_j : j \in \mathbb{Z}\}$. Dann gilt für den Spektraldichteschätzer \hat{f}_n gemäß (12.46) mit Lag Window k mit (12.47) und Truncation Point $M(n) < n$ derart, dass $M(n) \rightarrow \infty$ und $M(n)/n \rightarrow 0$ für $n \rightarrow \infty$:
Für alle $\omega \in [0, \pi]$

$$\lim_{n \rightarrow \infty} E\hat{f}_n(\omega) = f(\omega), \quad (12.48)$$

sowie

$$\lim_{n \rightarrow \infty} \frac{n}{M(n)} \text{Var} \hat{f}_n(\omega) = \begin{cases} 2f^2(\omega) \int_{-1}^{+1} k^2(u) du, & \omega = 0, \pi, \\ f^2(\omega) \int_{-1}^{+1} k^2(u) du, & 0 < \omega < \pi, \end{cases} \quad (12.49)$$

und für alle $\omega_1, \omega_2 \in [0, \pi]$ mit $\omega_1 \neq \omega_2$

$$\text{Cov} \left(\hat{f}_n(\omega_1), \hat{f}_n(\omega_2) \right) = \mathcal{O} \left(\frac{M(n)}{n} \right). \quad (12.50)$$

Dabei bezeichnet f die unter den Voraussetzungen des Satzes existierende Spektraldichte des linearen Prozesses.

Tabelle 12.1 Verschiedene Lag Windows

Name des Lag Window	$k(u)$ für $ u \leq 1$	$\int_{-1}^{+1} k^2(u) du$
Bartlett	1	2
Modified Bartlett	$1 - u $	2/3
Truncated Daniell	$\sin(\pi u)/\pi u$	0.9028
Blackman–Tukey	$1 - 2a + 2a \cos(\pi u)$	$2 - 8a + 12a^2$
$a = 0.25$ (Hanning)	$0.5 + 0.5 \cos(\pi u)$	3/4
$a = 0.23$ (Hamming)	$0.54 + 0.46 \cos(\pi u)$	0.7948
Parzen (1)	$1 - u^2$	16/15
Parzen (2)	$\begin{cases} 1 - 6u^2 + 6 u ^3, & u \leq \frac{1}{2} \\ 2(1 - u)^3, & \frac{1}{2} \leq u \leq 1 \end{cases}$	151/280

Beweis: Wir schreiben zur Abkürzung $M(n) = M$.

Zunächst zur asymptotischen Erwartungstreue (12.48). Es gilt

$$E \hat{f}_n(\omega) = \frac{1}{2\pi} \sum_{|h| \leq M} k \left(\frac{h}{M} \right) \left(1 - \frac{|h|}{n} \right) \gamma(|h|) e^{-ih\omega}.$$

Da aus $\sum_j |b_j| < \infty$ wegen $\gamma(h) = \sigma_e^2 \sum_j b_j b_{j+h}$ auch $\sum_h |\gamma(h)| < \infty$ folgt, impliziert der Satz von Lebesgue einerseits

$$\sum_{h=-\infty}^{\infty} 1\{|h| \leq M\} k \left(\frac{h}{M} \right) \frac{|h|}{n} |\gamma(|h|)| = o(1),$$

und andererseits (beachte $M \rightarrow \infty$ für $n \rightarrow \infty$),

$$\sum_{h=-\infty}^{\infty} 1\{|h| \leq M\} k\left(\frac{h}{M}\right) \gamma(|h|) e^{-ih\omega} \rightarrow \sum_{h=-\infty}^{\infty} \gamma(|h|) e^{-ih\omega}.$$

Mit der Darstellung der Spektraldichte f aus der Inversionsformel folgt (12.48).

Zur Berechnung der asymptotischen Varianz nehmen wir vereinfachend für ein festes $q \in \mathbb{N}$ an, dass $b_j = 0$ für $|j| > q$ gilt, d.h. wir unterstellen einen linearen Prozess endlicher Ordnung. Natürlich kann man jeden linearen Prozess durch einfaches Abschneiden der Summation in (12.33) durch einen Prozess endlicher Ordnung approximieren. Wir ersparen uns den technisch aufwändigen Nachweis, dass die Varianz von \hat{f}_n durch ein solches Abschneiden gleichmäßig in n beliebig gut approximiert werden kann.

Durch Einsetzen von $X_t = \sum_{|\nu| \leq q} b_\nu e_{t-\nu}$ erhalten wir

$$\hat{f}_n(\omega) = \frac{1}{2\pi n} \sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu \sum_{|r| \leq M} k\left(\frac{r}{M}\right) \sum_{t=1}^{n-|r|} e_{t-\nu} e_{t+|r|-\mu} e^{-ir\omega}.$$

Es ist leicht zu sehen, dass für den Term mit $r = 0$ gilt

$$\frac{1}{M \cdot n} \text{Var} \left(\sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu \sum_{t=1}^n e_{t-\nu} e_{t-\mu} \right) = o(1),$$

so dass die Betrachtung von

$$\tilde{h}(\omega) := \frac{1}{\pi n} \sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu \sum_{r=1}^M k\left(\frac{r}{M}\right) \sum_{t=1}^{n-r} e_{t-\nu} e_{t+r-\mu} \cos(r\omega) \quad (12.51)$$

hinreicht. Die Varianz dieses Ausdrucks ergibt sich bis auf einen Faktor $1/(\pi n)^2$ zu

$$\begin{aligned} & \sum_{\nu, \mu, \nu', \mu'=-q}^{+q} b_\nu b_\mu b_{\nu'} b_{\mu'} \sum_{r, r'=1}^M k\left(\frac{r}{M}\right) k\left(\frac{r'}{M}\right) \cdot \cos(r\omega) \cos(r'\omega) \\ & \times \sum_{t=1}^{n-r} \sum_{t'=1}^{n-r'} E(e_{t-\nu} e_{t+r-\mu} - \sigma_e^2 1_{\{\mu=\nu+r\}}) (e_{t'-\nu'} e_{t'+r'-\mu'} - \sigma_e^2 1_{\{\mu'=\nu'+r'\}}) \\ & = \sum_{\nu, \mu, \nu', \mu'=-q}^{+q} b_\nu b_\mu b_{\nu'} b_{\mu'} \sum_{r, r'=1}^M k\left(\frac{r}{M}\right) k\left(\frac{r'}{M}\right) \cdot \cos(r\omega) \cos(r'\omega) \times \\ & \sum_{t=1-\nu}^{n-r-\nu} \sum_{t'=1-\nu'}^{n-r'-\nu'} E(e_t e_{t+r+\nu-\mu} - \sigma_e^2 1_{\{\mu=\nu+r\}}) (e_{t'} e_{t'+r'+\nu'-\mu'} - \sigma_e^2 1_{\{\mu'=\nu'+r'\}}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\nu, \mu, \nu', \mu' = -q}^{+q} b_\nu b_\mu b_{\nu'} b_{\mu'} \sum_{r=1+\nu-\mu}^{M+\nu-\mu} \sum_{r'=1+\nu'-\mu'}^{M+\nu'-\mu'} k\left(\frac{r+\mu-\nu}{M}\right) k\left(\frac{r'+\mu'-\nu'}{M}\right) \\
&\quad \times \cos((r+\mu-\nu)\omega) \cos((r'+\mu'-\nu')\omega) \\
&\quad \times \sum_{t=1-\nu}^{n-r-\nu} \sum_{t'=1-\nu'}^{n-r'-\nu'} E(e_t e_{t+r} - \sigma_e^2 1_{\{r=0\}}) (e_{t'} e_{t'+r'} - \sigma_e^2 1_{\{r'=0\}}).
\end{aligned}$$

Für den Erwartungswert in diesem letzten Ausdruck gilt nun

$$\begin{aligned}
&E(e_t e_{t+r} - \sigma_e^2 1_{\{r=0\}}) (e_{t'} e_{t'+r'} - \sigma_e^2 1_{\{r'=0\}}) \\
&= \begin{cases} Ee_1^4 - \sigma_e^4, & t = t' \text{ und } r = r' = 0 \\ 0, & t \neq t' \text{ und } r = r' = 0 \\ \sigma_e^4, & t = t' \text{ und } r = r' \neq 0 \\ \sigma_e^4, & t = t' + r' \neq t' = t + r \\ 0, & \text{sonst.} \end{cases}
\end{aligned}$$

Der Fall $t = t' + r' \neq t' = t + r$ tritt höchstens dann auf, wenn $|t' - t| \leq 2q$ gilt. Durch grobe Abschätzung erkennt man, dass dieser Anteil asymptotisch keine Rolle spielt. Gleiches gilt für den Beitrag mit $t = t'$ und $r = r' = 0$. Wir können unsere weiteren Berechnungen also auf den Fall $t = t'$ und $r = r' \neq 0$ beschränken. Da sowohl die Doppelsumme über r, r' , wie diejenige über t', t zu jeweils einer Einfachsumme zusammenfällt und die Indizes ν, μ, ν' und μ' zwischen $-q$ und q variieren, ergibt sich erneut durch direkte grobe Abschätzung, dass gilt

$$\begin{aligned}
&Var \tilde{h}(\omega) \\
&= \frac{1}{\pi^2 n^2} \sum_{\nu, \mu, \nu', \mu' = -q}^{+q} b_\nu b_\mu b_{\nu'} b_{\mu'} \sum_{r=1}^{M-2q} k\left(\frac{r+\mu-\nu}{M}\right) k\left(\frac{r+\mu'-\nu'}{M}\right) \\
&\quad \cdot \cos((r+\mu-\nu)\omega) \cos((r+\mu'-\nu')\omega) \sum_{t=1}^{n-r} \sigma_e^4 + o\left(\frac{M}{n}\right).
\end{aligned}$$

Aufgrund der gleichmäßigen Stetigkeit von k auf $[-1, +1]$ und wegen der Ungleichungen $|\nu|, |\mu|, |\nu'|, |\mu'| \leq q$ erhalten wir weiter

$$Var \tilde{h}(\omega) = \frac{\sigma_e^4}{\pi^2 n} \sum_{r=1}^M k^2\left(\frac{r}{M}\right) \left(\sum_{\nu, \mu = -q}^{+q} b_\nu b_\mu \cos((r+\mu-\nu)\omega) \right)^2.$$

Mit $\cos(x) = (e^{ix} + e^{-ix})/2$ folgt nun

$$\begin{aligned}
& \sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu \cos((r + \mu - \nu)\omega) \\
&= \frac{1}{2} \sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu \left(e^{ir\omega} e^{i(\mu-\nu)\omega} + e^{-ir\omega} e^{i(\nu-\mu)\omega} \right) \\
&= \frac{e^{ir\omega} + e^{-ir\omega}}{2} \sum_{\nu, \mu=-q}^{+q} b_\nu b_\mu e^{i(\mu-\nu)\omega},
\end{aligned}$$

also schließlich

$$\begin{aligned}
& \frac{n}{M} \text{Var} \tilde{h}(\omega) \\
&= \frac{\sigma_e^4}{4\pi^2} \left| \sum_{\nu=-q}^q b_\nu e^{-i\nu\omega} \right|^4 \frac{1}{M} \sum_{r=1}^M k^2\left(\frac{r}{M}\right) (e^{2ir\omega} + e^{-2ir\omega} + 2) + o(1).
\end{aligned}$$

Wegen $f(\omega) = \frac{\sigma_e^2}{2\pi} \left| \sum_{\nu=-q}^q b_\nu e^{-i\nu\omega} \right|^2$ und da für $M \rightarrow \infty$

$$\frac{1}{M} \sum_{r=1}^M k^2\left(\frac{r}{M}\right) \xrightarrow{M \rightarrow \infty} \int_0^1 k^2(u) du = \frac{1}{2} \int_{-1}^{+1} k^2(u) du,$$

sowie für $0 < \omega < \pi$

$$\frac{1}{M} \sum_{r=1}^M k^2\left(\frac{r}{M}\right) e^{2ir\omega} \rightarrow 0, \quad (12.52)$$

(vgl. Aufgabe 12.17) folgt die Aussage (12.49). Zum Nachweis von (12.50) kann man abschließend einfach die Cauchy-Schwarz-Ungleichung auf (12.49) anwenden. \square

Da sich die Varianz des Spektraldichteschätzers \hat{f}_n multipliziert mit $n/M(n)$ asymptotisch stabilisiert, stellt sich die Frage, ob wir für

$$\sqrt{\frac{n}{M(n)}} \left(\hat{f}_n(\omega) - f(\omega) \right), \quad 0 \leq \omega \leq \pi, \quad (12.53)$$

eine zentrale Grenzwertaussage beweisen können. Macht man die für Kurvenschätzung typische Zerlegung in Varianzanteil $\hat{f}_n(\omega) - E\hat{f}_n(\omega)$ und Biasanteil $E\hat{f}_n(\omega) - f(\omega)$, so sind diese beiden Anteile getrennt voneinander zu untersuchen. Um den Biasanteil adäquat behandeln zu können, treffen wir folgende Annahme an das Lag Window k

$$\lim_{u \rightarrow 0} \frac{1 - k(u)}{u^2} = K > 0. \quad (12.54)$$

Für alle Lag Windows der Tabelle 12.1, mit der Ausnahme des Bartlett und des Modified Bartlett Windows, ist die Voraussetzung erfüllt. Dann kann das folgende Resultat gezeigt werden.

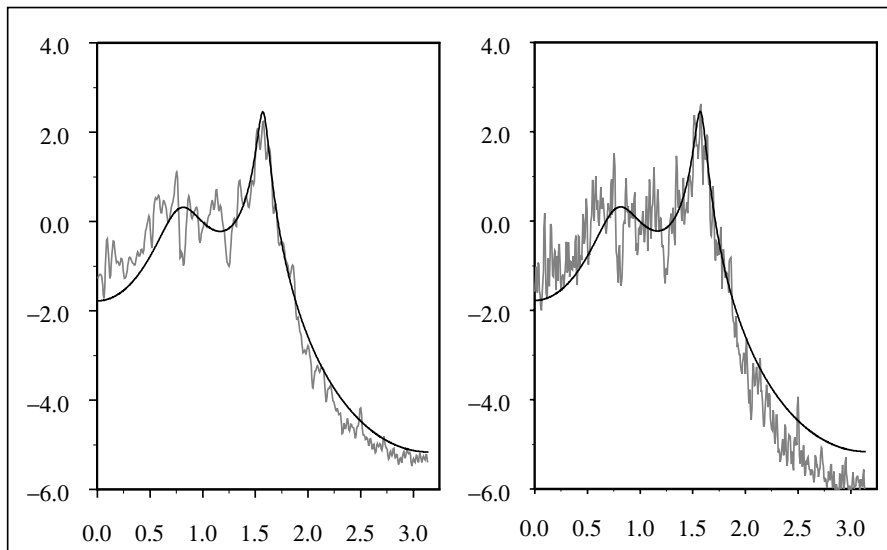


Abb. 12.6 Lag-Window-Schätzer $\hat{f}(\omega) = \frac{1}{2\pi} \sum_{h=-M}^M (1 - \frac{|h|}{M}) \hat{\gamma}(h) e^{-ih\omega}$ für ein simuliertes AR(4)-Modell $X_t - 1.13X_{t-2} + 1.54X_{t-2} - 1.02X_{t-3} + 0.58X_{t-4} = e_t$ der Länge $n = 1000$ mit standard normalverteiltem weißen Rauschen. In \hat{f} (links mit $M = 200$ und rechts mit $M = 500$) jeweils mit logarithmierter Spektraldichte (glatte Kurve)

Satz 12.15 (*Asymptot. Bias des Lag-Window-Spektraldichteschätzers*)

Unter den Voraussetzungen von Satz 12.14 mit $\sum_{\nu} \nu^2 |b_{\nu}| < \infty$ und falls k zusätzlich (12.54) erfüllt sowie $M^2(n)/n \rightarrow 0$ für $n \rightarrow \infty$ gilt, erhalten wir für $0 \leq \omega \leq \pi$

$$M^2(n) \cdot \left(E \hat{f}_n(\omega) - f(\omega) \right) \rightarrow K f''(\omega). \quad (12.55)$$

Beweis: (12.55) folgt aus den für $n \rightarrow \infty$ geltenden drei Aussagen

$$M^2(n) \sum_{|h| > M(n)} \gamma(h) e^{-ih\omega} \rightarrow 0, \quad (12.56)$$

$$M^2(n) \sum_{|h| \leq M(n)} k \left(\frac{h}{M(n)} \right) \frac{|h|}{n} \gamma(h) e^{-ih\omega} \rightarrow 0, \quad (12.57)$$

$$\frac{M^2(n)}{2\pi} \sum_{|h| \leq M(n)} \left(k \left(\frac{h}{M(n)} \right) - 1 \right) \gamma(h) e^{-ih\omega} \rightarrow K f''(\omega). \quad (12.58)$$

Die linke Seite von (12.56) ist beschränkt durch $\sum_{|h|>M(n)} h^2 |\gamma(h)| < \infty$, da $\sum_{\nu} \nu^2 |b_{\nu}| < \infty$, konvergiert also wegen $M(n) \rightarrow \infty$ gegen Null. Wegen der Konvergenz $M^2(n)/n \rightarrow 0$ gilt (12.57). Schließlich ist die linke Seite von (12.58), abgesehen vom Faktor $1/(2\pi)$, gleich

$$\sum_{|h| \leq M(n)} \frac{k\left(\frac{h}{M(n)}\right) - 1}{\frac{h^2}{M^2(n)}} h^2 \gamma(h) e^{-ih\omega},$$

wobei die Konvention $\frac{0}{0} = 0$ verwendet werde. Zu gegebenen $\varepsilon > 0$ existiert aufgrund von (12.54) ein $\delta > 0$, so dass

$$\left| \frac{k\left(\frac{h}{M(n)}\right) - 1}{\frac{h^2}{M^2(n)}} + K \right| \leq \varepsilon, \quad \forall \frac{|h|}{M(n)} \leq \delta.$$

Da

$$\begin{aligned} & \sum_{\delta \cdot M(n) < |h| \leq M(n)} \frac{\left| k\left(\frac{h}{M(n)}\right) - 1 \right|}{\frac{h^2}{M^2(n)}} h^2 |\gamma(h)| \\ & \leq \frac{\sup_{0 \leq u \leq 1} |k(u)| + 1}{\delta^2} \sum_{|h| > \delta \cdot M(n)} h^2 |\gamma(h)| \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

folgt

$$\lim_{n \rightarrow \infty} \sum_{|h| \leq M(n)} \frac{k\left(\frac{h}{M(n)}\right) - 1}{\frac{h^2}{M^2(n)}} h^2 \gamma(h) e^{-ih\omega} = -K \sum_{h=-\infty}^{\infty} h^2 \gamma(h) e^{-ih\omega}.$$

Beachtet man nun noch, dass in der Darstellung

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-ih\omega}$$

zweimal unter der Reihe differenziert werden darf, so folgt die Behauptung. \square

Nach diesen Vorarbeiten sind wir nun in der Lage, eine zentrale Grenzwertaussage für den Spektraldichteschätzer \hat{f}_n zu beweisen. Das Resultat lautet wie folgt.

Satz 12.16 (*Asymp. Normalität des Lag-Window-Spektraldichteschätzers*)
 $X = (X_t : t \in \mathbb{Z})$ sei ein linearer Prozeß mit Darstellung (12.33) mit i.i.d. weißem Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t^2 = \sigma_e^2$, $Ee_t^4 < \infty$ und absolut summierbaren Koeffizienten $(b_{\nu} : \nu \in \mathbb{Z})$. \hat{f}_n bezeichne den Spektraldichteschätzer

gemäß (12.46). Das zugehörige Lag Window k erfülle (12.47) und es gelte $M(n) \rightarrow \infty$ sowie $M(n)/n \rightarrow 0$ für $n \rightarrow \infty$. Dann gilt

$$\sqrt{\frac{n}{M(n)}} \left(\hat{f}_n(\omega) - E\hat{f}_n(\omega) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2(\omega)), \quad (12.59)$$

mit

$$\tau^2(\omega) = \begin{cases} 2f^2(\omega) \int_{-1}^{+1} k^2(u) du, & \omega = 0, \pi, \\ f^2(\omega) \int_{-1}^{+1} k^2(u) du, & 0 < \omega < \pi. \end{cases} \quad (12.60)$$

Gilt zusätzlich (12.54), $\sum_{\nu=1}^{\infty} \nu^2 |b_\nu| < \infty$ und $n/M^5(n) \rightarrow C^2 \geq 0$ für $n \rightarrow \infty$, so folgt darüberhinaus für alle $0 \leq \omega \leq \pi$ und K aus (12.54)

$$\sqrt{\frac{n}{M(n)}} \left(\hat{f}_n(\omega) - f(\omega) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(C \cdot K \cdot f''(\omega), \tau^2(\omega)). \quad (12.61)$$

Beweis: (12.61) folgt aus Satz 12.15 und (12.59) durch direktes Einsetzen. Zum Nachweis von (12.59) beschränken wir uns auf den einfachsten Fall eines weißen Rauschens, d.h. $b_0 = 1$ und $b_\nu = 0$ für alle $\nu \neq 0$. In dieser Situation gilt (wir schreiben wieder M statt $M(n)$)

$$\begin{aligned} & \sqrt{\frac{n}{M}} \left(\hat{f}_n(\omega) - E\hat{f}_n(\omega) \right) \\ &= \frac{1}{2\pi} \frac{1}{\sqrt{n}} \sum_{t=1}^n \underbrace{\frac{1}{\sqrt{M}} \sum_{|h| \leq M} k\left(\frac{h}{M}\right) e^{-ih\omega} (e_t e_{t+|h|} - \sigma_e^2 1_{\{h=0\}})}_{=: Z_{t,n}} + o_P(1) \end{aligned}$$

$Z_{t,n}$ hängt nur von e_t, \dots, e_{t+M} ab und es gilt $EZ_{t,n} = 0$ sowie

$$\begin{aligned} EZ_{t,n}^2 &= \frac{1}{M} \sum_{|r|, |s| \leq M} k\left(\frac{r}{M}\right) k\left(\frac{s}{M}\right) e^{-i(r+s)\omega} E\left(e_t e_{t+|r|} - \sigma_e^2 1_{\{r=0\}}\right) \\ &\quad \cdot \left(e_t e_{t+|s|} - \sigma_e^2 1_{\{s=0\}}\right) \\ &= \frac{Ee_1^4 - \sigma_e^4}{M} + \frac{1}{M} \sum_{|r| \leq M} k^2\left(\frac{r}{M}\right) (e^{-2ir\omega} + 1) \sigma_e^4 \\ &\xrightarrow{M \rightarrow \infty} \begin{cases} 2\sigma_e^4 \int_{-1}^{+1} k^2(u) du, & \omega = 0, \pi, \\ \sigma_e^4 \int_{-1}^{+1} k^2(u) du, & 0 < \omega < \pi, \end{cases} \end{aligned}$$

vgl. Aufgabe 12.17. Da

$$EZ_{s,n} Z_{t,n} = 0 \quad \text{für } s \neq t$$

erhalten wir

$$\lim_{n \rightarrow \infty} \text{Var} \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,n} = (2\pi)^2 \tau^2(\omega).$$

Die letzte Aussage folgt übrigens auch als Spezialfall aus Satz 12.14, (12.49).

Wir werden nun $\sum_{t=1}^n Z_{t,n}$ durch eine Summe unabhängig und identisch verteilter Zufallsvariablen approximieren. Man definiere dazu mit $N(n) = [\sqrt{nM(n)}]$ für $j = 1, \dots, [n/N(n)]$

$$Y_{j,n} := \frac{1}{\sqrt{N(n)}} (Z_{(j-1)N(n)+1,n} + \dots + Z_{jN(n)-M(n),n}). \quad (12.62)$$

Wegen $M(n)/N(n) \rightarrow 0$ für $n \rightarrow \infty$ gilt

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_{t,n} - \frac{1}{\sqrt{[n/N(n)]}} \sum_{j=1}^{[n/N(n)]} Y_{j,n} = o_P(1)$$

und außerdem sind die $Y_{j,n}$ tatsächlich i.i.d. Da $EY_{j,n}^4$ gleichmäßig in n beschränkt ist (vgl. Aufgabe 12.19) folgt schließlich mit dem Zentralen Grenzwertsatz von Lyapounov, siehe Bauer (1991), §28, die gewünschte asymptotische Normalität. \square

Bemerkung 12.17

Die asymptotische Normalverteilung des geglätteten Spektraldichteschätzers \hat{f}_n ist zentriert, falls $C = 0$, d.h. $n/M^5(n) \rightarrow 0$.

Da die Varianz des Spektraldichteschätzers mit Geschwindigkeit $M(n)/n$ konvergiert und der quadrierte Bias (vgl. Satz 12.15) mit Geschwindigkeit $1/M^4(n)$ konvergiert, ist zu erwarten, dass sich der Mean Square Error (MSE) $E(\hat{f}_n(\omega) - f(\omega))^2$ asymptotisch wie

$$C_1 \frac{M(n)}{n} + C_2 \frac{1}{M^4(n)}$$

(für zwei Konstanten $C_1, C_2 > 0$) verhält. Dieser asymptotische MSE wird gerade für $M(n) \sim n^{1/5}$ minimal, so dass tatsächlich die nicht zentrierte Limesverteilung einen besonders relevanten Fall darstellt. \square

Grundsätzlich besteht die Möglichkeit parametrische und nichtparametrische Spektraldichteschätzer zu kombinieren. Eine auf Tukey zurückgehende und als *prewhitening* bezeichnete Idee besteht darin, zunächst ein parametrisches Zeitreihenmodell an vorliegende Daten einer stationären Zeitreihe (X_t) mit Spektraldichte f_X anzupassen. Stellen wir uns vor, dass wir eine *AR*-Zeitreihe

der Ordnung p verwandt und Parameterschätzer $\hat{a}_1, \dots, \hat{a}_p$ erhalten haben. Wenden wir in einem nächsten Schritt den AR -Filter

$$Y_t := X_t + \sum_{j=1}^p \hat{a}_j X_{t-j} \quad (12.63)$$

an, so erhalten wir Beobachtungen einer Zeitreihe (Y_t) , deren Spektraldichte f_Y der Gleichung

$$f_Y(\omega) = \left| 1 + \sum_{j=1}^p \hat{a}_j e^{-ij\omega} \right|^2 f_X(\omega) \quad (12.64)$$

genügt. Falls die Approximation durch das autoregressive Modell hinreichend gut ist, stellt f_Y approximativ die Spektraldichte eines weißen Rauschens (*prewhitening*) dar, häufig aber zumindest eine recht glatte Spektraldichte (ohne ausgeprägte Spitzen und Senken). Eine solche Spektraldichte f_Y ist den vorgestellten nichtparametrischen Spektraldichteschätzmethode in der Regel weitaus besser zugänglich als die Ursprungsspektraldichte f_X . Wendet man nun basierend auf den gefilterten Daten Y_t eine nichtparametrische Schätzmethode (etwa einen Lag-Window-Schätzer) an und erhält den Schätzer \hat{f}_Y für die Spektraldichte f_Y , so können wir durch Auflösen nach f_X in (12.64) leicht einen gemischt parametrisch-nichtparametrischen Spektraldichteschätzer ableiten. In vielen Anwendungsfällen führt ein solches Vorgehen, nämlich parametrisches Schätzen der Grob- und nichtparametrisches Schätzen der Feinstruktur einer Spektraldichte, in der Tat zu guten Ergebnissen. Selbstverständlich ist es nicht zwingend ein autoregressives parametrisches Modell zu verwenden. Ganz genauso können MA -, $ARMA$ - und auch andere parametrische Modelle für den ersten Schätzschritt eingesetzt werden.

Zur Veranschaulichung der Idee des *prewhitening* wurde Daten aus einem stationären $AR(4)$ -Modell mit extremem peak der Spektraldichte in der Frequenz $\omega = \pi/2$ und moderatem peak in $\omega = \pi/4$ simuliert. Es ist in Abbildung 12.7 (rechts) zu sehen, dass ein Lag-Window-Schätzer durch den ausgeprägten peak der tatsächlichen Spektraldichte verfälscht wird. Da autoregressive Modelle der Ordnung 2 in der Lage sind eine ausgeprägte Frequenz in der Spektraldichte abzubilden, berechnen wir zunächst einen $AR(2)$ -Spektraldichteschätzer und wenden den sich gemäß (12.63) ergebenden Filter auf die simulierten Daten an. So eliminiert man in der Tat zu einem gewissen Grad den ausgeprägten peak an der Stelle $\omega = \pi/2$. Wendet man dann erst in einem zweiten Schritt auf die berechneten Werte Y_t den Lag-Window-Schätzer an, so ergibt sich eine deutliche bessere Approximation an die tatsächliche Spektraldichte (vgl. Abbildung 12.7).

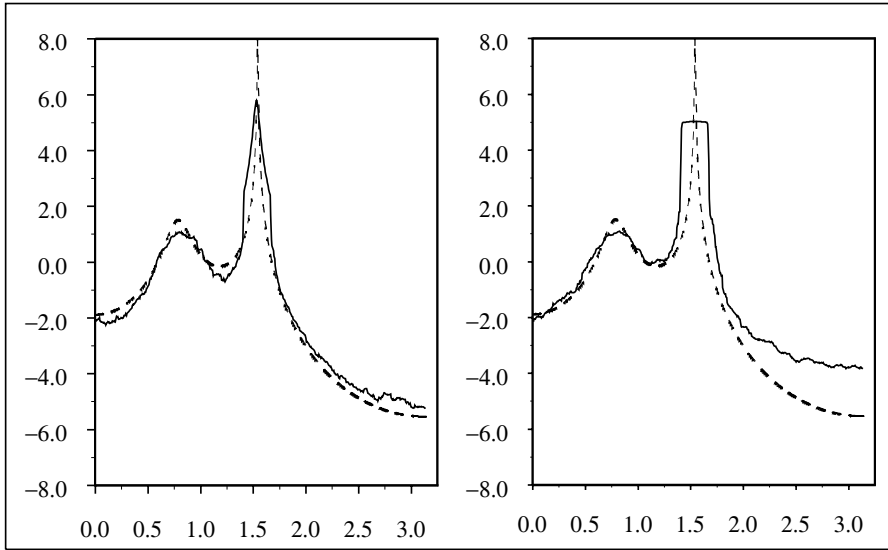


Abb. 12.7 Geschätzte Spektraldichte für ein $AR(4)$ -Modell ($n = 1000$) mit Hilfe des prewhitening über ein angepasstes $AR(2)$ -Modell kombiniert mit einem Lag-Window-Schätzer mit dem Truncated-Daniell-Fenster (vgl. Tabelle 12.1) für $M = 20$ (links). Zum Vergleich eine direkte Glättung (ebenfalls mit dem Truncated-Daniell-Fenster mit $M = 20$) des Periodogramms (rechts) jeweils zum Vergleich mit der logarithmierten tatsächlichen Spektraldichte (gestrichelt)

12.5 Das geglättete Periodogramm

Zu Beginn von Abschnitt 12.4 haben wir kurz Glättungen des Periodogramms von der Form

$$\frac{1}{2m+1} \sum_{|k| \leq m} I_n(\omega_{j+k}) \quad (12.65)$$

$\omega_j = 2\pi j/n$, bzw.

$$\int_{-\pi}^{\pi} W(\omega) I_n(\omega) d\omega \quad (12.66)$$

mit einer stark um eine Stelle $\omega_0 \in [0, \pi]$ herum konzentrierten Funktion W , vorgestellt. Wir nennen Schätzer der Form (12.66) *integrierte Periodogrammschätzer*. Wir werden nun zeigen, dass die in Abschnitt 12.4 behandelten konsistenten Spektraldichteschätzer \hat{f}_n (Lag-Window-Schätzer, vgl. (12.46)) und die obigen Glättungen des Periodogramms näher miteinander verwandt sind, als vielleicht auf den ersten Blick zu vermuten wäre. Es gilt nämlich wegen $\bar{Y}_n(h) = \int_{-\pi}^{\pi} e^{ih\omega} I_n(\omega) d\omega$ (beachte (12.6) und (12.15))

$$\begin{aligned}
\widehat{f}_n(\omega) &= \frac{1}{2\pi} \sum_{|h| \leq M} k\left(\frac{h}{M}\right) \bar{y}_n(h) e^{-ih\omega} \\
&= \int_{-\pi}^{\pi} I_n(v) W_n(\omega - v) dv = \int_{-\pi}^{\pi} I_n(\omega - v) W_n(v) dv \quad (12.67)
\end{aligned}$$

mit

$$W_n(v) := \frac{1}{2\pi} \sum_{|h| \leq M} k\left(\frac{h}{M}\right) e^{-ihv}.$$

(Man beachte für das letzte Gleichheitszeichen die 2π -Periodizität von I_n und W_n . Der Index n bei W_n rührt von der Tatsache her, dass $M = M(n)$ eine Funktion von n ist.) Diese Darstellung entspricht (12.66). Approximiert man nun das Integral in (12.67) mit Hilfe der Fourier-Frequenzen $\omega_j = 2\pi j/n$, so erhält man die (12.65) ähnelnde Darstellung

$$\frac{2\pi}{n} \sum_{j=-[(n-1)/2]}^{[n/2]} I_n(\omega_j) W_n(\omega - \omega_j) = \sum_{j=-[(n-1)/2]}^{[n/2]} W_{n,j} I_n(\omega_j) \quad (12.68)$$

mit allgemeinen Gewichten $W_{n,j} := \frac{2\pi}{n} W_n(\omega - \omega_j)$.

Bevor wir umgekehrt zeigen, dass integrierte Periodogrammschätzer auch als Lag-Window-Schätzer geschrieben werden können, zunächst zwei Beispiele.

Beispiel 12.18 (Bartlett Lag Window)

Für das Bartlett Lag Window $k(u) = 1_{[-1, +1]}(u)$ erhalten wir unter Beachtung von (A.6)

$$\begin{aligned}
W_n(v) &= \frac{1}{2\pi} \sum_{|h| \leq M} e^{-ihv} \\
&= \begin{cases} \frac{2M+1}{2\pi}, & v = 0, \pm 2\pi \\ \frac{1}{2\pi} \frac{\sin((M+1/2)v)}{\sin(v/2)}, & -2\pi < v < 2\pi \text{ und } v \neq 0 \end{cases} \\
&= \frac{1}{2\pi} \cdot \frac{\sin((M+1/2)v)}{\sin(v/2)}, -2\pi \leq v \leq 2\pi.
\end{aligned}$$

□

Aus der Tatsache, dass W_n für das Bartlett Lag Window auch negative Werte annehmen kann, folgt, dass sowohl die Version (12.67) als auch die Version (12.68) des geglätteten Periodogramms negative Werte annehmen können, was für einen Spektraldichteschätzer eine eher unerwünschte Eigenschaft ist.

Beispiel 12.19 (Modified Bartlett Lag Window)

Für das Lag Window $k(u) = 1 - |u|$, $-1 \leq u \leq 1$, ergibt sich (vgl. Aufgabe 12.20)

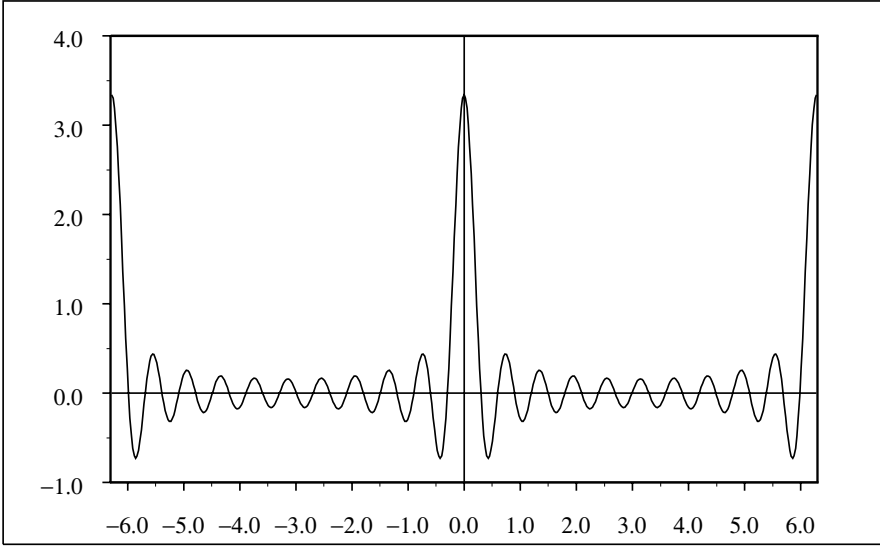


Abb. 12.8 Darstellung von $\frac{1}{2\pi} \frac{\sin((M+1/2)v)}{\sin(v/2)}$ für $M = 10$

$$W_n(v) = \frac{1}{2\pi M} \frac{\sin^2(Mv/2)}{\sin^2(v/2)}, \quad -2\pi \leq v \leq 2\pi.$$

Wegen $W_n(v) \geq 0$ folgt, dass sowohl der Schätzer (12.67) als auch (12.68) stets nichtnegativ sind. \square

Die Varianz eines Spektraldichteschätzers in der Darstellung (12.68) können wir wie folgt heuristisch approximieren. Zunächst gilt

$$\begin{aligned} \int_{-\pi}^{+\pi} W_n^2(\omega - v) dv &= \int_{-\pi}^{\pi} \frac{1}{4\pi^2} \sum_{|r|, |s| \leq M} k\left(\frac{r}{M}\right) k\left(\frac{s}{M}\right) e^{-i(r+s)(\omega-v)} dv \\ &= \frac{1}{2\pi} \sum_{|r| \leq M} k^2\left(\frac{r}{M}\right) \approx \frac{M}{2\pi} \cdot \int_{-1}^{+1} k^2(u) du. \end{aligned}$$

Für das letzte Gleichheitszeichen beachte man die Orthogonalität der Funktionen $e^{ij\cdot}$ und die Symmetrie von k . Damit folgt weiter für $0 < \omega < \pi$

$$\begin{aligned} \text{Var} \hat{f}_n(\omega) &\approx \frac{M}{n} \cdot \int_{-1}^{+1} k^2(u) du \cdot f^2(\omega) \\ &\approx \frac{2\pi}{n} \int_{-\pi}^{\pi} W_n^2(\omega - v) dv \cdot f^2(\omega). \end{aligned} \quad (12.69)$$

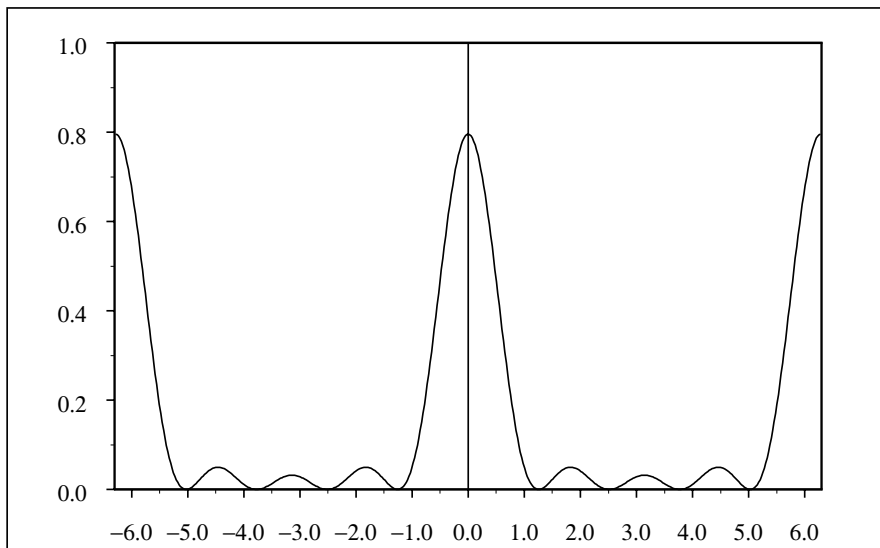


Abb. 12.9 Darstellung von $\frac{1}{2\pi M} \frac{\sin^2(Mv/2)}{\sin^2(v/2)}$ für $M = 5$

Wir zeigen nun, dass ein integrierter Periodogrammschätzer der Form

$$\int_{-\pi}^{\pi} W(\omega - v) I_n(v) dv = \int_{-\pi}^{\pi} W(v) I_n(\omega - v) dv$$

vgl. (12.67), in eine zu (12.46) vergleichbare Form gebracht werden kann. Dabei nehmen wir an, dass W eine 2π -periodische und gerade Funktion ist, für die die zugehörige Fourier Reihe

$$\frac{1}{2\pi} \sum_{h=-\infty}^{\infty} k(h) e^{-i\omega h},$$

mit $k(h) = \int_{-\pi}^{\pi} W(\omega) e^{ih\omega} d\omega$, $h \in \mathbb{Z}$, für alle $\omega \in [-\pi, \pi]$ gegen $W(\omega)$ konvergiert. Damit erhalten wir

$$\begin{aligned} \int_{-\pi}^{\pi} W(v) I_n(\omega - v) dv &= \frac{1}{2\pi} \sum_{|h| < n} \bar{\gamma}_n(h) \int_{-\pi}^{\pi} W(v) e^{ihv} dv e^{-ih\omega} \\ &= \frac{1}{2\pi} \sum_{|h| < n} k(h) \bar{\gamma}_n(h) e^{-ih\omega}, \end{aligned}$$

also eine Lag Window Form.

Beispiel 12.20 (Daniell Lag Window)

Verwendet man für $M > 1$

$$W_M(v) = \frac{M}{2\pi} 1_{[-\pi/M, \pi/M]}(v) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \frac{\sin(h\pi/M)}{h\pi/M} e^{-ihv},$$

so ergibt sich konkret

$$\begin{aligned} \frac{M}{2\pi} \int_{-\pi/M}^{\pi/M} I_n(\omega - v) dv &= \int_{-\pi}^{\pi} I_n(\omega - v) b_M(v) dv \\ &= \frac{1}{2\pi} \sum_{|h| < n} \frac{\sin(h\pi/M)}{h\pi/M} \cdot \bar{\gamma}_n(h) e^{-ih\omega}. \end{aligned}$$

Dies ist der sogenannte Daniell-Spektraldichteschätzer. Diskretisiert man das linke Integral für ein festes $\omega_j = 2\pi j/n$ über Stützstellen $\omega_k = 2\pi k/n$ so ergibt sich gerade

$$\frac{M}{n} \sum_{k=-[\frac{n}{2M}]+1}^{[\frac{n}{2M}]} I_n(\omega_j - \omega_k).$$

Falls $n/2M = m \in \mathbb{N}$, so erhält man im Wesentlichen den Spektraldichteschätzer (12.65). Bricht man umgekehrt in der Reihenentwicklung an der Stelle M (es sei $M \in \mathbb{N}$, $M < n$ angenommen) ab, so erhalten wir den Lag-Window-Spektraldichteschätzer \hat{f}_n gemäß (12.46) mit dem Truncated Daniell Lag Window $k(u) = \sin(\pi u)/(\pi u)$, $-1 \leq u \leq 1$. \square

12.6 Konfidenzintervalle für die Spektraldichte

Natürlich kann man die asymptotische Normalität des Spektraldichteschätzers \hat{f}_n (vgl. Satz 12.16) dazu nutzen, um punktweise Konfidenzintervalle für die Spektraldichte anzugeben. Konkret liefert

$$\hat{f}_n(\omega) \pm \left(\frac{M(n)}{n} \int_{-1}^{+1} k^2(u) du \right)^{1/2} \hat{f}_n(\omega) z_{1-\alpha/2}, \quad (12.70)$$

für $0 < \alpha < 1$, $0 < \omega < \pi$ und $z_{1-\alpha}$ das $(1-\alpha)$ -Quantil der Standardnormalverteilung ein asymptotisches Konfidenzintervall für $\hat{f}_n(\omega)$. Eine andere Variante ist

$$\hat{f}_n(\omega) \pm \left(\frac{1}{n} \sum_{|h| \leq M(n)} k^2\left(\frac{h}{M(n)}\right) \right)^{1/2} \hat{f}_n(\omega) z_{1-\alpha/2}. \quad (12.71)$$

Sowohl bei (12.70) als auch bei (12.71) wurde eine zentrierte asymptotische Normalverteilung zugrunde gelegt, d.h. von der Tatsache $n/M^5(n) \rightarrow 0$ für $n \rightarrow \infty$ ausgegangen. Macht man dies nicht, so hat man das Problem auch

$f''(\omega)$ konsistent schätzen zu müssen, was erneut eine Verkomplizierung darstellt, die wir im Rahmen dieser Einführung nicht diskutieren wollen. Beachtet man, dass aus (12.61) mit $C = 0$ und der Delta-Methode (siehe Satz A.31), angewandt auf die Logarithmusfunktion

$$\sqrt{\frac{n}{M(n)}} \left(\ln \hat{f}_n(\omega) - \ln f(\omega) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \int_{-1}^{+1} k^2(u) du \right)$$

für $0 < \omega < \pi$ folgt, so erhält man das folgende Konfidenzintervall für $\ln f(\omega)$:

$$\ln \hat{f}_n(\omega) \pm z_{1-\alpha/2} \left(\frac{M(n)}{n} \int_{-1}^{+1} k^2(u) du \right)^{1/2}. \quad (12.72)$$

Dieses Konfidenzintervall für $\ln f(\omega)$ hat den Vorteil, dass seine Breite im Gegensatz zu derjenigen der Intervalle (12.70) und (12.71) nicht mit $\omega \in (0, \pi)$ variiert. Daher wird häufig die logarithmierte Spektraldichte mit einem punktwweisen Konfidenzintervall über $(0, \pi)$ aufgetragen. Zur Veranschaulichung vergleiche mit Abbildung 12.10.

Abschließend soll noch eine auf Tukey zurückgehende alternative Metho-

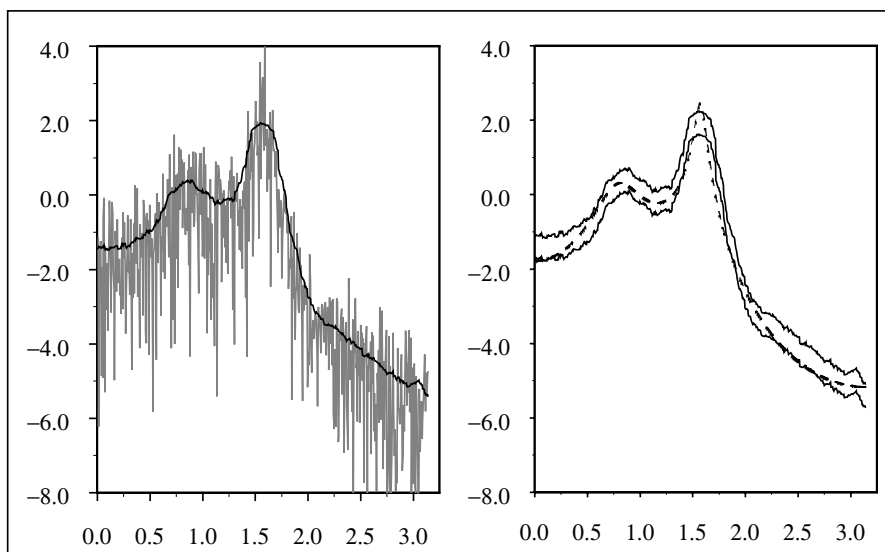


Abb. 12.10 Logarithmiertes geglättetes Periodogramm für das $AR(4)$ -Modell ($n = 1000$) aus Abbildung 12.5 (Daniell Lag Window (vgl. Beispiel 12.20 mit $M = 20$) mit logarithmiertem Periodogramm (links) und punktwweises asymptotisches 95%-Konfidenzintervall (12.72) mit logarithmierter Spektraldichte (gestrichelt) (rechts)

de zur Konstruktion von Konfidenzintervallen für die Spektraldichte vorgestellt werden. Da $I_n(\omega)$ nach Satz 12.11 für $0 < \omega < \pi$ asymptotisch exponentialverteilt mit Erwartungswert $f(\omega)$ ist und eine χ^2_2 -Verteilung einer Exponentialverteilung-Verteilung mit Parameter $1/2$ entspricht, erhält man (ebenfalls für $0 < \omega < \pi$)

$$2I_n(\omega)/f(\omega) \xrightarrow{D} \chi^2_2.$$

Die asymptotische Unabhängigkeit des Periodogramms an verschiedenen Frequenzen legt damit nahe, die Verteilung des folgenden geglätteten Periodogramms

$$\frac{1}{2m+1} \sum_{|k| \leq m} I_n(\omega_j + \omega_k)$$

durch die Verteilung von

$$\frac{1}{2(2m+1)} f(\omega_j) \sum_{|k| \leq m} \xi_k \quad (12.73)$$

mit stochastisch unabhängigen und jeweils χ^2_2 -verteilten Zufallsvariablen ξ_k zu approximieren. Die Verteilung von (12.73) stellt mit geeignetem $c > 0$ und $r \in \mathbb{N}$ aber eine $c\chi^2_r$ -verteilte Zufallsvariable dar. Da nun auch für allgemeine Glättungen des Periodogramms, etwa

$$\tilde{f}_n(\omega_j) := \sum_{j=-m}^m W_{n,j} I_n(\omega_{j+k}) \quad (12.74)$$

mit $\sum_{j=-m}^m W_{n,j} = 1$ (vgl. auch (12.68)), die Verteilungen recht gut durch eine gewichtete χ^2_r -Verteilung beschrieben werden, verwendet man approximativ

$$\mathcal{L}(\tilde{f}_n(\omega_j)) \approx c \cdot \chi^2_r,$$

wobei $c > 0$ und $r \in \mathbb{N}$ über die ersten beiden Momente angepasst werden und $\mathcal{L}(U)$ die Verteilung einer Zufallsvariablen U bedeutet. Da die χ^2_r -Verteilung den Erwartungswert r und die Varianz $2r$ besitzt, führt dies wegen $E\tilde{f}_n(\omega_j) \approx f(\omega_j)$ und $\text{Var} \tilde{f}_n(\omega_j) \approx \sum_{|j| \leq m} W_{n,j}^2 f^2(\omega_j)$ (vgl. (12.69)) zu

$$c r = f(\omega_j) \quad \text{und} \quad 2c^2 r = \sum_{|j| \leq m} W_{n,j}^2 \cdot f^2(\omega_j).$$

Hieraus ergeben sich die folgenden Festlegungen für c und r

$$c = \frac{1}{2} \sum_{|j| \leq m} W_{n,j}^2 f(\omega_j) \quad \text{und} \quad r = \left[2 / \sum_{|j| \leq m} W_{n,j}^2 \right].$$

r nennt man in diesem Zusammenhang auch *equivalent degrees of freedom* (EDF) des Spektraldichteschätzers \tilde{f}_n an der Stelle $\omega_j = 2\pi j/n$.

Im Hinblick auf Konfidenzintervalle ergibt sich hieraus

$$\mathcal{L}\left(r \cdot \frac{\tilde{f}_n(\omega_j)}{f(\omega_j)}\right) \approx \chi_r^2$$

und deshalb das folgende Konfidenzintervall für $f(\omega_j)$ zur approximativen Überdeckungswahrscheinlichkeit $1 - \alpha$

$$\left[\frac{r\tilde{f}_n(\omega_j)}{\chi_{r;1-\alpha/2}^2}, \frac{r\tilde{f}_n(\omega_j)}{\chi_{r;\alpha/2}^2} \right]. \quad (12.75)$$

Dabei bezeichnet $\chi_{r;\alpha}^2$ das α -Quantil der χ_r^2 -Verteilung ($0 < \alpha < 1$). Durch Logarithmieren erhält man wie oben ein Konfidenzintervall für $\ln f(\omega_j)$ mit fester Breite für alle Frequenzen, nämlich

$$\left[\ln \tilde{f}_n(\omega_j) + \ln r - \ln \chi_{r;1-\alpha/2}^2, \ln \tilde{f}_n(\omega_j) + \ln r - \ln \chi_{r;\alpha/2}^2 \right]. \quad (12.76)$$

12.7 Das integrierte Periodogramm

In Abschnitt 12.4 haben wir im Rahmen der notwendigen Glättung des in Definition 12.2 erklärten Periodogramms I_n sogenannte integrierte Periodogrammschätzer

$$\int_{-\pi}^{\pi} W(\omega) I_n(\omega) d\omega \quad (12.77)$$

mit einer um die Frequenz ω_0 konzentrierten Funktion $W : [-\pi, \pi] \rightarrow \mathbb{R}$ kurz eingeführt und in Abschnitt 12.5 ihren Zusammenhang zu den Lag-Window-Schätzern diskutiert. In diesem Abschnitt wollen wir nun integrierte Periodogrammschätzer (12.77) für beliebige Funktionen W detaillierter betrachten. Für weitergehende Untersuchungen sei auf Dahlhaus (1985) verwiesen. Schränkt man sich auf symmetrische Funktionen W ein, so reicht es wegen der Symmetrie des Periodogramms Ausdrücke der Form

$$M(I_n, W) = \int_0^{\pi} W(\omega) I_n(\omega) d\omega \quad (12.78)$$

für geeignete Funktionen $W : [0, \pi] \rightarrow \mathbb{R}$ zu betrachten.

Wir stellen nun einige integrierte Periodogrammschätzer vor.

Beispiel 12.21 (Verschiedene integrierte Periodogrammschätzer)

- (i) Für $W(\omega) = 2 \cos(\omega h)$, $h \in \mathbb{N}_0$, erhalten wir mit Hilfe der Darstellung (12.1) $M(I_n, W) = \bar{\gamma}_n(h)$, wobei $\bar{\gamma}_n(h)$ in (12.6) definiert wird. $\bar{\gamma}_n(h)$ ist bekanntlich unter geeigneten Voraussetzungen ein \sqrt{n} -konsistenter Schätzer für die Autokovarianz, die ihrerseits $\gamma(h) = \int_0^\pi 2 \cos(\omega h) f(\omega) d\omega$ erfüllt. Dabei bezeichnet f die zugrunde liegende und als existierend vorausgesetzte Spektraldichte.
- (ii) Für $W(\omega) = 1_{[0,x]}(\omega)$, $x \in [0, \pi]$ ist $M(I_n, W)$ das sogenannte *integrierte Periodogramm*

$$\int_0^x I_n(\omega) d\omega,$$

welches einen Schätzer für die sog. *Spektralverteilungsfunktion* $\int_0^x f(\omega) d\omega$ an der Stelle x darstellt.

- (iii) Stellt W hingegen eine mehr oder weniger stark um eine feste Frequenz $\omega_0 \in [0, \pi]$ konzentrierte Funktion dar, etwa

$$W(\omega) = \frac{1}{h} K\left(\frac{\omega_0 - \omega}{h}\right), \quad (12.79)$$

mit einer zentrierten Wahrscheinlichkeitsdichte K (ein sogenannter *Kern*) und $h > 0$ eine *Bandweite*, so stellt $M(I_n, W)$ eine Glättung des Periodogramms um die Frequenz ω_0 und damit einen Spektraldichteschätzer für $f(\omega_0)$ dar. \square

In vielen Fällen wird man $M(I_n, W)$ mit Hilfe einer Riemannschen Summe mit den Stützstellen der Fourier-Frequenzen $\omega_j = 2\pi j/n$, $j = 0, \dots, N = [n/2]$, approximieren. Dies führt zu der folgenden diskretisierten Version von (12.78)

$$\frac{2\pi}{n} \sum_{j=0}^N W(\omega_j) I_n(\omega_j). \quad (12.80)$$

Unser Ziel ist es, asymptotische Verteilungsaussagen für $M(I_n, W)$ herzuleiten. Das Beispiel 12.21 (i) läßt vermuten, dass wir für geeignete Funktionen W eine \sqrt{n} -Konsistenz mit einer asymptotischen Normalverteilung erwarten können. In der Tat hat Dahlhaus (1985) bewiesen, dass unter gewissen Annahmen an die Abhängigkeitsstruktur des zugrunde liegenden Prozesses für eine große Klasse von Funktionen W gezeigt werden kann, dass

$$\begin{aligned} & \sqrt{n}(M(I_n, W) - M(f, W)) \\ &= \sqrt{n} \left(\int_0^\pi W(\omega) I_n(\omega) d\omega - \int_0^\pi W(\omega) f(\omega) d\omega \right) \end{aligned} \quad (12.81)$$

eine zentrierte asymptotische Normalverteilung besitzt. Für die Klasse der linearen Zeitreihen mit i.i.d. weißem Rauschen (vgl. Definition 10.3) gilt das folgende Resultat.

Satz 12.22 (*Asymptot. Normalität für integrierte Periodogrammschätzer*)

$X = (X_t : t \in \mathbb{Z})$ sei eine kausale, lineare Zeitreihe mit i.i.d. weißem Rauschen im Sinne von Definition 10.3. Für die Koeffizienten b_j der $MA(\infty)$ -Darstellung gelte $\sum_{j=1}^{\infty} j |b_j| < \infty$ und für das weiße Rauschen wird neben $E e_1 = 0$ und $\text{Var } e_1 = \sigma_e^2$ auch $E e_1^4 =: \eta \sigma_e^4 < \infty$ vorausgesetzt. f bezeichne die zur Zeitreihe X gehörende Spektraldichte. Dann gilt für jede stetig differenzierbare Funktion $W : [0, \pi] \rightarrow \mathbb{R}$

$$\sqrt{n}(M(I_n, W) - M(f, W)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau^2) \quad (12.82)$$

mit der asymptotischen Varianz

$$\tau^2 = 2\pi \int_0^\pi W^2(\omega) f^2(\omega) d\omega + (\eta - 3) \left(\int_0^\pi W(\omega) f(\omega) d\omega \right)^2. \quad (12.83)$$

Beweis: Zunächst gilt aufgrund der Symmetrie von $\bar{\gamma}_n(h)$

$$\begin{aligned} I_n(\omega) &= \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} \bar{\gamma}_n(h) e^{-ih\omega} \\ &= \frac{1}{2\pi} \left(\bar{\gamma}_n(0) + 2 \sum_{h=1}^{n-1} \bar{\gamma}_n(h) \cos(h\omega) \right). \end{aligned}$$

Ganz analog erhält man aus der Inversionsformel Satz 3.8

$$f(\omega) = \frac{1}{2\pi} \left(\gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(h\omega) \right). \quad (12.84)$$

Wegen $\gamma(h) = \sigma_e^2 \sum_{j=0}^{\infty} b_j b_{j+h}$ gilt

$$\begin{aligned} \sqrt{n} \left| \sum_{h=n}^{\infty} \gamma(h) \cos(h\omega) \right| &\leq \sqrt{n} \cdot \sigma_e^2 \sum_{h=n}^{\infty} \sum_{j=0}^{\infty} |b_j b_{j+h}| \\ &\leq \sigma_e^2 \sum_{j=0}^{\infty} |b_j| \sum_{h=n}^{\infty} \sqrt{j+h} |b_{j+h}| \leq \sigma_e^2 \sum_{j=0}^{\infty} |b_j| \sum_{k=n}^{\infty} \sqrt{k} |b_k| \rightarrow 0, \text{ für } n \rightarrow \infty, \end{aligned}$$

denn $\sum_{k=1}^{\infty} \sqrt{k} |b_k| < \infty$. Deshalb erhalten wir

$$\begin{aligned} &\sqrt{n}(M(I_n, W) - M(f, W)) \\ &= \frac{\sqrt{n}}{2\pi} \left((\bar{\gamma}_n(0) - \gamma(0)) \cdot \int_0^\pi W(\omega) d\omega \right. \\ &\quad \left. + 2 \cdot \sum_{h=1}^{n-1} (\bar{\gamma}_n(h) - \gamma(h)) \cdot \int_0^\pi W(\omega) \cos(h\omega) d\omega \right) + o(1) \\ &=: Z_n + o(1) \end{aligned} \quad (12.85)$$

Man definiere nun für $M < n$

$$Z_{n,M} = \sqrt{n} \left((\bar{\gamma}_n(h) - \gamma(h) : h = 0, \dots, M) (w_0, w_1, \dots, w_M)^T \right) \quad (12.86)$$

mit

$$\begin{aligned} w_0 &:= \frac{1}{2\pi} \int_0^\pi W(\omega) d\omega \quad \text{und} \\ w_h &:= \frac{1}{\pi} \int_0^\pi W(\omega) \cos(h\omega) d\omega \quad (h \geq 1). \end{aligned} \quad (12.87)$$

Mit Hilfe von Satz A.29 aus Anhang A.5 folgt die Behauptung aus den folgenden drei Aussagen

$$Z_{n,M} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_M^2) \quad \text{für } n \rightarrow \infty \quad \forall M, \quad (12.88)$$

$$\tau_M^2 \rightarrow \tau^2 \quad \text{für } M \rightarrow \infty \quad (12.89)$$

und

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} P\{|Z_n - Z_{n,M}| > \varepsilon\} = 0 \quad \forall \varepsilon > 0. \quad (12.90)$$

Zunächst ist die Aussage (12.88) unter Beachtung von Satz 10.15 klar, wenn man τ_M^2 wie folgt definiert

$$\tau_M^2 := (w_0, \dots, w_M) \mathbf{S}_M (w_0, \dots, w_M)^T. \quad (12.91)$$

Es gilt (man beachte die Definition von \mathbf{S}_M aus Satz 10.15)

$$\begin{aligned} \tau_M^2 &= (\eta - 3) \cdot \left(\sum_{h=0}^M w_h \gamma(h) \right)^2 \\ &+ \sum_{r=-\infty}^{\infty} \sum_{h,k=0}^M w_h w_k \left\{ \gamma(r) \gamma(r+k-h) + \gamma(r+k) \gamma(r-h) \right\}. \end{aligned} \quad (12.92)$$

(12.90) sieht man wie folgt ein. Unter Verwendung der Markov-Ungleichung reicht die Betrachtung von $E|Z_n - Z_{n,M}|$ aus. Hierfür gilt

$$E|Z_n - Z_{n,M}| \leq \sum_{h=M+1}^{n-1} \sqrt{n} |w_h| E|\bar{\gamma}_n(h) - \gamma(h)|.$$

Da

$$(\sqrt{n} E|\bar{\gamma}_n(h) - \gamma(h)|)^2 \leq n \cdot E(\bar{\gamma}_n(h) - \gamma(h))^2$$

und die rechte Seite gleichmäßig in h nach oben beschränkt ist (vgl. Aufgabe 12.19), erhalten wir die gewünschte Aussage (12.95) aus der absoluten Summierbarkeit der w_h . Für die Summierbarkeitseigenschaft der w_h verweisen wir auf die dem Beweis folgende Bemerkung 12.23.

Es bleibt (12.89) zu zeigen. Da sowohl w_h als auch $\gamma(h)$ absolut summierbar sind, folgt zunächst aus (12.92) durch Grenzübergang $M \rightarrow \infty$

$$\begin{aligned} \tau_M^2 &\rightarrow (\eta - 3) \left(\sum_{h=0}^{\infty} w_h \gamma(h) \right)^2 \\ &+ \sum_{r=-\infty}^{\infty} \sum_{h,k=0}^{\infty} w_h w_k \left\{ \gamma(r) \gamma(r+k-h) + \gamma(r+k) \gamma(r-h) \right\}. \end{aligned} \quad (12.93)$$

Aufgrund der Darstellung (12.84) der Spektraldichte f aus der Inversionsformel gilt nun einerseits

$$\begin{aligned} \int_0^\pi W(\omega) f(\omega) d\omega &= \int_0^\pi W(\omega) \frac{1}{2\pi} \left\{ \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(h\omega) \right\} d\omega \\ &= \sum_{h=0}^{\infty} w_h \gamma(h). \end{aligned} \quad (12.94)$$

Andererseits erhalten wir wegen $W(\omega) = 2 \cdot \sum_{h=0}^{\infty} w_h \cos(h\omega)$ (man beachte hierfür ebenfalls die nachfolgende Bemerkung 12.23)

$$\begin{aligned} &2\pi \int_0^\pi W^2(\omega) f^2(\omega) d\omega \\ &= 4 \cdot \sum_{h,k=0}^{\infty} w_h w_k \int_0^\pi \cos(h\omega) \cos(k\omega) \left(\sum_{r=-\infty}^{\infty} \gamma(r) \cos(r\omega) \right) f(\omega) d\omega. \end{aligned} \quad (12.95)$$

Ein Additionstheorem für trigonometrische Funktionen lautet

$$\begin{aligned} &\cos(a) \cos(b) \cos(c) \\ &= \frac{1}{4} (\cos(a+b-c) + \cos(b+c-a) + \cos(c+a-b) + \cos(a+b+c)). \end{aligned}$$

Deshalb können wir (12.95) wegen $\gamma(h) = 2 \int_0^\pi \cos(h\omega) f(\omega) d\omega$, vgl. (3.17), wie folgt fortsetzen

$$\begin{aligned} &2\pi \int_0^\pi W^2(\omega) f^2(\omega) d\omega \\ &= 4 \sum_{h,k=0}^{\infty} w_h w_k \sum_{r=-\infty}^{\infty} \gamma(r) \int_0^\pi \cos(h\omega) \cos(k\omega) \cos(r\omega) f(\omega) d\omega \\ &= \frac{1}{2} \cdot \sum_{h,k=0}^{\infty} w_h w_k \sum_{r=-\infty}^{\infty} \left\{ \gamma(r) \gamma(r+k-h) + \gamma(r) \gamma(k+h-r) \right. \\ &\quad \left. + \gamma(r) \gamma(h+r-k) + \gamma(r) \gamma(r+k+h) \right\} \end{aligned} \quad (12.96)$$

$$\begin{aligned}
&= \sum_{h,k=0}^{\infty} w_h w_k \sum_{r=-\infty}^{\infty} \left\{ \gamma(r) \gamma(r+k-h) + \gamma(r) \gamma(r+k+h) \right\} \\
&\quad (\text{Ersetzung des Summationsindexes } r \text{ durch } r+k-h \\
&\quad \text{im dritten und } r+k+h \text{ im zweiten Summanden}) \\
&= \sum_{h,k=0}^{\infty} w_h w_k \sum_{r=-\infty}^{\infty} \left\{ \gamma(r) \gamma(r+k-h) + \gamma(r+k) \gamma(r-h) \right\} \\
&\quad (\text{Summationsverschiebung im zweiten Summanden}).
\end{aligned}$$

Durch Zusammenfügen von (12.93) - (12.96) erhalten wir mit τ^2 aus (12.83) also die Aussage (12.89) und damit die Behauptung. \square

Bemerkung 12.23

Die im Beweis von Satz 12.22 definierten Größen w_h für $h \in \mathbb{N}_0$, vergleiche (12.87), sind nichts anderes als die Fourierkoeffizienten der Funktion W . W können wir durch Spiegelung an der y -Achse (d.h. $W(-\omega) = W(\omega)$) zunächst symmetrisch auf $[-\pi, \pi]$ ausdehnen und anschließend periodisch auf \mathbb{R} fortsetzen. Es gilt $\frac{1}{\pi} \int_{-\pi}^{\pi} W(\omega) d\omega = 4w_0$,

$$\frac{1}{\pi} \int_{-\pi}^{\pi} W(\omega) \cos(h\omega) d\omega = 2w_h, \quad h = 1, 2, \dots, \quad (12.97)$$

sowie $\int_{-\pi}^{\pi} W(\omega) \sin(h\omega) d\omega = 0$ ($h = 1, 2, \dots$), aufgrund der Symmetrie von W . Für stetig differenzierbares W konvergiert die Fourier-Reihe

$$2w_0 + 2 \sum_{h=1}^{\infty} w_h \cos(h\omega)$$

nach Heuser (1981), Satz 136.5, absolut und damit gleichmäßig gegen W . Also gelten

$$2 \cdot \sum_{h=0}^{\infty} w_h \cos(h\omega) = W(\omega), \quad \omega \in [-\pi, \pi] \quad (12.98)$$

und

$$\sum_{h=0}^{\infty} |w_h| < \infty \quad (12.99)$$

Beide Aussagen wurden im Beweis von Satz 12.22 verwendet. \square

An dieser Stelle wollen wir auf das integrierte Periodogramm als Spektral-dichteschätzer zurückkommen, welches wir schon in Abschnitt 12.5 behandelt haben. Mit Hilfe von Satz 12.22 können wir die asymptotische Verteilung exakt angeben und damit die in Abschnitt 12.5 lediglich heuristisch hergeleiteten Ergebnisse untermauern.

Bemerkung 12.24 (Das integrierte Periodogr. als Spektraldichteschätzer)

Für

$$W(\omega) = K_h(\lambda - \omega) = \frac{1}{h} K\left(\frac{\lambda - \omega}{h}\right), \quad (12.100)$$

wobei $K : [-\pi, \pi] \rightarrow [0, \infty)$ eine stetig differenzierbare und symmetrische Wahrscheinlichkeitsdichte, $h > 0$ eine Bandweite und $\lambda \in (0, \pi)$ eine feste Frequenz bezeichne, ist W eine (in Abhängigkeit von h) mehr oder weniger stark um die gegebene Frequenz λ konzentrierte Funktion. Im Sinne von Beispiel 12.21 stellt

$$\begin{aligned} \tilde{f}_n(\lambda) &= \int_0^\pi K_h(\lambda - \omega) I_n(\omega) d\omega \\ &\approx \frac{2\pi}{n} \sum_{j=0}^N K_h(\lambda - \omega_j) I_n(\omega_j), \quad \omega_j = 2\pi j/n, j = 0, \dots, N = [n/2] \end{aligned}$$

also einen Spektraldichteschätzer dar. Dabei ist die Summe nichts anderes als eine Riemann-Approximation an das Integral mit den Fourier-Frequenzen $\omega_j = 2\pi j/n$ als Stützstellen. Für eine feste Bandweite h kann die asymptotische Verteilung von $\tilde{f}_n(\lambda)$ für lineare Zeitreihen mit Hilfe von Satz 12.22 bestimmt werden. Mit (12.100) gilt

$$\begin{aligned} &\sqrt{n} (M(I_n, W) - M(I_n, f)) \\ &= \sqrt{n} \left(\tilde{f}_n(\lambda) - \int_0^\pi K_h(\lambda - \omega) f(\omega) d\omega \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_h^2(\lambda)), \end{aligned}$$

mit

$$\tau_h^2(\lambda) = 2\pi \int_0^\pi K_h^2(\lambda - \omega) f^2(\omega) d\omega + (\eta - 3) \left(\int_0^\pi K_h(\lambda - \omega) f(\omega) d\omega \right)^2.$$

Obwohl wir den Fall $h = h(n) \rightarrow 0$ für $n \rightarrow \infty$ hier nicht behandeln, wollen wir im Folgenden den Varianzausdruck $\tau_h^2(\lambda)$ für kleine Bandweiten h approximieren. Mit Hilfe der Substitutionsregel für Integrale und entsprechenden Taylorentwicklungen erhalten wir für zweimal stetig differenzierbare Spektraldichten f sowohl

$$\begin{aligned} \int_0^\pi K_h^2(\lambda - \omega) f^2(\omega) d\omega &\approx \int_{(\lambda-\pi)/h}^{\lambda/h} h^{-1} K^2(u) \underbrace{f^2(\lambda - hu)}_{\approx f^2(\lambda) + \mathcal{O}(h)} du \\ &\approx h^{-1} f^2(\lambda) \int_{-\pi}^\pi K^2(u) du + \mathcal{O}(1) \quad (12.101) \end{aligned}$$

als auch

$$\begin{aligned}
& \int_0^\pi K_h(\lambda - \omega) f(\omega) d\omega \\
& \approx \int_{(\lambda-\pi)/h}^{\lambda/h} h^{-1} K(u) \underbrace{f(\lambda - hu)}_{\approx f(\lambda) - hu f'(\lambda) + \mathcal{O}(h^2)} du \\
& \approx f(\lambda) \int_{-\pi}^\pi K(u) du + \mathcal{O}(h^2)
\end{aligned}$$

man beachte $\int uK(u) du = 0$ wegen der Symmetrie von K
 $\approx f(\lambda) + \mathcal{O}(h^2)$ (K ist Wahrscheinlichkeitsdichte).

Die Approximationen mit den unterschiedlichen Abhängigkeiten von h für die beiden Varianzanteile legen für $\lambda \in (0, \pi)$ die folgende Asymptotik

$$\sqrt{nh} \left(\tilde{f}_n(\lambda) - f(\lambda) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_0^2(\lambda)) \quad (12.102)$$

mit $h = h(n) \rightarrow 0, nh(n) \rightarrow \infty$ für $n \rightarrow \infty$ und

$$\tau_0^2(\lambda) = 2\pi f^2(\lambda) \int_{-\pi}^\pi K^2(u) du.$$

nahe.

Die Überschlagsrechnungen sind im Einklang mit den heuristischen Varianzberechnungen für das geglättete Periodogramm in Abschnitt 12.5. Vergleiche insbesondere mit (12.69). \square

Die Form der Varianz der asymptotischen Verteilung für integrierte Periodogrammschätzer $M(I_n, W)$ in (12.83) wird vereinfacht, wenn $\int_0^\pi W(\omega) f(\omega) d\omega$ verschwindet. Insbesondere hängt in diesem Fall die resultierende asymptotische Varianz von $M(I_n, W)$ mit $2\pi \int_0^\pi W^2(\omega) f^2(\omega) d\omega$ nicht von dem vierten Moment Ee_1^4 der Innovationen der $MA(\infty)$ -Darstellung der Zeitreihe ab. Zu einer derartigen Situation gelangen wir, wenn wir wie in Janas/Dahlhaus (1986) vorgestellt, sogenannte Quotientenstatistiken

$$R(I_N, W) = \frac{\int_0^\pi W(\omega) I_n(\omega) d\omega}{\int_0^\pi I_n(\omega) d\omega} \quad (12.103)$$

betrachten. Im Spezialfall $W(\omega) = 2 \cos(\omega h)$ (vgl. Beispiel 12.21 (i)) führt die Quotientenstatistik (12.103) zu $M(I_n, W)/M(I_n, 1) = \bar{\gamma}_n(h)/\bar{\gamma}_n(0) =: \bar{\varrho}_n(h)$. In den Teilen (ii) und (iii) des Beispiels 12.21 führt die Division durch das integrierte Periodogramm dazu, dass die Schätzer für die Spektralverteilungsfunktion bzw. -dichte auf Wahrscheinlichkeitsverteilungsfunktionen bzw. -dichten auf $[0, \pi]$ normiert werden.

Wir erhalten für Quotientenstatistiken das folgende Korollar zu Satz 12.22

Korollar 12.25 (*Asymptotische Normalität für Quotientenstatistiken*)

Unter den Voraussetzungen von Satz 12.22 gilt

$$\sqrt{n}(R(I_n, W) - R(f, W)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tau_r^2) \quad (12.104)$$

mit der asymptotischen Varianz

$$\tau_r^2 = 2\pi \frac{\int_0^\pi \widetilde{W}^2(\omega) f(\omega) d\omega}{\left(\int_0^\pi f(\omega) d\omega\right)^4}, \quad (12.105)$$

wobei $\widetilde{W}(\omega) = W(\omega) \int_0^\pi f(\lambda) d\lambda - \int_0^\pi W(\lambda) f(\lambda) d\lambda$ gesetzt werde.

Beweis: Es gilt

$$\begin{aligned} & \sqrt{n}(R(I_n, W) - R(f, W)) \\ &= \frac{\sqrt{n}}{M(I_n, 1)M(f, 1)} \left(\int_0^\pi W(\omega) I_n(\omega) d\omega \cdot M(f, 1) - \int_0^\pi I_n(\omega) d\omega \cdot M(f, W) \right) \\ &= \frac{\sqrt{n}}{M(I_n, 1)M(f, 1)} \int_0^\pi \left\{ W(\omega) M(f, 1) - M(f, W) \right\} I_n(\omega) d\omega. \end{aligned}$$

Mit $\widetilde{W} = W \cdot M(f, 1) - M(f, W)$ gilt $M(f, \widetilde{W}) = \int_0^\pi \widetilde{W}(\omega) f(\omega) d\omega = 0$. Also folgt aus Satz 12.22

$$\sqrt{n} \int_0^\pi \widetilde{W}(\omega) I_n(\omega) d\omega \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, 2\pi \int_0^\pi \widetilde{W}^2(\omega) f^2(\omega) d\omega \right).$$

Speziell impliziert (12.82) natürlich die Konsistenz von $M(I_n, W)$, also auch $M(I_n, 1) \rightarrow M(f, 1)$ n.W. . Das Lemma von Slutsky (vgl. Bemerkung A.26 (ii)) ergibt nun die Behauptung. \square

Aufgaben

Aufgabe 12.1 Seien $x_1, \dots, x_n \in \mathbb{R}$. Für ein weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $e_t \sim (0, \frac{1}{n})$ und $b_i := x_{i+1}$, $i = 0, \dots, n-1$, bilde man die Zeitreihe $Y_t := b_0 e_t + b_1 e_{t-1} + \dots + b_{n-1} e_{t-n+1}$, $\forall t \in \mathbb{Z}$, und zeige für ihre Autokovarianzfunktion $\gamma_Y(h)$ die Gleichheit

$$\gamma_Y(h) = \sum_{i=1}^{n-|h|} x_i x_{i+|h|} \quad \text{für } |h| \leq n-1$$

und $\gamma_Y(h) = 0$ für $|h| \geq n$. Als Folgerung erhält man einen anderen Beweis für die positive Semidefinitheit der Autokovarianzschätzer $\widehat{\gamma}_n$ und $\bar{\gamma}_n$ (zur Definition von $\widehat{\gamma}_n$ und $\bar{\gamma}_n$ siehe (12.14) bzw. (12.15)).

Aufgabe 12.2 Seien $x_1, \dots, x_n \in \mathbb{R}$.

- a) Man zeige durch direkte Rechnung, dass für $\bar{\gamma}_n(k) := \frac{1}{n} \sum_{t=1}^{n-k} x_t x_{t+k}$, $k = 0, \dots, n-1$, $\bar{\gamma}_n(-k) := \bar{\gamma}_n(k)$, die Ungleichung

$$f_n(\omega) := \frac{1}{2\pi} \sum_{h=-(n-1)}^{n-1} e^{-ih\omega} \bar{\gamma}_n(h) \geq 0$$

gilt.

- b) Man leite aus a) die Beziehung

$$\int_{-\pi}^{\pi} f_n(\omega) d\omega = \bar{\gamma}_n(0) < \infty,$$

ab, sowie

- c)

$$\int_{-\pi}^{\pi} e^{ik\omega} f_n(\omega) d\omega = \begin{cases} \bar{\gamma}_n(k), & |k| < n, \\ 0, & |k| \geq n. \end{cases}$$

Nach dem Herglotz-Lemma ist also $\bar{\gamma}_n$ positiv semidefinit mit Spektraldichte f_n . Dies liefert wiederum einen Beweis für die positive Semidefinitheit von $\hat{\gamma}_n$ aus (12.14) und $\bar{\gamma}_n$ aus (12.15).

Aufgabe 12.3 Seien X_1, \dots, X_n i.i.d. $\sim \mathcal{N}(0, \sigma^2)$, mit $n = 2m + 1$, $m \in \mathbb{N}$. Für $A(\omega) := (2/n)^{1/2} \sum_{t=1}^n X_t \cos(\omega t)$ und $B(\omega) := (2/n)^{1/2} \sum_{t=1}^n X_t \sin(\omega t)$ zeige man mit $\omega_k := 2\pi k/n$

$$(A(\omega_1), \dots, A(\omega_m), B(\omega_1), \dots, B(\omega_m)) \sim \mathcal{N}(\mathbf{0}, \sigma^2 E_{2m}),$$

wobei E_k die $k \times k$ -Einheitsmatrix bezeichne.

Aufgabe 12.4 Man beweise Lemma 12.5.

Aufgabe 12.5 Man beweise die Beziehungen (12.31) und (12.32).

Aufgabe 12.6 Man beweise Aussage (12.37).

Aufgabe 12.7 Man zeige für den zentrierten Autokovarianzschätzer $\hat{\gamma}_n$ aus Gleichung (12.14) die Aussage

$$\sum_{|h| < n} \hat{\gamma}_n(h) = 0.$$

Aufgabe 12.8 Man beweise das *Frequency Domain Sampling Theorem*, d.h. für

$$Z^{(n)}(\omega) := \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{-i\omega t}, \quad \forall \omega \in [-\pi, \pi],$$

und $Z_k^{(n)} := Z^{(n)}(\omega_k)$ mit den Fourierfrequenzen $\omega_k = 2\pi k/n$ gilt die Gleichheit

$$Z^{(n)}(\omega) = \sum_{k=-[(n-1)/2]}^{[n/2]} Z_k^{(n)} e^{-i(\omega-\omega_k)(n+1)/2} \frac{\sin(n(\omega-\omega_k)/2)}{\sin((\omega-\omega_k)/2)}.$$

Alle Eigenschaften der Funktion $Z^{(n)}$ können also aus denen der endlich vielen Zufallsvariablen $Z_k^{(n)}$ hergeleitet werden.

Aufgabe 12.9 Seien $x_1, \dots, x_n \in \mathbb{R}$ und

$$Z^{(n)}(\omega) := \frac{1}{\sqrt{2\pi n}} \sum_{\kappa=1}^n x_\kappa e^{-i\omega \kappa}, \quad \forall \omega \in [-\pi, \pi].$$

Man zeige die Gleichheit

$$x_t = \sqrt{\frac{n}{2\pi}} \int_{-\pi}^{\pi} Z^{(n)}(\omega) e^{it\omega} d\omega, \quad t = 1, \dots, n.$$

Aufgabe 12.10 Seien $x_1, \dots, x_n \in \mathbb{R}$. Man zeige für $\bar{\gamma}_n(k) = \frac{1}{n} \sum_{t=1}^{n-k} x_t x_{t+k}$, $k = 0, \dots, n-1$, und $I_n(\omega) := |Z^{(n)}(\omega)|^2$ (siehe vorherige Aufgabe) die Gleichheit ($r > 2n-1$; $\tilde{\omega}_h := 2\pi h/r$)

$$\bar{\gamma}_n(k) = \frac{2\pi}{r} \sum_{h=-[(r-1)/2]}^{[r/2]} e^{i\tilde{\omega}_h k} I_n(\tilde{\omega}_h), \quad k = 0, \dots, n-1.$$

Aufgabe 12.11 Seien X_1, \dots, X_n reelle Zufallsvariablen, $n = 2m+1$, $m \in \mathbb{N}$. Mit den Bezeichnungen aus (12.3) zeige man

$$\frac{4\pi}{n} \sum_{k=1}^m I_{n, \bar{X}}(\omega_k) = \frac{1}{n} \sum_{t=1}^n (X_t - \bar{X}_n)^2, \quad \omega_k := 2\pi k/n.$$

Aufgabe 12.12 Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, stationäre, zentrierte Zeitreihe mit Spektraldichte f .

a) Man zeige für das Periodogramm I_n die Gleichung

$$EI_n(\omega) = \int_{[-\pi, \pi]} k_n(x - \omega) f(x) dx, \quad \forall \omega \in [-\pi, \pi]. \quad (12.106)$$

Dabei bezeichnet k_n den *Fejér-Kern*

$$k_n(u) := \frac{\sin^2(u \cdot n/2)}{2\pi n \sin^2(u/2)}. \quad (12.107)$$

b) Man zeige

$$\int_{[-\pi, \pi]} k_n(x - \omega) dx = 1 \quad \forall \omega \in \mathbb{R}. \quad (12.108)$$

c) Sei g Lebesgue-integrierbar auf $[-\pi, \pi]$ und stetig im Punkt $\omega \in (-\pi, \pi]$.
Man zeige

$$\int_{[-\pi, \pi]} g k_n(\cdot - \omega) d\lambda \rightarrow g(\omega) \quad \text{für } n \rightarrow \infty. \quad (12.109)$$

Aufgabe 12.13 Für eine reelle, zentrierte, stationäre Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit Spektraldichte f zeige man für das zentrierte Periodogramm (vgl. (12.3))

$$EI_{n, \bar{X}}(\omega) = \frac{1}{2\pi n} \int_{[-\pi, \pi]} \left\{ \frac{\sin(\frac{1}{2}(x - \omega)n)}{\sin(\frac{1}{2}(x - \omega))} - \frac{\sin(\frac{1}{2}\omega n)}{n \sin(\frac{1}{2}\omega)} \frac{\sin(\frac{1}{2}n)}{\sin(\frac{1}{2}x)} \right\}^2 f(x) dx.$$

Aufgabe 12.14 Unter den Voraussetzungen der vorherigen Aufgabe und der zusätzlichen Annahme, dass f stetig in 0 und ω ist, zeige man

$$\begin{aligned} & n(EI_{n, \bar{X}}(\omega) - EI_n(\omega)) \\ &= \int_{[-\pi, \pi]} \left[-2 \frac{\sin(\omega n/2)}{\sin(\omega/2)} k_n(x - \omega, x) + \frac{1}{n} \frac{\sin^2(\omega n/2)}{\sin^2(\omega/2)} k_n(x) \right] f(x) dx \end{aligned}$$

und beweise, dass dieser Ausdruck für $n \rightarrow \infty$ gegen Null konvergiert für alle $\omega \neq 0$. Dabei sei k_n der Fejér-Kern aus Aufgabe 12.12 und $k_n(\lambda, \mu)$ gegeben durch

$$k_n(\lambda, \nu) := \frac{\sin(\lambda n/2) \sin(\nu n/2)}{2\pi n \sin(\lambda/2) \sin(\nu/2)}.$$

Aufgabe 12.15 Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle Zeitreihe in $MA(\infty)$ -Darstellung $X_t = \sum_{j=-\infty}^{+\infty} b_j e_{t-j}$, $\forall t \in \mathbb{Z}$, für ein i.i.d. weißes Rauschen $e = (e_t : t \in \mathbb{Z})$ mit $Ee_t^4 < \infty$ und absolut summierbaren Koeffizienten $b_j, j \in \mathbb{Z}$. Man zeige, dass es eine von h und n unabhängige Konstante c gibt, sodass für alle $n \in \mathbb{N}$ und $h \in \mathbb{N}_0$ gilt

$$\text{Var}\left(\frac{1}{n} \sum_{t=1}^{n-h} X_t X_{t+h}\right) \leq \frac{c}{n}.$$

Aufgabe 12.16 Sei $X = (X_t : t \in \mathbb{Z})$ eine reelle, stationäre, zentrierte Zeitreihe mit absolut summierbarer Autokovarianzfunktion γ und Spektraldichte f .

a) Man zeige für $\tilde{\gamma}_n(h) = \frac{1}{n} \sum_{t=1}^n X_t X_{t+h}$ und den Dichteschätzer

$$\tilde{f}_n(\omega) = \frac{1}{2\pi} \sum_{k=-M}^M e^{-ik\omega} \tilde{\gamma}_n(|k|) \quad (12.110)$$

die Aussage

$$\tilde{f}_n(\omega) \rightarrow f(\omega), \quad n.W. \quad \forall \omega \in [-\pi, \pi], \quad (12.111)$$

falls die folgenden Voraussetzungen gelten: Es gibt eine Konstante $c > 0$ mit

$$\text{Var} \tilde{\gamma}_n(h) \leq \frac{c}{n}, \quad \forall n, h, \quad (12.112)$$

sowie

$$M = M(n) \rightarrow \infty \text{ und } M(n)^2/n \rightarrow 0 \text{ für } n \rightarrow \infty. \quad (12.113)$$

b) Man zeige, dass dann auch für $\hat{\gamma}_n$ mit einer (12.110) entsprechenden Definition gilt

$$E\hat{f}_n(\omega) \rightarrow f(\omega), \quad \forall \omega \in [-\pi, \pi], \quad (12.114)$$

und

$$\hat{f}_n(\omega) \rightarrow f(\omega), \quad n.W. \quad \forall \omega \in [-\pi, \pi]. \quad (12.115)$$

Aufgabe 12.17 Sei $f : [-1, 1] \rightarrow \mathbb{R}$ stetig. Für $d \notin 2\pi\mathbb{Z}$ zeige man

$$\frac{1}{M} \sum_{k=-M}^M f\left(\frac{k}{M}\right) e^{idk} \rightarrow 0 \quad \text{für } M \rightarrow \infty.$$

Hinweis: Für Indikatorfunktionen $f = 1_{(a,b]}$ ergibt sich die Aussage aus (A.6). Sie gilt damit für primitive Funktionen. Schließlich approximiere man eine stetige Funktion f gleichmäßig durch primitive Funktionen.

Aufgabe 12.18 Seien e_t , $t = 1, \dots, n$, stochastisch unabhängige reelle Zufallsvariable mit $e_t \sim (0, \sigma_e^2)$ und $Ee_t^4 =: \eta\sigma_e^4 < \infty \quad \forall t$. Man zeige für $0 \leq \omega \leq \pi$

$$E \left| \sum_{t=1}^n e_t e^{i\omega t} \right|^4 \leq n(\eta - 3)\sigma_e^4 + n^2 3\sigma_e^4.$$

Aufgabe 12.19 Man zeige, dass für lineare Zeitreihen $X = (X_t : t \in \mathbb{Z})$ mit i.i.d. weißem Rauschen $e = (e_t : t \in \mathbb{Z})$ für das $Ee_1^4 < \infty$ erfüllt ist, eine Konstante $C > 0$ existiert, so dass für alle n, h gilt

$$n \cdot E (\bar{\gamma}_n(h) - \gamma(h))^2 \leq C.$$

Aufgabe 12.20 (Whittle-Schätzer)

$f_{ARMA}(\omega; \mathbf{a}, \mathbf{b},)$ bezeichne die parametrische Form der Spektraldichte einer $ARMA$ -Zeitreihe. Mit dem Periodogramm I_n führt eine Minimierung des Integrals

$$\int_{-\pi}^{\pi} \frac{I_n(\omega)}{f_{ARMA}(\omega; \mathbf{a}, \mathbf{b})} d\omega$$

in den Parametern \mathbf{a} und \mathbf{b} zum sogenannten *Whittle-Schätzer*.

Man zeige, dass für den Fall einer reinen AR -Zeitreihe dieses Vorgehen, nämlich eine Minimierung von

$$\int_{-\pi}^{\pi} \frac{I_n(\omega)}{|1 + \sum_{j=1}^p a_j e^{-ij\omega}|^2} d\omega,$$

exakt zu den *Yule-Walker-Parameterschätzern* führt.

Modellierung mit ARMA-Zeitreihen

Bisher haben wir uns im Rahmen der Parameterschätzung von *ARMA*-Zeitreihen auf den Standpunkt gestellt, dass die zugehörigen Ordnungen p und q bekannt sind. Darüberhinaus haben wir unterstellt, dass die zur Verfügung stehenden Daten X_1, \dots, X_n tatsächlich von einem *ARMA*(p, q)-Modell generiert wurden. Beide Annahmen sind natürlich für die tatsächliche Anwendung unrealistisch. Grundsätzlich sollte man sich bei statistischen Test- und Schätzverfahren immer fragen, wie sich die unter einer gewissen Modellvorstellung entwickelten Methoden verhalten, wenn das unterstellte Modell nicht vorliegt. Einen gewissen Schritt in diese Richtung haben wir mit Satz 11.3 getan, in dem wir das asymptotische Verhalten des unter der Annahme einer *AR*(p)-Zeitreihe abgeleiteten Yule-Walker-Parameterschätzers in der Modellklasse der *MA*(∞)-Zeitreihen abgeleitet haben. Eine weitergehende Untersuchung in dieser Richtung soll im Rahmen dieser Einführung allerdings unterbleiben.

Vielmehr wollen wir uns in diesem Kapitel mit der Frage auseinander setzen, wie wir auf der Basis realer Daten ein adäquates *ARMA*-Modell auswählen können. Erste Schritte bei der Untersuchung von Zeitreihendaten sind sicherlich eine visuelle Betrachtung und die etwaige Bereinigung von Saison- und/oder Trendkomponenten. Hierfür haben wir in Kapitel 1 einige Methoden vorgestellt. Durch derartige Saison- und/oder Trendbereinigungen soll erreicht werden, dass wir zumindest approximativ von stationär erzeugten Daten ausgehen können, wenn wir damit beginnen *ARMA*-Modelle anzupassen.

In Beispiel 6.7 haben wir darüberhinaus die Differenzenbildung (als einen speziellen Filter) für Zeitreihen kennen gelernt. Wie dort ausgeführt, versucht man mit dieser Differenzenbildung ebenfalls nichtstationäre Reihen in stationäre zu überführen. Verbunden mit der *ARMA*-Modellklasse führt die Differenzenbildung zu sogenannten *ARIMA*-Zeitreihen, die wir in Abschnitt 13.1 vorstellen werden. Danach gehen wir im Abschnitt 13.2 auf verschiedene Methoden zur Wahl adäquater Ordnungen p und q von *ARMA*-Zeitreihen ein.

Neben eher graphischen und teilweise auch heuristischen Methoden stehen sogenannte Ordnungswahlverfahren im Vordergrund. Zu nennen sind hier die auf Akaike zurückgehenden Methoden des *Final Prediction Errors (FPE)* und das *Akaike Information Criterion (AIC)*.

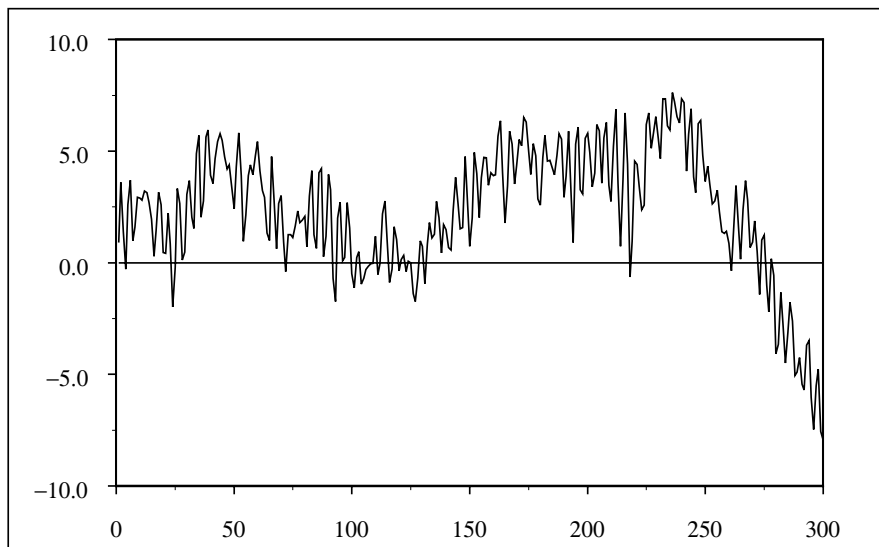


Abb. 13.1 Simulation einer $ARIMA(2, 1, 0)$ -Zeitreihe (vgl. Beispiel 13.2)

13.1 ARIMA-Zeitserien

Eine in der Praxis gängige Klasse zur Modellierung auch nichtstationärer Zeitreihen stellen die im Folgenden definierten *ARIMA* (autoregressive integrated moving average)-Zeitreihen dar.

Definition 13.1 (*ARIMA* (p, d, q) -Zeitreihe)

Für gegebene $p, d, q \in \mathbb{N}_0$ heißt $(Y_t : t \in \mathbb{Z})$ *ARIMA* (p, d, q) -Zeitreihe, falls $X_t := (1 - L)^d Y_t, t \in \mathbb{Z}$, eine kausale und stationäre *ARMA* (p, q) -Zeitreihe darstellt. L ist dabei der Lag-Operator.

Zur Verdeutlichung der Definition betrachten wir das folgende Beispiel.

Beispiel 13.2 (*ARIMA* $(p, 1, q)$ -Zeitreihe)

Ausgehend von einer stationären und kausalen *ARMA* (p, q) -Zeitreihe (X_t) erfüllt die *ARIMA* $(p, 1, q)$ -Zeitreihe (Y_t) die Beziehung

$$X_t = (1 - L)Y_t = Y_t - Y_{t-1}, \quad t \in \mathbb{Z}.$$

Durch rekursives Einsetzen erhalten wir

$$Y_t = Y_{t-1} + X_t = Y_0 + \sum_{k=1}^t X_k. \quad (13.1)$$

In Abbildung 13.1 ist für den Spezialfall $p = 2, q = 0$ sowie $a_1 = 0, a_2 = 0.8$ und $Y_0 = 0$ eine Realisierung einer $ARIMA(2, 1, 0)$ -Zeitreihe der Länge $n = 300$ zu sehen. Deutlich zu erkennen ist das nichtstationäre Verhalten. \square

Unter Verwendung der Darstellung einer $ARMA(p, q)$ -Zeitreihe X_t in Kompaktform (vgl. Bemerkung 7.3)

$$A(L)X_t = B(L)e_t \quad (13.2)$$

mit $A(L) = \sum_{j=0}^p a_j L^j$ sowie $B(L) = \sum_{j=0}^q b_j L^j$, ergibt sich ganz entsprechend für $ARIMA(p, d, q)$ -Zeitreihen (Y_t) die Darstellung

$$A(L)(1 - L)^d Y_t = B(L)e_t, \quad t \in \mathbb{Z}. \quad (13.3)$$

Dies ist die kompakte Darstellung einer $ARMA(p+d, q)$ -Gleichung, wobei das autoregressive Polynom $A(z)(1 - z)^d$ eine d -fache Nullstelle in $z = 1$ besitzt.

Bereits aus Beispiel 6.7 ist bekannt, dass die Anwendung des Operators $(1 - L)^d$ auf ein Polynom $(d - 1)$ -ten Grades mit zufälligen oder konstanten Koeffizienten stets die Nullfunktion liefert. D.h. aber, dass wir zu einer $ARIMA(p, d, q)$ -Zeitreihe (Y_t) stets ein solches Polynom $(d - 1)$ -ten Grades addieren können, ohne den nach d -facher Differenzenbildung entstehenden $ARMA(p, q)$ -Prozeß zu verändern. Damit ist klar, dass die in Definition 13.1 eingeführten $ARIMA$ -Zeitreihen keineswegs eindeutig bestimmt sind. Diese Tatsache tritt in (13.1) durch die vollständige Freiheit bei der Wahl von Y_0 zu Tage.

Wir schließen die kurze Einführung in die Klasse der $ARIMA$ -Zeitreihen ab und wenden uns im nächsten Abschnitt dem Problem der Ordnungswahl in $ARMA$ -Zeitreihen zu.

13.2 Ordnungswahl in ARMA-Zeitreihen

Zunächst wollen wir einige visuelle Möglichkeiten zur Wahl geeigneter Ordnungen in speziellen $ARMA$ -Modellen vorstellen. Aus Beispiel 2.6 wissen wir, dass die Autokovarianz- und damit auch die Autokorrelationsfunktion für eine $MA(q)$ -Zeitreihe für Lags $h > q$ gleich Null ist. Diese Tatsache ermöglicht es uns, anhand der empirischen Autokorrelationsfunktion $\hat{\varrho}(h)$ Informationen über die Ordnungen q einer zugrunde liegenden moving average Zeitreihe zu gewinnen.

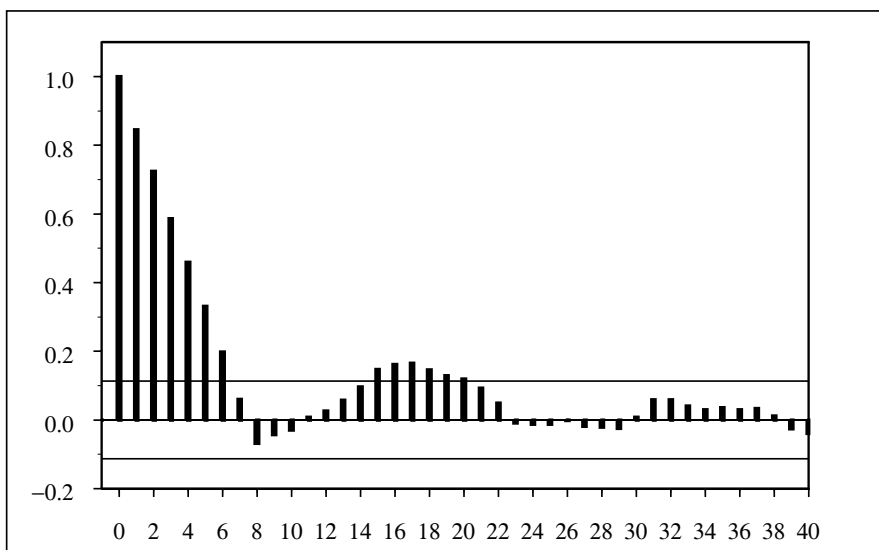


Abb. 13.2 Empirisches Korrelogramm einer $MA(7)$ -Zeitreihe gemäß Beispiel 13.3 mit Länge $n = 300$ und Konfidenzschranken $\pm 1.96\sqrt{n}$

Beispiel 13.3 (Visuelle Ordnungswahl bei MA -Zeitreihen)

In Abbildung 13.2 ist für eine $MA(7)$ -Zeitreihe mit $b_1 = \dots = b_7 = 1$ und $\sigma_e^2 = 1$ aufgrund von $n = 300$ simulierten Beobachtungen ein empirisches Korrelogramm gegeben. Mit Hilfe von 10.100) ergibt sich die Varianz der asymptotischen Normalverteilung von $\sqrt{n} \cdot \hat{\varrho}(h)$ für $h > 7$ zu $43/8$. Wir erwarten also für größere Stichprobenumfänge im besagten $MA(7)$ -Modell, dass $\hat{\varrho}(h)$ für $h > 7$ mit approximativer Wahrscheinlichkeit von 95% im Konfidenzintervall $\pm 1.96 \cdot \sqrt{43/(8n)}$ liegt (beachte $z_{0.975} = 1.96$). Hieraus leitet man ab, dass als Schätzer für die Ordnung q der MA -Zeitreihe (unter der Annahme, dass überhaupt eine moving average Zeitreihe zugrunde liegt) das größte Lag h zu verwenden ist, für das $\hat{\varrho}(h)$ nicht im oben genannten asymptotischen 95%-Konfidenzintervall liegt. Da die für die Bestimmung des Konfidenzintervalls benötigten Autokorrelationen aber in aller Regel nicht bekannt sind, verwendet man in der tatsächlichen Anwendung häufig das kürzere Intervall $\pm 1.96/\sqrt{n}$. Dieses engere Intervall wird in der Regel dazu führen, dass nach dem gerade beschriebenen Vorgehen tendenziell eine größere Modellordnung gewählt wird. Falls die zugrunde liegende Zeitreihe ein unabhängiges weißes Rauschen darstellt, so sind die engeren Konfidenzschranken asymptotisch exakt.

In Abbildung 13.2 sind diese schärfere Konfidenzschranken eingezeichnet. Man würde auf die beschriebene Weise für die der Abbildung 13.2 zugrunde liegen-

den Daten vermutlich den Ordnungsschätzer $\hat{q} = 6$ wählen.

□

Bei autoregressiven Zeitreihen können wir ähnlich vorgehen, wobei statt der Autokorrelation die partielle Autokorrelation zu verwenden ist.

Beispiel 13.4 (Visuelle Ordnungswahl bei AR-Zeitreihen)

(X_t) sei eine kausale $AR(p)$ -Zeitreihe. Dann gilt für die zugehörige partielle Autokorrelationsfunktion π gemäß Aufgabe 7.7 $\pi(h) = 0$ für $h > p$. Ebenso wie im Beispiel 13.3 für moving average Zeitreihen können wir nun anhand der empirischen partiellen Autokorrelationsfunktion $\hat{\pi}(h)$ (vgl. (2.40)) einen Schätzer für die Ordnung der angenommenen AR-Zeitreihe ableiten. Satz 11.4 und die anschließende Bemerkung 11.6 besagen, dass $\sqrt{n}\hat{\pi}(h)$ für $h > p$ eine asymptotische Standardnormalverteilung besitzt. Deshalb ist es gerechtfertigt, \hat{p} als größtes Lag h zu wählen, für welches $\hat{\pi}(h)$ nicht im asymptotischen Konfidenzintervall $\pm 1.96/\sqrt{n}$ liegt. Für $n = 200$ simulierte Werte der $AR(2)$ -Zeitreihe (X_t) gemäß $X_t + 0.8X_{t-2} = e_t, t \in \mathbb{Z}$, führt dieses Vorgehen zur Ordnungswahl $\hat{p} = 2$ (vgl. Abbildung 13.3). □

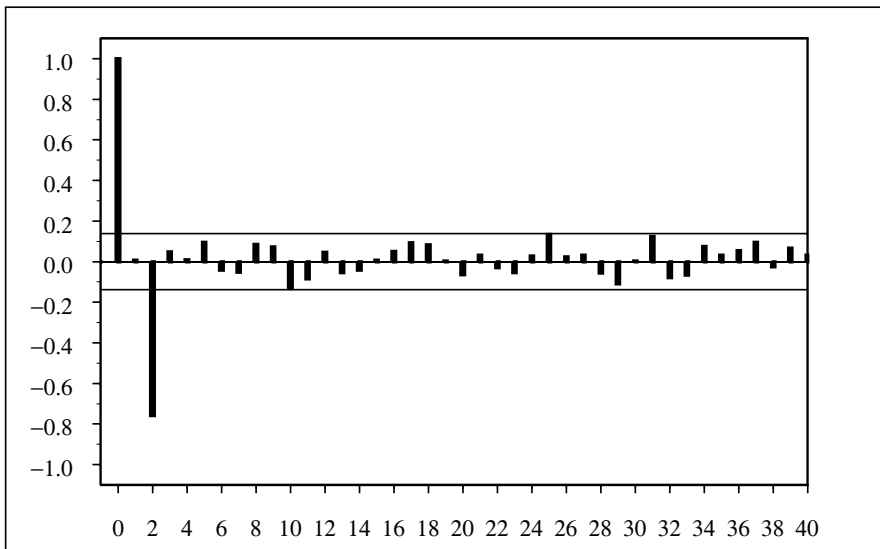


Abb. 13.3 Empirisches partielles Autokorrelogramm einer simulierten $AR(2)$ -Zeitreihe der Länge $n = 200$ gemäß Beispiel 13.4 mit den asymptotischen 95%-Konfidenzschranken $\pm 1.96/\sqrt{n}$

Für allgemeine $ARMA(p, q)$ -Zeitreihen ist die Situation komplizierter. Hier wissen wir aufgrund von Bemerkung 8.1 nur, dass die Autokorrelation für

wachsendes Lag h geometrisch abklingt. Auf dieser Basis ist mit Hilfe der jeweiligen empirischen Größen kaum eine adäquate Ordnungswahl möglich.

Einen ganz anderen Ansatz für die Wahl adäquater Modellordnungen erlaubt die Betrachtung der Spektraldichte f einer kausalen $ARMA(p, q)$ -Zeitreihe (X_t) mit

$$X_t + a_1 X_{t-1} + \dots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \dots + b_q e_{t-q}, t \in \mathbb{Z}. \quad (13.4)$$

Die zugehörige Spektraldichte f besitzt gemäß (7.15) die Darstellung

$$\begin{aligned} f(\omega) &= \frac{\sigma_e^2}{2\pi} \frac{|1 + b_1 e^{-i\omega} + \dots + b_q e^{-iq\omega}|^2}{|1 + a_1 e^{-i\omega} + \dots + a_p e^{-ip\omega}|^2} \\ &= \frac{\sigma_e^2}{2\pi} \frac{|B(e^{-i\omega})|^2}{|A(e^{-i\omega})|^2}, \quad \omega \in [-\pi, \pi]. \end{aligned} \quad (13.5)$$

Da wir uns in diesem Abschnitt auf reelle Zeitreihen beschränken ist f symmetrisch um Null und wir betrachten f somit lediglich auf $[0, \pi]$. Im Falle der Invertibilität transformiert eine Anwendung des Filters $A(e^{-i\omega})/B(e^{-i\omega})$ auf die $ARMA$ -Zeitreihe diese bekanntlich in das zugehörige weiße Rauschen mit konstanter Spektraldichte.

Durch Ausprobieren können wir nun für eine Reihe von Ordnungen p' und q' die Koeffizienten der zugehörigen $ARMA(p, q)$ -Modelle schätzen und – solange die geschätzten Modelle invertibel sind – die Filter

$$\frac{1 + \hat{a}_1(p')e^{-i\omega} + \dots + \hat{a}_{p'}(p')e^{-ip'\omega}}{1 + \hat{b}_1(q')e^{-i\omega} + \dots + \hat{b}_1(q')e^{-iq'\omega}} \quad (13.6)$$

auf unsere Daten anwenden. Auf die gefilterte Zeitreihe

$$\hat{e}_t(p', q'), t = 1 + \max(p', q'), \dots, n,$$

(die sog. Residuen) wenden wir dann einen konsistenten Spektraldichteschätzer gemäß Kapitel 12 an und wählen schlussendlich diejenigen Ordnungen \hat{p} und \hat{q} , für die der Spektraldichteschätzer der zugehörigen Residuen die visuell beste Approximation an eine gewünschte konstante Funktion auf $[0, \pi]$ liefert.

Grundsätzlich spricht für ein solches Vorgehen, dass beliebige stetige Spektraldichten auf $[0, \pi]$ sowohl durch Spektraldichten eines kausalen AR -Modells als auch eines invertiblen MA -Modells (und somit auch eines kausalen und invertiblen $ARMA$ -Modells) beliebig genau approximiert werden können (vgl. Abschnitt 9.2). Deshalb können wir selbst bei Daten, die nicht von einem $ARMA$ -Modell generiert wurden, auf die beschriebene Art und Weise vorgehen, solange nur eine stetige Spektraldichte existiert. Vergleiche hierzu auch die Überlegungen in Abschnitt 12.1 über die parametrische Spektraldichteschätzung mittels Anpassung von $ARMA$ -Modellen.

Im Folgenden wollen wir uns eher quantitativen Methoden der Ordnungswahl zuwenden. Wir beginnen erneut mit den einfacher zugänglichen autoregressiven Zeitreihen. Auf Akaike (1969) geht die folgende, *Final Prediction Error* (FPE) genannte, Methode zurück. Ausgehend von Beobachtungen X_1, \dots, X_n einer $AR(p_0)$ -Zeitreihe bestimmen wir für verschiedene $p \in \{1, \dots, P\}$ (wobei P typischerweise mit n wächst) Parameterschätzer $\hat{a}_1(p), \dots, \hat{a}_p(p)$, $p = 1, \dots, P$, für jeweils entsprechende $AR(p)$ -Modelle. Dies könnte etwa über die Yule-Walker-Methode (vgl. (11.3)) oder über die LS-Methode (vgl. (11.24), (11.25) und (11.12)) geschehen. Die Yule-Walker-Methode hat den Vorteil, dass sie uns stets auf kausale geschätzte autoregressive Modelle führt.

Würden wir p nun derart wählen, dass

$$\hat{\sigma}_e^2(p) := \frac{1}{n} \sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p \hat{a}_j(p) X_{t-j} \right)^2 \quad (13.7)$$

minimal wird, so würde das Ergebnis in aller Regel $\hat{p} = P$ lauten, da eine höhere Modellordnung grundsätzlich eine bessere Approximation erlaubt. Wählt man für jedes feste p die Schätzer $\hat{a}_j(p)$, $j = 1, \dots, p$, gerade so, dass die rechte Seite von (13.7) minimal wird (LS-Methode), so ist die Gültigkeit der Aussage $\hat{p} = P$ offensichtlich. Akaike (1969) hat nun vorgeschlagen, gedanklich eine identische aber unabhängige Kopie (Y_t) des zugrunde liegenden $AR(p_0)$ -Modells anzunehmen. Eine optimale Ordnung im Sinne des final prediction error (FPE) ist dann als Minimalstelle (bzgl. p) von

$$E_Y \left(Y_t + \sum_{j=1}^p \hat{a}_j(p) Y_{t-p} \right)^2 \quad (13.8)$$

erklärt. Dabei bedeutet $E_Y(\cdot)$ eine Erwartungsbildung bezüglich (Y_t) bei festgehaltenen X_1, \dots, X_n (also eine bedingte Erwartungswertbildung). Der Unterschied zu (13.7) ist, dass in (13.8) eine unabhängige Kopie (Y_t) der Ausgangszeitreihe möglichst gut mittels des geschätzten $AR(p)$ -Modells prognostiziert werden soll und nicht die zugrunde liegende Zeitreihe, die zur Parameterschätzung verwendet wurde, selbst. Dadurch wird ein „overfitting“ an den einen vorliegenden Datensatz durch eine zu große Modellordnung vermieden und ein Ausgleich zwischen der Komplexität des Modells (eine höhere Ordnung liefert theoretisch immer bessere Anpassung) und des stochastischen Fehlers bei der Parameterschätzung (der natürlich mit wachsender Ordnung immer größer wird) geschaffen.

Zunächst können wir (13.8) unter Verwendung von $\mathbf{a}(p) = (a_1(p), \dots, a_p(p))^T$ aus Lemma 11.13 wie folgt vereinfachen

$$\begin{aligned}
& E_Y \left(Y_t + \sum_{j=1}^p \hat{a}_j(p) Y_{t-j} \right)^2 \\
&= E_Y \left(Y_t + \sum_{j=1}^p a_j(p) Y_{t-j} + \sum_{j=1}^p (\hat{a}_j(p) - a_j(p)) Y_{t-j} \right)^2 \\
&= E_Y \left(Y_t + \sum_{j=1}^p a_j(p) Y_{t-j} \right)^2 \\
&\quad + (\hat{\mathbf{a}}(p) - \mathbf{a}(p))^T \left(\begin{matrix} E Y_{t-j} Y_{t-k} \\ j, k = 1, \dots, p \end{matrix} \right) (\hat{\mathbf{a}}(p) - \mathbf{a}(p)), \quad (13.9)
\end{aligned}$$

mit $\hat{\mathbf{a}}(p) = (\hat{a}_1(p), \dots, \hat{a}_p(p))^T$. Die Begründung für das letzte Gleichheitszeichen, also das Verschwinden der gemischten Terme, liegt in der Definition der $a_j(p)$, $j = 1, \dots, p$. Es gilt nämlich für $h = 1, \dots, p$ (vgl. Lemma 11.13)

$$\begin{aligned}
& E_Y \left(\left(Y_t + \sum_{j=1}^p a_j(p) Y_{t-j} \right) Y_{t-h} \right) \\
&= \gamma(h) + \sum_{j=1}^p a_j(p) \gamma(h-j) = 0.
\end{aligned}$$

Beide in (13.9) auftretenden Ausdrücke müssen aus den vorliegenden Beobachtungen X_1, \dots, X_n geschätzt werden. Den Ausdruck

$$\sigma^2(p) := E_Y \left(Y_t + \sum_{j=1}^p a_j(p) Y_{t-j} \right)^2 = E \left(X_t + \sum_{j=1}^p a_j(p) X_{t-j} \right)^2$$

wird man zunächst durch eine Stichprobenversion, also etwa durch

$$\tilde{\sigma}^2(p) := \frac{1}{n} \sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p a_j(p) X_{t-j} \right)^2$$

zu approximieren versuchen. Der Versuch, die $a_j(p)$ durch entsprechende Schätzer $\hat{a}_j(p)$, $j = 1, \dots, p$, zu ersetzen, die sich durch Minimierung von (13.7) ergeben (*LS-Schätzer*), führt dann zu der folgenden Rechnung

$$\begin{aligned}
\tilde{\sigma}^2(p) &= \frac{1}{n} \sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p \hat{a}_j(p) X_{t-j} - \sum_{j=1}^p (\hat{a}_j(p) - a_j(p)) X_{t-j} \right)^2 \\
&= \frac{1}{n} \sum_{t=p+1}^n \underbrace{\left(X_t + \sum_{j=1}^p \hat{a}_j(p) X_{t-j} \right)^2}_{=\hat{\sigma}_e^2(p)} \\
&\quad + (\hat{\mathbf{a}}(p) - \mathbf{a}(p))^T \left(\frac{1}{n} \sum_{t=p+1}^n \begin{matrix} X_{t-j} X_{t-k} \\ j, k = 1, \dots, p \end{matrix} \right) (\hat{\mathbf{a}}(p) - \mathbf{a}(p)), \quad (13.10)
\end{aligned}$$

denn wieder fallen die gemischten Terme weg, da wir die $\hat{a}_j(p)$, $j = 1, \dots, p$, gerade als Minimalstellen von $\hat{\sigma}_e^2(p)$ definiert haben und diese erfüllen für $k = 1, \dots, p$

$$\sum_{t=p+1}^n \left(X_t + \sum_{j=1}^p \hat{a}_j(p) X_{t-j} \right) X_{t-k} = 0 \quad (13.11)$$

$$\iff \sum_{t=p+1}^n X_t X_{t-k} + \left(\sum_{t=p+1}^n X_{t-j} X_{t-k} : j = 1, \dots, p \right) \hat{\mathbf{a}}(p) = 0.$$

Nun ist aber der zweite Summand in (13.10) nichts anderes als eine Art

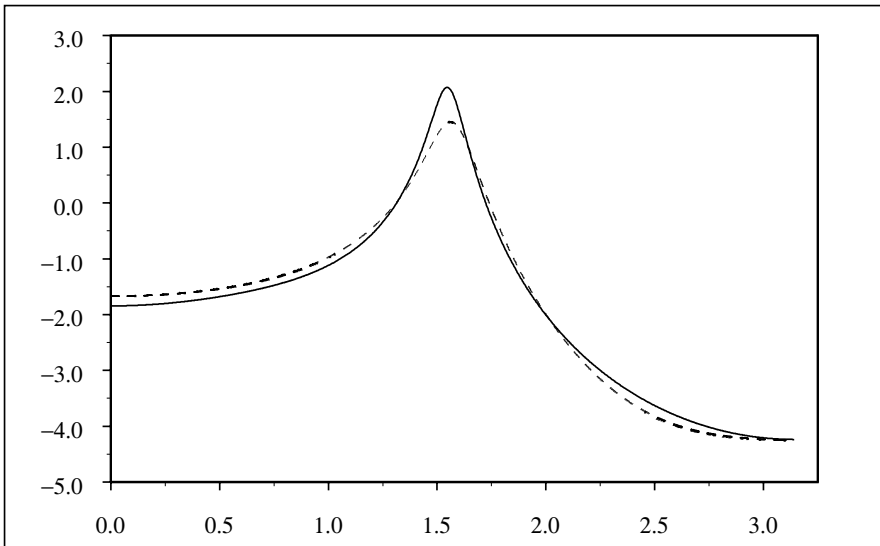


Abb. 13.4 Tatsächliche logarithmierte Spektraldichte des $ARMA(2,2)$ -Modells (13.13) (gestrichelt) und die logarithmierte Spektraldichte eines auf Basis eines simulierten Datensatzes der Länge $n = 400$ angepassten $AR(4)$ -Modells (vgl. Beispiel 13.5)

Stichprobenversion des zweiten Summanden in (13.9), den wir ja auch noch zu schätzen haben. Hierzu gehen wir lediglich heuristisch vor. Falls das zugrunde liegende Modell tatsächlich eine $AR(p)$ -Zeitreihe mit Parametern a_1, \dots, a_p ist, so gilt $\mathbf{a}(p) = (a_1, \dots, a_p)^T$ und wir wissen aufgrund von Satz 11.4 das $\sqrt{n}(\hat{\mathbf{a}}(p) - \mathbf{a}(p))$ asymptotisch $\mathcal{N}(\mathbf{0}, \sigma_e^2 \Gamma(p)^{-1})$ -verteilt ist. $\sqrt{n} \Gamma(p)^{1/2} (\hat{\mathbf{a}}(p) - \mathbf{a}(p)) / \sigma_e$ besitzt folglich eine asymptotische p -variante Standardnormalverteilung. Bezeichnet nun Z eine Zufallsvariable mit p -variaten Standardnormalverteilung, so approximieren wir den in Frage stehenden Ausdruck (vgl. (13.9))

$$(\hat{\mathbf{a}}(p) - \mathbf{a}(p))^T \Gamma(p) (\hat{\mathbf{a}}(p) - \mathbf{a}(p)) = \left\| \Gamma(p)^{1/2} (\hat{\mathbf{a}}(p) - \mathbf{a}(p)) \right\|^2$$

und die zugehörige Stichprobenversion aus (13.10) jeweils durch

$$\frac{\sigma_e^2}{n} E \|Z\|^2 = \frac{\sigma_e^2}{n} E Z^T Z.$$

Wegen $E Z^T Z = p$ und mit Ersetzung von σ_e^2 durch $\hat{\sigma}_e^2(p)$ gelangen wir schließlich auf die folgende berechenbare Approximation von (13.8)

$$\hat{\sigma}_e^2(p) \left(1 + \frac{2p}{n} \right) =: FPE(p). \quad (13.12)$$

Eine Minimierung von $FPE(p)$ in p führt auf die Ordnungswahl gemäß der Methode des Final Prediction Error.

Natürlich können wir das FPE -Kriterium zur Auswahl eines geeigneten $AR(p)$ -Modells auch anwenden, wenn die zugrunde liegende Zeitreihe nicht selbst von autoregressiver Gestalt ist. Für die Klasse der $AR(\infty)$ -Zeitreihen, zu der alle invertiblen $ARMA$ -Prozesse gehören, hat Shibata (1980) eine gewisse Optimalität des FPE -Kriteriums nachgewiesen. In der $AR(\infty)$ -Situation ist es so, dass mit wachsendem Stichprobenumfang immer größere Modellordnungen ausgewählt werden. Theoretisch beschreiben natürlich größere Modellordnungen das zugrunde liegende $AR(\infty)$ -Modell besser. Dem entgegen steht allerdings die Tatsache, dass die Schätzung einer großen Zahl von Parametern eine wachsende stochastische Fluktuation bedeutet. Durch den sogenannten Penalty-Term $1 + 2p/n$ in (13.12) versucht das FPE -Kriterium hier einen Ausgleich zwischen diesen beiden konkurrierenden Anforderungen zu erreichen. Selbst wenn der zugrunde liegende Prozess eine autoregressive Zeitreihe mit endlicher Ordnung p_0 ist, ist es nicht zwingend, dass auch für jeden Datensatz genau die tatsächliche Modellordnung gewählt wird. Wir werden dies in Beispiel 13.6 für ein modifiziertes Ordnungswahlkriterium noch näher betrachten.

Beispiel 13.5 (Ordnungswahl gemäß FPE)

In diesem Beispiel wenden wir die FPE -Methodik an, um an einen simulierten Datensatz der Länge $n = 400$ ein autoregressives Zeitreihenmodell anzupassen. Die Daten entstammen einer autoregressiven Zeitreihe unendlicher Ordnung, gegeben durch das folgende invertible $ARMA(2, 2)$ -Schema

$$X_t + 0.8X_{t-2} = e_t + 0.71e_{t-1} + 0.25e_{t-2}, \quad t \in \mathbb{Z}. \quad (13.13)$$

In Tabelle 13.1 haben wir für verschiedene Ordnungen p sowohl die Ausdrücke $\hat{\sigma}_e^2(p)$ (vgl. (13.7)) als auch $FPE(p)$ (vgl. (13.12)) berechnet. Man erkennt, dass $\hat{\sigma}_e^2(p)$ in der Tat in p fällt, $FPE(p)$ dagegen ein Minimum in $p_{FPE} = 4$ aufweist. Nach der Methode des Final Prediction Error würden wir also ein $AR(4)$ -Modell zur Beschreibung der Daten auswählen. Die zugehörigen Para-

Tabelle 13.1 Simulierte Werte von $\hat{\sigma}_e^2(p)$ und $FPE(p)$ für einen Datensatz der Länge $n = 400$ des $ARMA(2, 2)$ –Modells aus Beispiel 13.5

p	1	2	3	4	5	6
$\hat{\sigma}_e^2(p)$	3.864	1.227	0.959	0.941	0.941	0.940
$FPE(p)$	3.883	1.239	0.974	0.960	0.964	0.968

meterschätzer $\hat{a}_j(p), j = 1, \dots, p, p = 1, \dots, 6$, (vgl. (13.11)) sind in Tabelle 13.2 angegeben.

Um die Güte der Anpassung beurteilen zu können, vergleichen wir in Abbildung 13.4 die Spektraldichte der Ausgangszeitreihe (13.13) mit derjenigen des geschätzten $AR(4)$ –Modells (mit Parametern gemäß Tabelle 13.2). Man erkennt, dass das $AR(4)$ –Modell den zugrunde liegenden $ARMA(2, 2)$ –Prozess in diesem Sinne recht gut beschreibt. \square

Tabelle 13.2 Geschätzte Parameter für verschiedene AR –Modelle, die an einen simulierten Datensatz der Länge $n = 400$ des $ARMA(2, 2)$ –Prozesses aus Beispiel 13.5 angepasst wurden

p	$\hat{a}_j(p), j = 1, \dots, p$					
1	−0.118					
2	−0.214	+0.822				
3	−0.581	+0.918	−0.446			
4	−0.638	+1.034	−0.520	+0.127		
5	−0.640	−1.044	−0.540	+0.139	−0.019	
6	−0.640	−1.039	−0.520	+0.100	+0.005	−0.038

Ein sehr verbreitetes weiteres Modellwahlkriterium ist das sogenannte *Akaike Information Criterion (AIC)*, vergleiche Akaike (1973) und Akaike (1974). Das *AIC* läßt sich unabhängig vom Zeitreihenkontext formulieren und basiert auf einer empirischen Minimierung der *Kullback–Leibler–Information*. Bevor wir kurz auf eine wiederum nur heuristische Ableitung des *AIC* eingehen, soll das resultierende Kriterium für $ARMA$ –Modelle genannt werden. Dazu bezeichne $L_n(\mathbf{a}, \mathbf{b}, \sigma_e^2)$ die Likelihood–Funktion auf Basis von Beobachtungen X_1, \dots, X_n einer kausalen und normalverteilten $ARMA(p, q)$ –Zeitreihe mit Parametern $\mathbf{a} = (a_1, \dots, a_p)^T$ und $\mathbf{b} = (b_1, \dots, b_q)^T$ sowie Varianz σ_e^2 des weißen Rauschens (vgl. (11.71)).

Mit den Maximum–Likelihood–Schätzern $\hat{\mathbf{a}}, \hat{\mathbf{b}}$ und $\hat{\sigma}_e^2$ (vgl. mit Abschnitt

11.4) lautet das AIC -Kriterium

$$AIC(p, q) = -\frac{2}{n} \ln L_n(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\sigma}_\epsilon^2) + 2 \frac{p+q+1}{n}. \quad (13.14)$$

Tabelle 13.3 Simulierte Werte von $AIC(p, 0)$ für verschiedene an einen Datensatz der Länge $n = 400$ des $ARMA(2, 2)$ -Modells aus Beispiel 13.5 angepasste $AR(p)$ -Modelle

p	1	2	3	4	5	6	7
$AIC(p, 0)$	4.199	3.063	2.824	2.810	2.814	2.818	2.815

Tabelle 13.4 Simulierte Werte von $AIC(p, q)$ für verschiedene an einen Datensatz der Länge $n = 400$ des $ARMA(2, 2)$ -Modells aus Beispiel 13.5 angepasste $ARMA(p, q)$ -Modelle

(p, q)	(1, 1)	(2, 1)	(1, 2)	(2, 2)	(3, 2)	(2, 3)	(3, 3)
$AIC(p, q)$	3.851	2.853	3.552	2.809	2.821	2.818	2.818

Zur Anpassung eines $ARMA$ -Modells an die Daten X_1, \dots, X_n werden nun die Ordnungen p und q gewählt, die den Ausdruck $AIC(p, q)$ minimieren. Für die gleichen Daten wie im Beispiel 13.5 sind in Tabelle 13.3 zunächst die Werte $AIC(p, 0)$ für verschiedene p gegeben und in Tabelle 13.4 für verschiedene p und q die Werte $AIC(p, q)$. Man erkennt, dass zur Beschreibung der simulierten Daten des $ARMA(2, 2)$ -Modells vom Umfang $n = 400$ laut Tabelle 13.3 (wie schon im Beispiel 13.5 auf Basis des FPE -Kriteriums) eine $AR(4)$ -Zeitreihe gewählt werden sollte, wenn man das AIC -Kriterium zugrunde legt. Minimiert man dagegen das AIC -Kriterium in der Klasse der $ARMA(p, q)$ -Modelle, so zeigt Tabelle 13.4, dass dann die tatsächlichen Modellordnungen $p = q = 2$ identifiziert werden. Die zum $ARMA(2, 2)$ -Modell gehörenden Schätzwerte der Parameter sind in Tabelle 13.5 wiedergegeben.

Auch hier vergleichen wir die tatsächliche mit der über das $ARMA(2, 2)$ -Modell (d.h. mit Parametern gemäß Tabelle 13.5) geschätzten Spektraldichte (vgl. Abbildung 13.5). Eine entsprechende Grafik für das geschätzte $AR(4)$ -Modell ist bereits in Abbildung 13.4 zu finden.

Brockwell/Davis (1991) geben darüber hinaus ein auf Hurvich/Tsai (1989) zurückgehendes korrigiertes AIC -Kriterium (das sog. $AICC$ -Kriterium) an, was sich durch einen veränderten Penalty-Term (d.h. durch die additive Konstante zur Likelihood-Funktion) von (13.14) unterscheidet. Es lautet

Tabelle 13.5 Geschätzte Parameter für das $ARMA(2,2)$ -Modell aus Beispiel 13.5 aufgrund eines simulierten Datensatzes der Länge $n = 400$

$\hat{a}_1(2)$	$\hat{a}_2(2)$	$\hat{b}_1(2)$	$\hat{b}_2(2)$
-0.07	+0.85	+0.61	+0.17

$$AICC(p, q) = -\frac{2}{n} \ln L_n(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\sigma}^2) + 2 \frac{p + q + 1}{n - p - q - 2} \quad (13.15)$$

Der Penalty-Term $2(p + q + 1)/(n - p - q - 2)$ beim $AICC$, der zur Likelihood-Funktion addiert wird, ist größer als der entsprechende Term $2(p + q + 1)/n$ im AIC -Kriterium. Dies führt dazu, dass mit Hilfe des $AICC$ tendenziell kleinere Modellordnungen gewählt werden. Dies trägt der Tatsache Rechnung, dass das AIC (wie übrigens auch das FPE) typischerweise zu einer Überschätzung der tatsächlichen Modellordnungen führt (wenn es denn überhaupt eine tatsächliche Modellordnung gibt). Wir wollen dies an einem kleinen Simulationsbeispiel belegen.

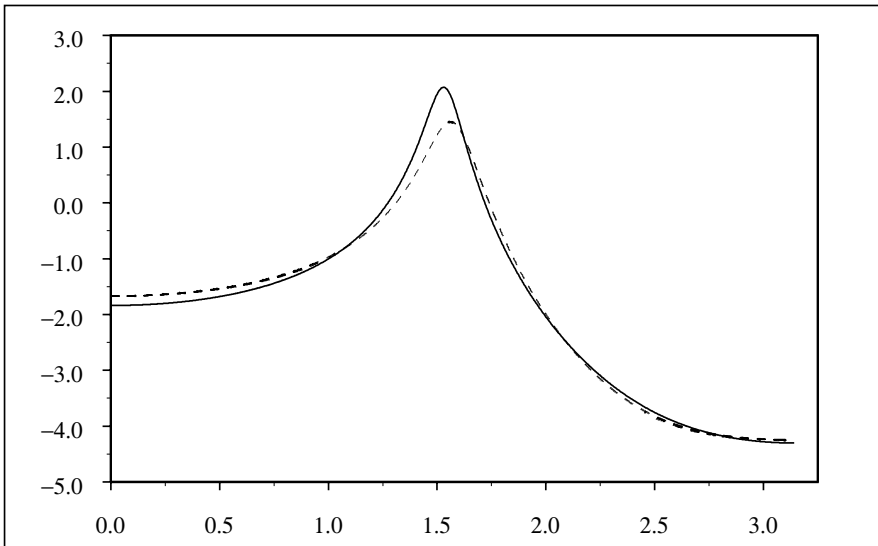


Abb. 13.5 Tatsächliche logarithmierte Spektraldichte des $ARMA(2,2)$ -Modells (13.13) (gestrichelt) und die logarithmierte Spektraldichte eines auf Basis eines simulierten Datensatzes der Länge $n = 400$ geschätzten $ARMA(2,2)$ -Modells

Beispiel 13.6 (Ordnungswahl mit dem *AIC*-Kriterium)

X_1, \dots, X_n seien Realisierungen der folgenden $AR(4)$ -Zeitreihe

$$X_t - 1.34X_{t-1} + 1.88X_{t-2} - 1.32X_{t-3} + 0.80X_{t-4} = e_t, \quad t \in \mathbb{Z},$$

mit standard normalverteiltem weißen Rauschen (e_t). Für jeweils 100 simulierte Datensätze der Länge n haben wir gemäß *AIC* die geschätzten Ordnungen bestimmt, um einen Eindruck von der Verteilung der Ordnungswahlschätzer zu gewinnen. Tabelle 13.6 gibt die Resultate für verschiedene Werte von n an. Sehr deutlich ist in Tabelle 13.6 die Tendenz zum Überschätzen der tatsächlichen Modellordnung $p_0 = 4$ zu erkennen.

Tabelle 13.6 Relative Häufigkeiten der gemäß *AIC* gewählten Modellordnung eines an simulierte Beobachtungen der $AR(4)$ -Zeitreihe aus Beispiel 13.6 der Länge $n = 50, 100, 200, 500$ und 1000 angepassten autoregressiven Modells (jeweils 100 Monte-Carlo Wiederholungen)

Ordnung p	≤ 2	3	4	5	6	≥ 7
$n = 50$	0	1	36	31	13	19
$n = 100$	0	0	41	38	8	13
$n = 200$	0	0	55	31	2	12
$n = 500$	0	0	78	11	3	8
$n = 1000$	0	0	88	6	2	4

Diese grundsätzliche Tendenz zum Überschätzen der tatsächlichen Modellordnungen wird in abgeschwächter Form allerdings auch vom *AICC*-Kriterium geteilt.

□

Akaike (1978) hat ein weiteres Ordnungwahlkriterium vorgeschlagen (das sogenannte *BIC*), dass sich vom *AIC* dahingehend unterscheidet, dass der additive Penalty-Term $2(p + q + 1)/n$ durch den mit n langsamer fallenden Ausdruck $\ln(n) \cdot (p + q + 1)/n$ ersetzt wird. Hannan (1980) konnte für tatsächlich zugrunde liegende $ARMA(p, q)$ -Zeitreihen zeigen, dass gemäß *BIC* gewählte Modellordnungen konsistente Schätzer für p und q darstellen.

Nun wollen wir die oben angekündigte, allerdings nur heuristische Begründung für das *AIC* nachreichen. Gehen wir dazu ganz allgemein von einer Klasse $\{f(\cdot; \vartheta) | \vartheta \in \Theta\}$ von Wahrscheinlichkeitsdichten auf \mathbb{R}^n aus, zu der die Verteilung des Datensatzes $\mathbf{X} = (X_1, \dots, X_n)$ gehören möge. ϑ_0 bezeichne den tatsächlichen Parameter der zugrunde liegenden Verteilung.

$$\begin{aligned}
I(\vartheta_0, \vartheta) &= 2E_{\vartheta_0} \ln \frac{f(\mathbf{X}; \vartheta_0)}{f(\mathbf{X}; \vartheta)} \\
&= 2E_{\vartheta_0} \ln f(\mathbf{X}; \vartheta_0) - 2E_{\vartheta_0} \ln f(\mathbf{X}; \vartheta)
\end{aligned} \tag{13.16}$$

heißt *Kullback–Leibler–Information* von ϑ (in Bezug auf ϑ_0). Mit Hilfe der Jensen–Ungleichung folgt wegen der Konvexität von $-\ln(x)$

$$\begin{aligned}
I(\vartheta_0, \vartheta) &= 2E_{\vartheta_0} \left(-\ln \frac{f(\mathbf{X}; \vartheta)}{f(\mathbf{X}; \vartheta_0)} \right) \\
&\geq -2 \ln E_{\vartheta_0} \left(\frac{f(\mathbf{X}; \vartheta)}{f(\mathbf{X}; \vartheta_0)} \right) \\
&= -2 \ln \int_{\mathbb{R}^n} \frac{f(\mathbf{X}; \vartheta)}{f(\mathbf{X}; \vartheta_0)} f(\mathbf{X}; \vartheta_0) d\mu(x) \\
&\geq -2 \ln 1 = 0.
\end{aligned}$$

Dabei gilt $I(\vartheta_0, \vartheta) = 0$ genau dann, wenn $f(\cdot; \vartheta) = f(\cdot; \vartheta_0)$ P_{ϑ_0} -fast sicher.

$I(\vartheta_0, \vartheta)/n$ wird als Maß für die Approximationsgüte der Dichte $f(\cdot; \vartheta)$ an die zugrunde liegende Verteilung angesehen. Die Idee des *AIC* ist nun, einen geeigneten Schätzer für $I(\vartheta_0, \vartheta)/n$ zu konstruieren, der nur von den zur Verfügung stehenden Daten abhängt, und ϑ dann als Minimalstelle dieses Schätzers zu wählen. Offensichtlich reicht es aufgrund von (13.16) das genannte Programm lediglich für einen Schätzer von

$$-\frac{2}{n} E_{\vartheta_0} \ln f(\mathbf{X}, \vartheta)$$

durchzuführen, da der verbleibende Ausdruck in (13.16) unabhängig von ϑ ist. Den Erwartungswert E_{ϑ_0} ersetzt man durch das Stichprobenmittel $(-2/n) \ln f(\mathbf{X}; \vartheta)$, was für unabhängige Beobachtungen tatsächlich die Form eines Mittels besitzt. Problematischer ist die Ersetzung von ϑ durch einen Schätzer $\hat{\vartheta}$, da hierdurch typischerweise eine wesentliche Verfälschung (ein sogenannter Bias–Term) in Kauf genommen werden muss. Akaike schlägt zur Korrektur dieses Bias–Terms den folgenden Ausdruck vor

$$2 \cdot (\text{Anzahl der Parameter}) / \text{Stichprobenumfang}$$

und erhält so das in (13.14) angegebene *AIC*–Kriterium.

Auf weiterführende Erläuterungen soll im Rahmen dieser Einführung verzichtet werden. Es sei lediglich darauf hingewiesen, dass wir im Rahmen der heuristischen Herleitung des *FPE*–Kriteriums (vgl. (13.10)) und nachfolgende Ausführungen) einer sehr ähnlichen Problematik begegnet sind.

Bevor wir diesen bewusst kurzen Einblick in die Welt der Modellwahl bei Zeitreihen beenden, soll noch darauf hingewiesen werden, dass (mit welcher Methode auch immer) gewählte Zeitreihenmodelle anhand der zugrunde liegenden Daten validiert werden sollten. Eine nahe liegende Idee ist, mit Hilfe des gewählten *ARMA*–Modells und der zur Verfügung stehenden Daten die

Terme e_t des zum geschätzten ARMA-Modell gehörenden weißen Rauschens zu approximieren. Diese geschätzten Innovationen \hat{e}_t , die sich über eine Filterung aus den Ausgangsdaten gewinnen lassen, sollte man dann zumindest visuell oder auch mit statistischen Methoden auf approximative Unabhängigkeit überprüfen.

13.3 Threshold Zeitreihenmodelle

Eine für die Anwendung wichtige Klasse nichtlinearer Zeitreihenmodelle wird durch die sogenannten *Threshold (Schwellen)-Modelle* gegeben. Die Grundidee ist, dass die Modellparameter zur Berechnung des folgenden Wertes abhängig von tatsächlichen Werten zurück liegender Beobachtungen sind. Man erreicht dadurch, dass der Verlauf der Zeitreihe beobachtungsabhängig von wechselnden Regimen gesteuert wird. Für autoregressive Zeitreihen lässt sich diese Idee am einfachsten darstellen und führt zu den *Threshold AR-* (kurz: *TAR-*) Zeitreihenmodellen. Seien dazu für ein $R \in \mathbb{N}$ Ordnungen p_r und Parametersätze $a_1(r), \dots, a_{p_r}(r), \sigma_r^2$ für $r = 1, \dots, R$ gegeben. Im einfachsten Fall erfüllt eine *TAR-Zeitreihe* (X_t) für ein Lag d , eine Zerlegung $-\infty = u_0 < u_1 < \dots < u_R = \infty$ von \mathbb{R} und ein weißes Rauschen (e_t) dann die folgenden Gleichungen

$$X_t + \sum_{j=1}^{p_r} a_j(r) X_{t-j} = e_t, \text{ falls } u_{r-1} < X_{t-d} \leq u_r. \quad (13.17)$$

In diesem Fall steuert also der zurückliegende Wert X_{t-d} die Auswahl eines entsprechenden autoregressiven Modells zur Berechnung des nächsten Wertes X_t . Durch diese beobachtungsabhängige Parameterwahl wird etwa die Modellierung von Sättigungsproblemen möglich. Wir gehen auf diese und weitere Modellklassen parametrischer nichtlinearer Zeitreihen nicht weiter ein. Den interessierten Leser verweisen wir auf Tong (1983) und Tong (1990), der insbesondere die Threshold Modelle eingeführt und untersucht hat.

Es gibt noch zahlreiche weitere Modellklassen, insbesondere für nichtlineare Zeitreihen, die wir in dieser Einführung nicht behandeln können. Dazu gehören sowohl bilineare als auch die für die ökonomischen Anwendungen sehr wichtigen sogenannten kointegrierten Zeitreihenmodelle.

Aufgaben

Aufgabe 13.1 Man betrachte das bei $t = 1$ startende $AR(1)$ -Modell

$$X_0 := 0 \text{ und } X_t = X_{t-1} + e_t, \quad t = 1, 2, \dots,$$

für i.i.d. Zufallsvariable $e_t \sim (0, \sigma_e^2)$ und berechne die Autokorrelationen für die Zeitreihe $U_t := X_t - X_{t-m}$, $t = m, 2m, \dots$, sowie mit $Y_t := (X_t + \dots +$

X_{t+m-1}) für $V_t := Y_t - Y_{t-m}$, $t = m, 2m, \dots$.

Dies einfache Beispiel zeigt, dass bei der Betrachtung von Differenzen (wie sie häufig bei Preiszeitreihen gebildet werden) durch die Bildung von Durchschnitts Korrelationen entstehen können, die in der Ausgangsreihe nicht vorhanden waren.

Aufgabe 13.2 Man betrachte das folgende Threshold $AR(1)$ -Modell

$$X_t = \begin{cases} a_1 X_{t-1} + e_t, & X_{t-1} \leq r \\ a_2 X_{t-1} + e_t, & X_{t-1} > r, \end{cases}$$

für bekanntes $r \in \mathbb{R}$ und a_1, a_2 mit $|a_1|, |a_2| < 1$.

Man leite mit Hilfe eines Kleinste-Quadrate-Ansatzes Schätzer für a_1 und a_2 her.

Aufgabe 13.3 Sei (X_t) eine $ARIMA(p, d, q)$ -Zeitreihe. Man zeige, dass die Zeitreihe $X_t + \sum_{j=0}^{d-1} A_j t^j$ mit beliebigen Zufallsvariablen A_0, A_1, \dots, A_{d-1} dieselbe $ARIMA$ -Gleichung wie (X_t) erfüllt.

Grundlagen finanzieller Zeitreihen

Seit Mitte der siebziger Jahre nimmt die Modellierung von Preisen von Finanzgütern eine wichtige Rolle im Bereich der Zeitreihenanalyse ein. Insbesondere für das Finanzrisikomanagement ist die adäquate Modellierung von diskret beobachteten Finanzdaten von großer Bedeutung. Da man den Realisierungen von vielen Finanzzeitreihen auf den ersten Blick ansieht, dass sie die Annahme der Stationarität in der Regel in keiner Weise erfüllen (vgl. etwa den zeitlichen Verlauf des Deutschen Aktienindex DAX in Abbildung 1.5 oder den des US-Dollar/DM-Wechselkurs in Abbildung 1.6), können stationäre Zeitreihenmodelle höchstens auf geeignet transformierte Beobachtungen zur Modellierung eingesetzt werden. Eine gängige Transformation für den Preis P_t , $t = 0, 1, 2, \dots$, eines bestimmten Finanzgutes (z.B. den Preis einer Aktie, ein Aktienindex oder ein Wechselkurs) stellen die sogenannten log-returns

$$X_t = \ln P_t - \ln P_{t-1}, \quad t = 1, 2, \dots, \quad (14.1)$$

dar. In erster Näherung beschreibt ein log-return gerade die relative Änderung

$$(P_t - P_{t-1})/P_{t-1}, \quad t = 1, 2, \dots, \quad (14.2)$$

des Preises vom Zeitpunkt $t - 1$ zu t . Dies sieht man mit Hilfe der Taylor-Approximation

$$\ln(1 + x) \approx x$$

für betragsmäßig kleine x sofort ein. Für den Deutschen Aktienindex (DAX) sind die zugehörigen log-returns in Abbildung 14.1 dargestellt (vgl. auch mit den Renditen des DAX aus Abbildung 1.7). Man erkennt, dass eine stationäre Modellierung hierfür schon eher in Frage kommt. Allerdings ist die Schwankungsbreite der log-returns des DAX deutlich zeitlich veränderlich. In den Anfängen der Modellierung von Finanzzeitreihen war es durchaus üblich anzunehmen, dass log-returns sich approximativ wie i.i.d. Zufallsvariable verhalten. Darüberhinaus wurde vielfach auch eine Normalverteilung angenommen.

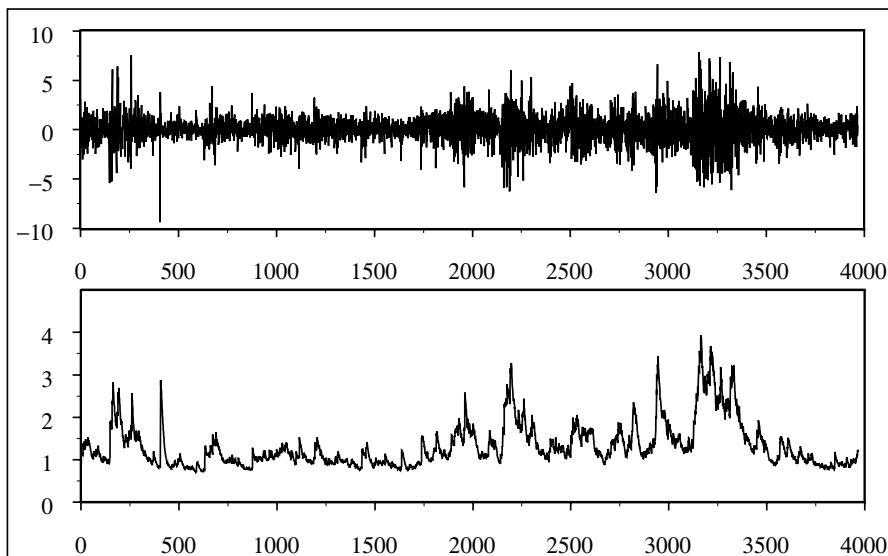


Abb. 14.1 Log-returns des DAX für den Zeitraum Januar 1990 bis September 2005 ($n = 3966$) (oben) zusammen mit der über ein $GARCH(1, 1)$ -Modell geschätzten Volatilität σ_t (unten)

Die 1997 durch die Verleihung des Preises der schwedischen Nationalbank zu Ehren Alfred Nobels gewürdigten Arbeiten von Black/Scholes (1973) und Merton (1973), die bahnbrechend für die Optionspreistheorie waren, nahmen eine sogenannte geometrische Brownsche Bewegung für das zugrunde liegende Finanzgut an. Ohne näher auf diese Theorie einzugehen sei hier lediglich erwähnt, dass diese Annahme zu einer normalverteilten i.i.d. Struktur für die log-returns des Finanzgutes führt und somit eine gewisse Rechtfertigung für die Annahme darstellt, dass log-returns in der Sprache der Zeitreihenanalyse nichts anderes sind als ein unabhängiges weißes Rauschen. Für den logarithmierten Preisprozess $\ln P_t$ führt eine solche i.i.d. Annahme für die log-returns (etwa $\ln P_t - \ln P_{t-1} = e_t$) zu einer klassischen random walk Struktur, genauer

$$\ln P_t = \ln P_0 + \sum_{j=1}^t e_j, \quad t = 1, 2, \dots \quad (14.3)$$

Zahlreiche empirische Studien über Finanzzeitreihen legen nun aber nachdrücklich nahe, dass die Unabhängigkeit von log-returns und auch die Normalverteilungsannahme kaum zu rechtfertigen sind. Ohne hier in die Tiefe zu gehen, betrachten wir lediglich ein empirisches Korrelogramm für die zentrierten log-returns, die Absolutbeträge und die quadrierten log-returns des DAX. Eine Unabhängigkeitsannahme für die log-returns allein würde sich direkt auf die Beträge und auch die Quadrate übertragen und müsste in allen Fällen

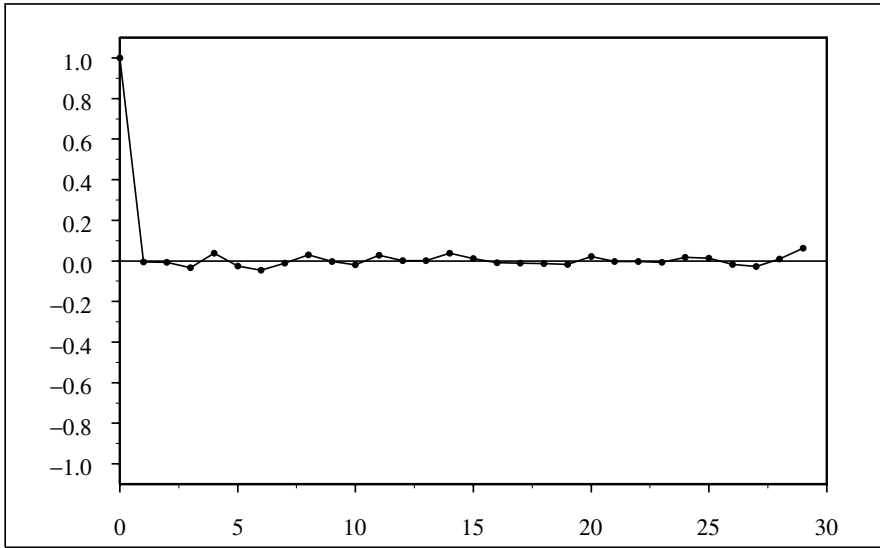


Abb. 14.2 Korrelogramm für die log-returns des DAX aus Abbildung 14.1

zu verschwindenden Autokorrelationen führen, was zwar für die log-returns selbst aber keineswegs für die Beträge und die Quadrate gerechtfertigt werden kann (vgl. Abbildung 14.2 und Abbildung 14.3). Diese Tatsache widerspricht der Modellierung von log-returns als weißem Rauschen zunächst nicht, allerdings einer möglichen *Unabhängigkeitsannahme*. Wenn nun aber log-returns als lediglich *unkorreliertes* aber nicht *unabhängiges* weißes Rauschen modelliert werden sollen, dann stellt dies die im Rahmen der Black-Scholes Theorie gemachte Normalverteilungsannahme in Frage, da für multivariate Normalverteilungen die Unabhängigkeit stets aus der Unkorreliertheit folgt.

Robert Engle (1982) hat mit der Klasse der autoregressiven bedingt heteroskedastischen Zeitreihenmodelle (autoregressive with conditional heteroscedasticity, kurz: *ARCH*-Modelle) eine Modellklasse vorgeschlagen, die trotz ihrer recht einfachen Struktur in der Lage ist, real auftretende Phänomene bei Finanzzeitreihen einigermaßen zufriedenstellend beschreiben zu können. Hierfür wurde Engle 2003 mit dem schon oben erwähnten Nobelpreis für Wirtschaftswissenschaften ausgezeichnet. Zusammen mit einer von Bollerslev (1986) vorgeschlagenen Verallgemeinerung (generalized *ARCH* oder kurz *GARCH*) hat die von Engle initiierte bedingt heteroskedastische Modellierung eine überragende Bedeutung in der Finanzzeitreihenanalyse erlangt. Aus und neben den *GARCH*-Modellen haben sich zahlreiche weitere bedeutende bedingt heteroskedastische Modellansätze entwickelt. Besonders genannt werden sollen die stochastischen Volatilitätsmodelle. Shephard (1996) gibt eine gute Übersicht über die fast unüberschaubare Literatur im Bereich der Modellierung von Finanzzeitreihen mit *ARCH*-Typ und stochastischen Vo-

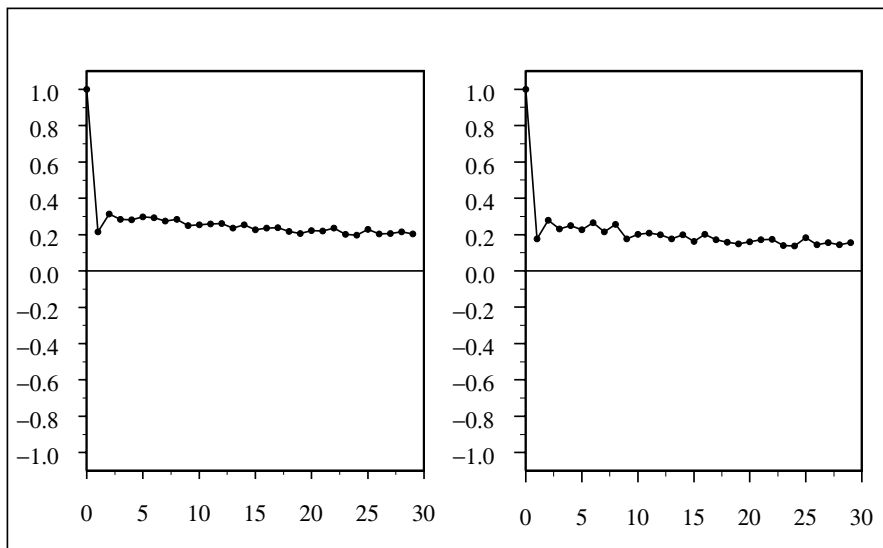


Abb. 14.3 Korrelogramm der Beträge (links) und der Quadrate (rechts) der log-returns des DAX aus Abbildung 14.1

latilitätsmodellen. Wir können im Rahmen dieser Einführung nur einen kleinen Ausschnitt dieses bedeutenden Anwendungsfeldes der Zeitreihenanalyse beleuchten und werden uns auf die *GARCH*-Modelle beschränken. In Abschnitt 14.1 untersuchen wir die grundsätzlichen wahrscheinlichkeitstheoretischen Eigenschaften des einfachsten *GARCH*-Modells erster Ordnung und gehen auch auf *GARCH*-Prozesse höherer Ordnung ein. Die Möglichkeiten der Parameterschätzung in *GARCH*-Modellen diskutieren wir in Abschnitt 14.2, bevor wir dieses Kapitel in Abschnitt 14.3 mit einer Anwendung der *GARCH*-Modellierung auf einen realen Datensatz (nämlich den Deutschen Aktienindex) beenden.

14.1 GARCH-Modelle

Zunächst wollen wir die Klasse der verallgemeinerten *ARCH*-Modelle definieren.

Definition 14.1 (*GARCH*(p, q)-Zeitreihe)

Eine reelle Zeitreihe heißt *GARCH*(p, q)-Zeitreihe, falls für i.i.d. Zufallsvariable ($e_t : t \in \mathbb{Z}$) mit $Ee_t = 0$ und $\text{Var } e_t = 1$, also $e_t \sim (0, 1)$, die Gleichung

$$X_t = \sigma_t \cdot e_t, \quad \forall t \in \mathbb{Z}, \quad (14.4)$$

erfüllt ist, wobei ($\sigma_t : t \in \mathbb{Z}$) eine nichtnegative Zeitreihe bezeichnet, die für reelle und nichtnegative Parameter a_0, a_1, \dots, a_p und b_1, \dots, b_q mit $a_p, b_q \neq 0$

der Rekursion

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2 \quad (14.5)$$

genügt. (σ_t) heißt *Volatilität der GARCH-Zeitreihe* (X_t) .

Im Falle $q = 0$ nennen wir $X = (X_t : t \in \mathbb{Z})$ eine *ARCH(p)-Zeitreihe*.

Bemerkung 14.2

Die Abkürzung *GARCH* steht wie schon oben erwähnt für *generalized autoregressive with conditional heteroscedasticity*. Falls die Zeitreihe (σ_t) streng stationär ist, so überträgt sich diese Eigenschaft wegen der i.i.d.-Struktur der Innovationen (e_t) auch auf die *GARCH-Zeitreihe* (X_t) selbst. Für die folgenden Überlegungen nehmen wir an, dass σ_t eine Funktion von $(e_s : s \leq t-1)$, also messbar bzgl. der von diesen Zufallsvariablen aufgepannten σ -Algebra ist. Wir werden weiter unten in Satz 14.3 sehen, dass jede streng stationäre Lösung von (14.5) automatisch diese Art Kausalität erfüllt. Weiterhin nehmen wir für den Augenblick an, dass die vierten Momente von σ_t und e_t endlich sind. Warum wir im Rahmen der *GARCH-Modelle* speziell nach streng stationären Lösungen suchen, wird weiter unten klar werden.

Durch Quadrieren der Gleichung (14.4) erhalten wir zunächst

$$X_t^2 = \sigma_t^2 + \sigma_t^2(e_t^2 - 1), \quad \forall t \in \mathbb{Z}. \quad (14.6)$$

Aufgrund der eben getroffenen Annahme sind e_t und σ_t stochastisch unabhängig, so dass die Zufallsvariable $\eta_t = X_t^2 - \sigma_t^2 = \sigma_t^2(e_t^2 - 1)$ zentriert ist. Wegen der angenommenen Existenz der vierten Momente von σ_t und e_t stellt $(\eta_t : t \in \mathbb{Z})$ im Sinne der klassischen Zeitreihenanalyse sogar ein weißes Rauschen dar, wie man leicht nachrechnet. Unter Verwendung von (14.5) ergibt (14.6) schließlich die folgende *ARMA*(max(p, q), q)-Darstellung für den quadrierten *GARCH-Prozess*

$$\begin{aligned} X_t^2 &= a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j (X_{t-j}^2 - \eta_{t-j}) + \eta_t \\ &= a_0 + \sum_{i=1}^{\max(p,q)} (a_i + b_i) X_{t-i}^2 - \sum_{j=1}^q b_j \eta_{t-j} + \eta_t, \end{aligned} \quad (14.7)$$

wobei $a_i = 0$ für $i > p$ und $b_i = 0$ für $i > q$ zu setzen ist. Man beachte aber, dass das zugehörige weiße Rauschen (η_t) lediglich aus unkorrelierten Zufallsvariablen besteht.

In Aufgabe 14.1 wird gezeigt, dass unter der Nichtnegativitätsvoraussetzung $a_1, \dots, a_p, b_1, \dots, b_q \geq 0$ die Bedingung

$$\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1 \quad (14.8)$$

notwendig und hinreichend für die Existenz einer (schwach) stationären und kausalen Lösung der *ARMA*-Gleichung (14.7) ist.

Man könnte geneigt sein zu vermuten, dass (14.8) auch notwendig und hinreichend für die Existenz eines (schwach) stationären *GARCH*-Prozesses als Lösung von (14.4) und (14.5) ist, was allerdings nicht korrekt ist, wie wir im Laufe dieses Abschnitts noch sehen werden. \square

Unter der in Bemerkung 14.2 gemachten Annahme der Unabhängigkeit von σ_t und e_s für $s \geq t$ und den Integrabilitätsbedingungen folgt nun, dass für einen *GARCH*-Prozess (X_t) stets

$$EX_t = E\sigma_t Ee_t = 0, \quad (14.9)$$

sowie für $h > 0$

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= EE[\sigma_t e_t \sigma_{t+h} e_{t+h} | e_s : s < t + h] \\ &= E(\sigma_t e_t \sigma_{t+h} E[e_{t+h} | e_s : s < t + h]) = 0 \end{aligned} \quad (14.10)$$

gilt. Ein *GARCH*-Prozess besteht somit aus unkorrelierten und zentrierten Zufallsvariablen und ist mit anderen Worten nichts anderes als ein weißes Rauschen. Da aber der quadrierte *GARCH*-Prozess einer *ARMA*-Gleichung gehorcht, liegt für diesen eine nicht verschwindende Autokovarianzfunktion vor. Diese beiden Eigenschaften sind in Übereinstimmung mit der bei den log-returns des DAX in den Abbildungen 14.2 und 14.3 gemachten empirischen Beobachtung. Insofern sind *GARCH*-Prozesse also grundsätzlich in der Lage dieses scheinbare Phänomen der Unkorreliertheit der log-returns aber der deutlichen Korreliertheit der Beträge und Quadrate der log-returns zu beschreiben. Aus wahrscheinlichkeitstheoretischer Sicht beinhalten *GARCH*-Prozesse damit eine nicht-lineare Abhängigkeitsstruktur, die sich bekanntlich nicht über die in der klassischen Zeitreihenanalyse üblichen Autokorrelationen darstellen läßt. Deshalb müssen wir bei der Untersuchung der Eigenschaften von *GARCH*-Prozessen über die Klasse der schwach stationären L_2 -Prozesse hinausgehen und werden hier nach streng stationären Lösungen der *GARCH*-Gleichungen suchen und wollen auch von vornherein keine Annahme an die Endlichkeit gewisser Momente dieser streng stationären Lösungen stellen. Festzuhalten bleibt, dass *GARCH*-Prozesse im Rahmen der klassischen Zeitreihenanalyse nicht über ein weißes Rauschen hinausgehen.

Von besonderer Bedeutung ist die sogenannte *Volatilität* σ_t einer *GARCH*-Zeitreihe (X_t) . Ihr Quadrat stimmt unter der zu Beginn von Bemerkung 14.2 gemachten Messbarkeitsannahme mit der bedingten Varianz von X_t bei gegebener Vergangenheit $\mathcal{F}_{t-1} = \sigma(e_s : s \leq t-1)$ überein. Wegen $E[X_t | \mathcal{F}_{t-1}] = 0$ gilt nämlich

$$\begin{aligned} \text{Var}[X_t | \mathcal{F}_{t-1}] &= E[X_t^2 | \mathcal{F}_{t-1}] \\ &= \sigma_t^2 = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2. \end{aligned} \quad (14.11)$$

Diese Eigenschaft impliziert, dass die bedingte Varianz einer Beobachtung X_t , also die quadrierte Volatilität, aus den p vorangegangenen log-returns $X_{t-i}, i = 1, \dots, p$, und den q vorangegangenen Volatilitäten $\sigma_{t-j}, j = 1, \dots, q$, explizit berechnet werden kann. Befindet sich unter den vergangenen returns ein größerer Wert (also eine hohe positive oder negative Rendite), so wirkt sich dieser aufgrund der Nichtnegativität der Koeffizienten in der Zukunft über höhere Streuungen der Renditen aus. Auch besitzt die quadrierte Volatilität die Eigenschaft, dass sie sich in autoregressiver Weise fortschreibt. Dies eröffnet die Möglichkeit Phasen höherer und niedriger Volatilitäten modellmäßig in einer *GARCH*-Zeitreihe zu erfassen. Hierin ist eine Stärke der *GARCH*-Zeitreihen zu sehen.

Wir wollen nun genauer die Frage der Existenz von stationären *GARCH*-Zeitreihen untersuchen. Dazu betrachten wir das einfache *GARCH*(1,1)-Modell im Detail. Hier können wir mit einem recht übersichtlichen Zugang, der auf Nelson (1990) zurückgeht, die Situation vollständig aufklären. Für *GARCH*-Prozesse höherer Ordnung haben Bougerol (1993) die Fragestellung nach stationären Lösungen abschließend geklärt. Eine ansprechende Zusammenstellung notwendiger und hinreichender Bedingungen für die Existenz streng stationärer Lösungen von *GARCH*-Gleichungen (und weiteren Verallgemeinerungen) findet man in Straumann (2005).

Es wird sich zeigen, dass es möglich ist, streng stationäre *GARCH*-Zeitreihen zu erhalten, für die keine endlichen (unbedingten) vierten Momente existieren. In einem solchen Fall besäße der in Bemerkung (14.2) abgeleitete *ARMA*-Prozess ein weißes Rauschen η_t mit nicht existierender Varianz und dann ist in jedem Fall eine Übertragung der notwendigen Bedingungen für schwache Stationarität aus der *ARMA*-Welt nicht möglich. Auch suchen wir wie gesagt bei *GARCH*-Zeitreihen nach streng stationären Lösungen und können schon allein deshalb nicht auf die *ARMA*-Theorie aufsetzen.

Konkret sind wir also im *GARCH*(1,1)-Fall an einer streng stationären Lösung der Gleichungen

$$X_t = \sigma_t e_t, \quad \forall t \in \mathbb{Z}, \quad (14.12)$$

$$\sigma_t^2 = a_0 + aX_{t-1}^2 + b\sigma_{t-1}^2, \quad \forall t \in \mathbb{Z}, \quad (14.13)$$

für gegebene i.i.d. Zufallsvariable $(e_t : t \in \mathbb{Z})$ mit $e_t \sim (0, 1)$ interessiert.

Satz 14.3 (*Strenge Stationarität im GARCH(1,1)-Modell*)

Sei $a_0 > 0$ und $a, b \geq 0$. Dann besitzen die *GARCH*(1,1)-Gleichungen (14.12) und (14.13) genau dann eine eindeutige streng stationäre Lösung, wenn

$$E \ln(ae_1^2 + b) < 0 \quad (14.14)$$

gilt. In diesem Fall besitzt die streng stationäre Lösung für σ_t^2 die Darstellung

$$\sigma_t^2 = a_0 \cdot \left(1 + \sum_{j=1}^{\infty} \prod_{i=1}^j (ae_{t-i}^2 + b)\right), \quad \forall t \in \mathbb{Z}. \quad (14.15)$$

Bevor wir Satz 14.3 beweisen, geben wir nachfolgendes Korollar an.

Korollar 14.4

Seien $a_0 > 0$ und $a, b \geq 0$. Falls $a + b < 1$, dann existiert eine eindeutige streng stationäre Lösung der GARCH(1, 1)–Gleichungen (14.12) und (14.13) mit Darstellung (14.15). Darüberhinaus besitzt diese Lösung für σ_t^2 einen endlichen Erwartungswert. Es gilt

$$EX_t^2 = E\sigma_t^2 = \frac{a_0}{1 - a - b}, \quad \forall t \in \mathbb{Z}. \quad (14.16)$$

Beweis: Zunächst folgt mit Hilfe der Jensen–Ungleichung

$$\exp(E \ln(ae_1^2 + b)) \leq E(ae_1^2 + b) = a + b.$$

Die Eigenschaft $a + b < 1$ sichert also (14.14) und damit die Existenz einer eindeutigen streng stationären Lösung von (14.12) und (14.13) mit Darstellung (14.15). Aus (14.15) folgt für $a + b < 1$ dann direkt

$$\begin{aligned} E\sigma_t^2 &= a_0 \cdot \left(1 + \sum_{j=1}^{\infty} E \prod_{i=1}^j (ae_{t-i}^2 + b)\right) \\ &= a_0 \cdot \left(1 + \sum_{j=1}^{\infty} (a + b)^j\right) = \frac{a_0}{1 - a - b}. \end{aligned}$$

□

Kommen wir nun zum

Beweis: (von Satz 14.3)

Die GARCH(1, 1)–Gleichungen (14.12) und (14.13) führen durch Ersetzen von X_{t-1}^2 durch $\sigma_{t-1}^2 e_{t-1}^2$ auf das folgende rekursive Schema für die quadrierte Volatilität σ_t^2

$$\sigma_t^2 = a_0 + (ae_{t-1}^2 + b) \cdot \sigma_{t-1}^2. \quad (14.17)$$

Analog zu den Überlegungen zur Stationarität im AR(1)–Modell, vgl. Beispiel 2.7, erhalten wir durch wiederholtes Einsetzen in (14.17)

$$\sigma_t^2 = a_0 \left(1 + \sum_{j=1}^k \prod_{i=1}^j (ae_{t-i}^2 + b)\right) + \prod_{i=1}^{k+1} (ae_{t-i}^2 + b) \cdot \sigma_{t-k-1}^2. \quad (14.18)$$

Diese Gleichung verwenden wir zunächst nur, um hieraus die Vermutung abzuleiten, dass durch (14.15) eine streng stationäre Lösung der Rekursion (14.17) und damit gemäß $X_t = \sigma_t e_t$ auch der $GARCH(1, 1)$ -Gleichungen (14.12) und (14.13) gegeben ist. Anders als z.B. in der ähnlichen Überlegung zur $AR(1)$ Reihen in Beispiel 2.7 untersuchen wir das Restglied, also den zweiten Summanden in (14.18), hier nicht explizit.

Zu zeigen ist neben der Konvergenz der Reihe in (14.15) die Tatsache, dass (σ_t) gemäß (14.15) tatsächlich (14.17) löst und dass jede Lösung von (14.17) die Darstellung (14.15) besitzt, natürlich nur unter der Voraussetzung $E \ln(ae_1^2 + b) < 0$. Nicht zu vergessen ist auch die Notwendigkeit dieser Bedingung für die Lösbarkeit der Rekursion (14.17). Die strenge Stationarität des Prozesses (σ_t^2) gemäß (14.15) ist hingegen eine direkte Folge der i.i.d.-Struktur der Innovationen (e_t) .

Der Schlüssel für die Konvergenz der Reihe in (14.15) ist die Tatsache, dass ein $\rho > 1$ existiert mit

$$\varrho^n \prod_{i=1}^n (ae_{t-i}^2 + b) \rightarrow 0 \quad \text{f.s. für } n \rightarrow \infty. \quad (14.19)$$

Hieraus folgt nämlich für jedes $t \in \mathbb{Z}$ wegen

$$\sum_{j=1}^n \prod_{i=1}^j (ae_{t-i}^2 + b) = \sum_{j=1}^n \varrho^{-j} \cdot (\varrho^j \prod_{i=1}^j (ae_{t-i}^2 + b))$$

die fast sichere absolute Konvergenz der Reihendarstellung (14.15). Man rechnet nun leicht nach, dass (14.15) wirklich eine Lösung der Rekursion (14.17) ist.

Der Nachweis von (14.19) wird mit Hilfe des Starken Gesetzes der großen Zahlen (SLLN) geführt. Wegen $E \ln(ae_1^2 + b) < 0$ existiert ein $\varrho > 1$ mit $\ln \varrho + E \ln(ae_1^2 + b) < 0$. Aufgrund der i.i.d.-Struktur der (e_t) und der Integrierbarkeit von $\ln(ae_1^2 + b)$ folgt aus dem SLLN für jedes $t \in \mathbb{Z}$

$$\begin{aligned} \ln \varrho + \frac{1}{n} \sum_{i=1}^n \ln(ae_{t-i}^2 + b) \\ \rightarrow \ln \varrho + E \ln(ae_1^2 + b) < 0 \quad \text{f.s. für } n \rightarrow \infty. \end{aligned}$$

Das heißt aber

$$\begin{aligned} \ln \left(\varrho^n \prod_{i=1}^n (ae_{t-i}^2 + b) \right) \\ = n \left(\ln \varrho + \frac{1}{n} \sum_{i=1}^n \ln(ae_{t-i}^2 + b) \right) \rightarrow -\infty \quad \text{f.s. für } n \rightarrow \infty, \end{aligned}$$

also

$$\varrho^n \prod_{i=1}^n (ae_{t-i}^2 + b) \rightarrow 0 \quad \text{f.s. für } n \rightarrow \infty,$$

wie in (14.19) behauptet.

Kommen wir nun zum Nachweis der Eindeutigkeit streng stationärer Lösungen von (14.17). Dazu sei neben (σ_t^2) gemäß (14.15) $(\tilde{\sigma}_t^2)$ eine weitere streng stationäre Lösung von (14.17). Aus (14.17) folgt mit $\varrho > 1$ gemäß (14.19)

$$\begin{aligned} |\sigma_t^2 - \tilde{\sigma}_t^2| &= (ae_{t-1}^2 + b) |\sigma_{t-1}^2 - \tilde{\sigma}_{t-1}^2| \\ &= \varrho^{-n} \cdot \varrho^n \cdot \prod_{i=1}^n (ae_{t-i}^2 + b) \cdot |\sigma_{t-n}^2 - \tilde{\sigma}_{t-n}^2| \\ &\rightarrow 0 \quad \text{n.W. für } n \rightarrow \infty. \end{aligned}$$

Für den Nachweis der Konvergenz nach Wahrscheinlichkeit beachte man die strenge Stationarität von σ_t^2 und $\tilde{\sigma}_t^2$. Aus $P\{|\sigma_t^2 - \tilde{\sigma}_t^2| > \varepsilon\} = 0 \quad \forall \varepsilon > 0$ folgt nun die fast sichere Gleichheit $\sigma_t^2 = \tilde{\sigma}_t^2$ der beiden Lösungen von (14.17).

Abschließend befassen wir uns nun noch mit der Notwendigkeit der Bedingung $E \ln(ae_1^2 + b) < 0$ für die Existenz streng stationärer Lösungen der GARCH(1, 1)-Gleichungen. Dazu sei (σ_t^2) eine beliebige streng stationäre Lösung von (14.17). Für einen Widerspruchsbeweis nehmen wir an, dass $E \ln(ae_1^2 + b) \geq 0$ gilt. Aufgrund der Nichtnegativität der auftretenden Ausdrücke erhalten wir (vgl. (14.18))

$$\sigma_t^2 \geq a_0 \left(1 + \sum_{j=1}^k \prod_{i=1}^j (ae_{t-i}^2 + b) \right).$$

Wegen

$$\ln \prod_{i=1}^j (ae_{t-i}^2 + b) = \sum_{i=1}^j \ln(ae_{t-i}^2 + b)$$

konvergiert dieser Ausdruck im Falle $E \ln(ae_1^2 + b) > 0$ wiederum aufgrund des SLLN fast sicher gegen $+\infty$. Falls $E \ln(ae_1^2 + b) = 0$ besagen Resultate für random walks (und $\sum_{i=1}^j \ln(ae_{t-i}^2 + b)$ ist ein solcher), vgl. Spitzer (1976), dass

$$\limsup_{j \rightarrow \infty} \sum_{i=1}^j \ln(ae_{t-i}^2 + b) = +\infty \quad \text{f.s.}$$

gilt. In jedem Fall haben wir also

$$\limsup_{j \rightarrow \infty} \prod_{i=1}^j (ae_{t-i}^2 + b) = +\infty \quad \text{f.s.}$$

Wegen der Nichtnegativität folgt sofort

$$\sigma_t^2 \geq \limsup_{j \rightarrow \infty} a_0 \prod_{i=1}^j (ae_{t-i}^2 + b) = +\infty \quad \text{f.s.},$$

also $\sigma_t^2 = +\infty$ f.s., im Widerspruch zu den Voraussetzungen. Damit ist Satz 14.3 bewiesen. \square

Anmerkung 14.5

Die Beweismethode zu Satz 14.3 für das $GARCH(1, 1)$ -Modell geht auf Nelson (1990) zurück und findet sich z.B. auch in Straumann (2005). \square

Wir sind nun in der Situation, dass wir im streng stationären Fall sowohl für die ganz allgemeine Situation (vgl. Satz 14.3), wie unter der Einschränkung der Existenz zweiter Momente für X_t (vgl. Korollar 14.4) notwendige und hinreichende Bedingungen für die eindeutige Existenz von Lösungen der $GARCH(1, 1)$ -Gleichungen haben. Die Frage nach streng stationären $GARCH(1, 1)$ -Prozessen mit $a + b > 1$ beantwortet das folgende Beispiel

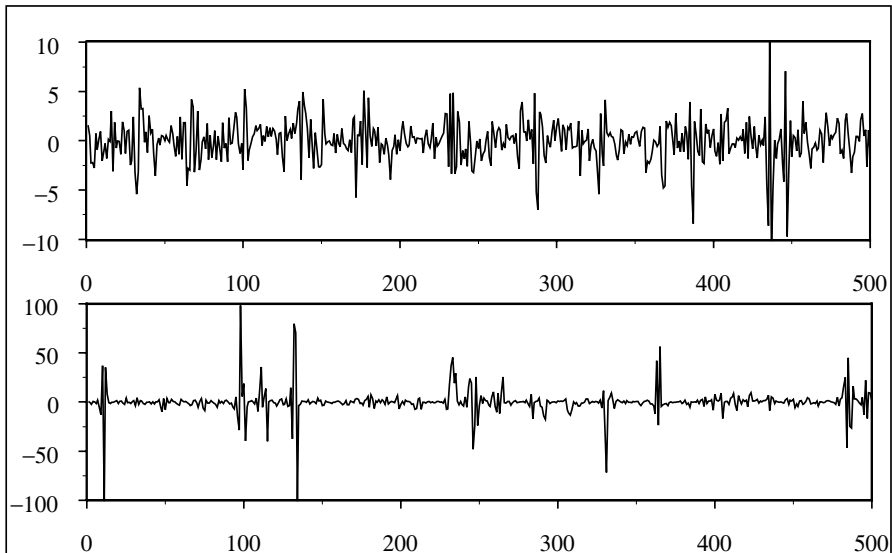


Abb. 14.4 Simulation jeweils einer $ARCH(1)$ -Zeitreihe mit Parameter $a = 0.8$ (oben) bzw. Parameter $a = 2.0$ (unten) und normalverteilten Innovationen. Man beachte die unterschiedliche Skala der y -Achse.

Beispiel 14.6

- (i) Man betrachte zunächst den einfacheren $ARCH(1)$ -Fall mit standardnormalverteilten Innovationen (e_t) . Die Stationaritätsbedingung (14.14) lautet in diesem Fall

$$\begin{aligned} E \ln(ae_1^2) &= \ln(a) + \frac{4}{\sqrt{2\pi}} \int_0^\infty \ln(x) \cdot e^{-x^2/2} dx \\ &= \ln(a) - (C + \ln(2)) < 0. \end{aligned}$$

Dabei ist $C \approx 0.5772$ die Eulersche Konstante (vgl. Gradstein/Ryshik (1981), Nr. 4.333). Also ist die Bedingung $E \ln(ae_1^2) < 0$ in dieser Situation äquivalent zu $a < 2 \cdot \exp(C) \approx 3.562$. Alle diese Parameter a führen somit zu streng stationären $ARCH(1)$ -Prozessen, aber nur für $a < 1$ besitzt ein solcher $ARCH(1)$ -Prozess endliche zweite Momente.

Eine Forderung nach endlichen vierten Momenten des $ARCH(1)$ -Prozesses führt zur weiteren Einschränkung des Parameterbereiches auf $a < \frac{1}{\sqrt{3}} \approx 0.577$ (wie man aus der Darstellung (14.15) nachrechnen kann, vgl. Aufgabe 14.4). Diese Überlegungen zeigen erneut, dass selbst einfache $ARCH$ -Zeitreihen erster Ordnung mit normalverteilten Innovationen (e_t) keine Normalverteilung als stationäre Verteilung besitzen können und dass die Forderung nach endlichen Momenten höherer Ordnung den Parameterraum immer weiter einschränkt. Man vergleiche auch die Simulation von $ARCH(1)$ -Prozessen mit Parameter $a = 0.8$ bzw. $a = 2.0$ in Abbildung 14.4.

- (ii) Jetzt betrachten wir einen $GARCH(1, 1)$ -Prozess mit der Rechteck-Innovationsverteilung $\mathcal{R}(-\sqrt{3}, \sqrt{3})$. Dann gilt $Ee_1 = 0, Ee_1^2 = 1$ sowie

$$\begin{aligned} E \ln(ae_1^2 + b) &= \ln(a) + \frac{1}{\sqrt{3}} \int_0^{\sqrt{3}} \ln\left(x^2 + \frac{b}{a}\right) dx \\ &= \ln(a) + \frac{1}{\sqrt{3}} \left[x \ln\left(x^2 + \frac{b}{a}\right) - 2x + 2\sqrt{\frac{b}{a}} \arctan\left(x\sqrt{\frac{a}{b}}\right) \right]_0^{\sqrt{3}} \\ &= \ln(a) + \ln\left(3 + \frac{b}{a}\right) - 2 + 2\sqrt{\frac{b}{3a}} \arctan\left(\sqrt{\frac{a}{3b}}\right). \end{aligned}$$

Im Falle $a = b$ führt die Forderung $E \ln(ae_1^2 + b) < 0$ auf

$$\ln(a) < -0.595 \dots \iff a < 0.551 \dots,$$

d.h. es gibt in der Tat streng stationäre $GARCH(1, 1)$ -Prozesse mit $a + b > 1$

□

Nachdem wir nun den $GARCH(1, 1)$ -Fall ausführlich diskutiert haben, wollen wir für den allgemeinen $GARCH(p, q)$ -Fall auf ein ebenso scharfes Resultat

lediglich hinweisen. Der Beweis dieses Resultates geht zurück auf Bougerol (1993) und greift auf eine interessante stochastische Version des Banachschen Fixpunktsatzes für sogenannte Stochastische Rekurrenzgleichungen zurück (vgl. Bougerol (1993)). Eine vollständige Darstellung dieser Argumentation würde den Rahmen einer Einführung in die Zeitreihenanalyse sprengen. Dem interessierten Leser sei neben den oben genannten Originalarbeiten erneut das Buch von Straumann (2005), insbesondere die Abschnitte 2.6 und 3.3 (Theorem 3.3.1), zur vertiefenden Lektüre empfohlen.

Wir beschränken uns in diesem Text auf das folgende Resultat, welches sich um $GARCH(p, q)$ -Zeitreihen mit existierenden zweiten Momenten kümmert.

Satz 14.7 (*Strenge Stationarität von $GARCH(p, q)$ -Zeitreihen*)

Es seien $p, q \in \mathbb{N}$ und $a_0 > 0$ sowie $a_i, i = 1, \dots, p$, und $b_j, j = 1, \dots, q$, nichtnegative reelle Zahlen. $(e_t : t \in \mathbb{Z})$ seien i.i.d. Zufallsvariable mit $e_t \sim (0, 1)$.

Dann existiert eine streng stationäre Lösung $X = (X_t : t \in \mathbb{Z})$ mit $EX_t^2 < \infty$ für die $GARCH(p, q)$ -Gleichungen

$$X_t = \sigma_t \cdot e_t, \quad \forall t \in \mathbb{Z}, \quad (14.20)$$

$$\sigma_t^2 = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2, \quad \forall t \in \mathbb{Z}, \quad (14.21)$$

genau dann, wenn gilt

$$\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1. \quad (14.22)$$

In diesem Falle gilt für alle $t \in \mathbb{Z}$

$$EX_t^2 = E\sigma_t^2 = \frac{a_0}{1 - \sum_{i=1}^p a_i - \sum_{j=1}^q b_j}. \quad (14.23)$$

Beweis: Wir folgen erneut der Darstellung in Straumann (2005) vgl. Proposition 3.3.3 und 3.3.5. Zunächst sei (14.22) erfüllt. Um Schreibarbeit zu sparen, nehmen wir $p = q$ an, was durch Auffüllen mit Nullen stets zu erreichen ist und somit keine Einschränkung darstellt. (14.21) bedeutet wegen $X_{t-i}^2 = e_{t-i}^2 \sigma_{t-i}^2$

$$\sigma_t^2 = a_0 + \sum_{i=1}^p (a_i e_{t-i}^2 + b_i) \cdot \sigma_{t-i}^2. \quad (14.24)$$

Durch rekursives Einsetzen in (14.24) erhalten wir

$$\sigma_t^2 = a_0 \left[1 + \sum_{i=1}^p (a_i e_{t-i}^2 + b_i) \right] + \sum_{i,j=1}^p (a_i e_{t-i}^2 + b_i)(a_j e_{t-i-j}^2 + b_j) \sigma_{t-i-j}^2.$$

Dies legt nahe, dass eine streng stationäre Lösung der *GARCH*-Gleichungen gemäß

$$\begin{aligned} \sigma_t^2 := & a_0 \left[1 + \sum_{i_1=1}^p (a_{i_1} e_{t-i_1}^2 + b_{i_1}) \right. \\ & + \sum_{i_1, i_2=1}^p (a_{i_1} e_{t-i_1}^2 + b_{i_1})(a_{i_2} e_{t-i_1-i_2}^2 + b_{i_2}) \\ & + \sum_{i_1, i_2, i_3=1}^p (a_{i_1} e_{t-i_1}^2 + b_{i_1})(a_{i_2} e_{t-i_1-i_2}^2 + b_{i_2})(a_{i_3} e_{t-i_1-i_2-i_3}^2 + b_{i_3}) \\ & \left. + \dots \right] \end{aligned} \quad (14.25)$$

gewonnen werden kann. Hierzu ist natürlich zunächst die fast sichere Konvergenz auf der rechten Seite von (14.25) zu klären. Ist diese nachgewiesen, so erhält man durch Einsetzen, dass der streng stationäre Prozess (σ_t^2) gemäß (14.25) tatsächlich (14.24) löst und damit zu einem streng stationären *GARCH*-Prozess führt.

Zum Nachweis der fast sicheren Konvergenz der rechten Seite von (14.25) genügt es wegen der Nichtnegativität aller auftretenden Terme, die Aussage $E\sigma_t^2 < \infty$ nachzuweisen. Durch Inspektion der rechten Seite von (14.25) erhalten wir wegen der Unabhängigkeit der Innovationen e_t

$$\begin{aligned} E\sigma_t^2 &= a_0 \left[1 + \sum_{i_1=1}^p (a_{i_1} + b_{i_1}) + \sum_{i_1, i_2=1}^p (a_{i_1} + b_{i_1})(a_{i_2} + b_{i_2}) + \dots \right] \\ &= a_0 \sum_{k=0}^{\infty} \left(\sum_{i=1}^p (a_i + b_i) \right)^k = \frac{a_0}{1 - \sum_{i=1}^p (a_i + b_i)}, \end{aligned}$$

also das gewünschte Resultat. Sei nun umgekehrt (X_t) eine streng stationäre Lösung von (14.21) mit $EX_t^2 = E\sigma_t^2 = E\sigma_0^2 < \infty$. Dann folgt durch Erwartungswertbildung in (14.21) unter Beachtung von $EX_t^2 = E\sigma_t^2$

$$E\sigma_0^2 = a_0 + \left(\sum_{i=1}^p a_i + \sum_{j=1}^q b_j \right) \cdot E\sigma_0^2.$$

Wegen $a_0 > 0$ folgt zwingend (14.22). Satz 14.7 ist somit bewiesen. \square

Bemerkung 14.8 (Eindeutigkeit von $GARCH(p, q)$ -Prozessen)

Unter der Bedingung (14.22) ist die im Beweis von Satz 14.7 angegebene Lösung (14.25) die einzige derartige streng stationäre Lösung der $GARCH$ -Gleichungen (14.20) und (14.21). Vgl. dazu Straumann (2005), Proposition 3.3.3 und Theorem 3.3.1. Ausdrücklich sei aber darauf hingewiesen, dass die eingängige Bedingung (14.22) an die Koeffizienten des $GARCH(p, q)$ -Modells keineswegs notwendig für die Existenz streng stationärer Lösungen der $GARCH(p, q)$ -Gleichungen ist. Es ist ausdrücklich möglich, dass stationäre $GARCH(p, q)$ -Zeitreihen mit $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j \geq 1$ existieren. Für derartige Lösungen gilt dann aber stets $EX_t^2 = E\sigma_t^2 = \infty$.

□

14.2 Parameterschätzung in GARCH-Modellen

Bevor wir uns mit der Schätzung der Modellparameter in $GARCH(p, q)$ -Zeitreihen auseinandersetzen können, müssen wir die grundsätzliche Frage der Identifizierbarkeit der Modellparameter bei diesen Zeitreihen beleuchten. Falls verschiedene Parameterkonstellationen zu identischen $GARCH$ -Prozessen führen können, dann müssen wir die Menge der zulässigen Parameter so weit einschränken, dass wir eine eindeutige Parametrisierung erreichen.

Beispiel 14.9

- (i) Das $GARCH$ -Modell (14.20), (14.21) ist für den Fall $a_1 = \dots = a_p = 0$ nicht identifizierbar. Die im Beweis von Satz 14.7 in (14.25) angegebene Lösung lautet für diesen Fall (beachte $\sum_{j=1}^q b_j < 1$)

$$\begin{aligned}\sigma_t^2 &= a_0 \left[1 + \sum_{j=1}^q b_j + \left(\sum_{j=1}^q b_j \right)^2 + \dots \right] \\ &= a_0 / \left(1 - \sum_{j=1}^q b_j \right), \quad t \in \mathbb{Z}.\end{aligned}\tag{14.26}$$

Die Heteroskedastizität verschwindet also und es gibt offensichtlich verschiedene Parameterwerte a_0, b_1, \dots, b_q mit $\sum_{j=1}^q b_j < 1$, die zu identischen Varianzen σ_t^2 in (14.26) führen. Speziell ist auch das $GARCH(1, 1)$ -Modell mit $a_0 > 0, a = 0$ und $0 \leq b < 1$ nicht eindeutig parametrisiert und damit nicht identifizierbar.

- (ii) Für gegebene i.i.d. Zufallsvariable $(e_t : t \in \mathbb{Z})$ sind die beiden stationären $GARCH$ -Zeitreihen $X_t = \sigma_t \cdot e_t$ mit

$$\sigma_t^2 = \frac{3}{4} + \frac{1}{3}X_{t-1}^2 + \frac{1}{2}\sigma_{t-1}^2$$

bzw.

$$\sigma_t^2 = 1 + \frac{1}{3}X_{t-1}^2 + \frac{1}{9}X_{t-2}^2 + \frac{1}{6}\sigma_{t-1}^2 + \frac{1}{6}\sigma_{t-2}^2$$

nicht zu unterscheiden. Dies erkennt man beispielsweise dadurch, dass im $GARCH(1, 1)$ -Modell ein weiteres Einsetzen der $GARCH(1, 1)$ -Rekursion für $(1/3)\sigma_{t-1}^2$ in diese Rekursion exakt auf die angegebene $GARCH(2, 2)$ -Rekursion führt und das obige $GARCH(1, 1)$ -Modell somit mit der eindeutigen Lösung der genannten $GARCH(2, 2)$ -Gleichung übereinstimmt. Generell könnte man in der Gleichung (14.21) z.B. σ_{t-q}^2 durch die rechte Seite von (14.21) mit t ersetzt durch $t - q$ ausdrücken und erhielte eine $GARCH$ -Darstellung mit höherem p und q . □

Um die Fragen der Identifizierbarkeit allgemein beantworten zu können, beachte man, dass die Volatilitätsrekursion einer $GARCH(p, q)$ -Zeitreihe in folgende Form umgeschrieben werden kann

$$\sigma_t^2 - \sum_{j=1}^q b_j \sigma_{t-j}^2 = \sum_{i=1}^p a_i \left(X_{t-i}^2 + \frac{a_0}{\sum_{k=1}^p a_k} \right). \quad (14.27)$$

In Analogie zu den bei den ARMA-Modellen durchgeführten Überlegungen, könnte man im Fall, dass die Polynome $B(z) = 1 - \sum_{j=1}^q b_j z^j$ und $A(z) = \sum_{i=1}^p a_i z^i$ gemeinsame Nullstellen besitzen, auf beiden Seiten der Gleichung (14.27) Linearfaktoren abspalten und gelangte so eventuell zu einem identischen $GARCH$ -Modell mit niedriger Ordnung. Dies ist etwa im Beispiel 14.9 (ii) der Fall.

Die Frage ist, ob die Bedingung

$$A(z) = \sum_{i=1}^p a_i z^i \quad \text{und} \quad B(z) = 1 - \sum_{j=1}^q b_j z^j \quad (14.28)$$

besitzen keine gemeinsamen Nullstellen

schon zur Identifizierbarkeit führt. Da (14.28) die Eindeutigkeit der Parametrisierung des aus der $GARCH$ -Gleichung abgeleiteten $ARMA(\max(p, q), q)$ -Modells (vgl. (14.17)) für X_t^2 bedeutet, könnte man geneigt sein, die Frage ohne weitere Prüfung mit ja zu beantworten, denn es gilt, dass A und B genau dann keine Nullstellen besitzen, wenn $A + B$ und B keine gemeinsamen Nullstellen haben. An dieser Stelle ist aber erneut zu beachten, dass die für $ARMA$ -Zeitreihen benutzten Techniken auf L_2 -Zeitreihen basierten und wir die einschränkende Voraussetzung $EX_t^4 < \infty$ im $GARCH$ -Kontext vermeiden wollen. Deshalb müssen wir erneut argumentieren und erhalten das folgende Resultat.

Satz 14.10 (*Identifizierbarkeit von GARCH-Modellen*)

Sei (X_t) eine stationäre $GARCH(p, q)$ -Zeitreihe (14.20), (14.21) mit Parametern $a_0 > 0, a_1, \dots, a_p, b_1, \dots, b_q \geq 0$, für die $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$ gelte.

Hinreichend für eine eindeutige Parametrisierung ist dann, dass für die zugehörigen i.i.d. Zufallsvariablen (e_t) die Größen e_t^2 keine Einpunktverteilung besitzen, dass zudem a_p oder $b_q \neq 0$ ist, sowie dass ein Index $i > 0$ existiert mit $a_i > 0$ und dass (14.28) erfüllt ist.

Bevor wir zum Beweis kommen, wollen wir einige Anmerkungen zu den Annahmen machen.

Bemerkung 14.11

Im Fall $a_p = b_q = 0$ kann man durch Multiplikation von beiden Seiten mit einem geeigneten Linearfaktor $1 + \varepsilon L$ (L bezeichne wie früher den Lag-Operator) einen identischen stationären GARCH-Prozess mit unterschiedlichen Parametern erzeugen. Auch der Fall $a_1 = \dots = a_p = 0$ führt, wie im Beispiel 14.9 (i) gesehen, zur Nichtidentifizierbarkeit. Schließlich können gemeinsame Nullstellen in den Polynomen A und B zur Möglichkeit der Abspaltung eines Linearfaktors auf beiden Seiten von (14.27) führen und so zu einem identischen, aber von den Parametern her unterschiedlichen, GARCH-Modell.

Beweis: (von Satz 14.10) Wir nehmen an, dass ein und derselbe Prozess (X_t) für unterschiedliche Parameterkonstellationen $a_0, a_1, \dots, a_p, b_1, \dots, b_q$ und $\tilde{a}_0, \tilde{a}_1, \dots, \tilde{a}_p, \tilde{b}_1, \dots, \tilde{b}_q$ mit (14.21) der Darstellung $X_t = \sigma_t \cdot e_t$, $t \in \mathbb{Z}$, genügt. Da sowohl $\sum_{j=1}^q b_j < 1$ als auch $\sum_{j=1}^q \tilde{b}_j < 1$ gilt, besitzen die zugehörigen Polynome $B(z)$ und $\tilde{B}(z)$, vgl. (14.28) keine Nullstellen mit $|z| \leq 1$. D.h. sowohl $1/B(z)$ wie $1/\tilde{B}(z)$ besitzen geometrisch konvergente Potenzreihendarstellungen. Für ein geeignetes $\delta > 0$ gilt also

$$\frac{1}{B(z)} = \sum_{j=0}^{\infty} \beta_j z^j, \quad \frac{1}{\tilde{B}(z)} = \sum_{j=0}^{\infty} \tilde{\beta}_j z^j, \quad \forall |z| \leq 1 + \delta.$$

Damit kann man aber nun für beide Parameterkonstellationen auf (14.27) den linearen L_1 -Filter $1/B(z)$ bzw. $1/\tilde{B}(z)$ anwenden (siehe Abschnitt 7.3) und erhält einerseits

$$\sigma_t^2 = c_0 + \sum_{i=1}^{\infty} c_i X_{t-i}^2 \quad (14.29)$$

sowie andererseits

$$\sigma_t^2 = \tilde{c}_0 + \sum_{i=1}^{\infty} \tilde{c}_i X_{t-i}^2 \quad (14.30)$$

wobei die c_i für $i \geq 1$ die Potenzreihenkoeffizienten von $A(z)/B(z)$ sind (analog für die \tilde{c}_i , $i \geq 1$), die geometrisch abklingen, also insbesondere absolut summierbar sind. Außerdem gilt

$$c_0 = a_0 / \left(1 - \sum_{j=1}^q b_j \right) \quad (14.31)$$

(analog für \tilde{c}_0). Aus (14.29) und (14.30) folgt (wie wir gleich sehen werden) $c_i = \tilde{c}_i \forall i = 1, 2, \dots$, woraus sich dann sofort auch $c_0 = \tilde{c}_0$ ergibt. Nehmen wir zum Nachweis an, dass $c_i = \tilde{c}_i$ für alle $i = 1, \dots, i_0 - 1$ und $c_{i_0} \neq \tilde{c}_{i_0}$ gilt. Wegen $\sigma_{t-i_0}^2 \geq a_0 > 0$ folgt hieraus mit $e_{t-i_0}^2 = X_{t-i_0}^2 / \sigma_{t-i_0}^2$

$$e_{t-i_0}^2 = \left(c_0 - \tilde{c}_0 + \sum_{i=i_0+1}^{\infty} (c_i - \tilde{c}_i) X_{t-i}^2 \right) / \left((\tilde{c}_{i_0} - c_{i_0}) \sigma_{t-i_0}^2 \right).$$

Bezeichnet \mathcal{F} die von $X_s, s \leq t-i_0-1$ aufgespannte σ -Algebra, so ist die rechte Seite der Gleichung messbar bzgl. \mathcal{F} . Andererseits ist e_{t-i_0} aber unabhängig von \mathcal{F} , wegen der i.i.d.-Struktur der Innovationen e_t . Nach Aufgabe 14.2 ist $e_{t-i_0}^2$ damit fast sicher konstant, was im Widerspruch zu unseren Annahmen steht. Es stimmen also alle Koeffizienten c_i und \tilde{c}_i überein. Aufgrund der Definition der c_i bzw. \tilde{c}_i folgt hieraus weiter

$$\frac{A(z)}{B(z)} = \frac{\tilde{A}(z)}{\tilde{B}(z)}, \quad \forall |z| \leq 1 + \delta.$$

Also gilt $A(z)\tilde{B}(z) = \tilde{A}(z)B(z)$ zunächst für alle $|z| \leq 1 + \delta$ und damit sogar für alle z . Da $A(z)$ und $B(z)$ sowie $\tilde{A}(z)$ und $\tilde{B}(z)$ keine gemeinsamen Nullstellen besitzen, müssen sich alle Nullstellen von $A(z)$ in $\tilde{A}(z)$ wiederfinden und umgekehrt. Analoges gilt für B und \tilde{B} . Die beiden Polynome B und \tilde{B} stimmen dann aber wegen der jeweils gleichen Konstanten 1 vollständig überein. Schließlich folgt daraus $A(z) = \tilde{A}(z)$, was die Gleichheiten $p = \tilde{p}$, $q = \tilde{q}$, $a_i = \tilde{a}_i$ ($i = 1, \dots, p$), $b_j = \tilde{b}_j$ ($j = 1, \dots, q$) und somit auch $a_0 = \tilde{a}_0$ nach sich zieht. Der Beweis von Satz 14.10 ist damit vollständig. \square

Wir kommen nun zur Schätzung der Modellparameter $(\underline{a}; \underline{b})^T = (a_0, a_1, \dots, a_p; b_1, \dots, b_q)^T$ in einer $GARCH(p, q)$ -Zeitreihe (X_t) . In (14.7) haben wir gesehen, dass der quadrierte Prozess (X_t^2) einer $ARMA(\max(p, q), q)$ -Gleichung genügt. Deshalb liegt es nahe, bereits bekannte Techniken für die Parameterschätzung in $ARMA$ -Modellen zu übertragen, um auf diesem Weg Schätzer für die $GARCH$ -Parameter zu gewinnen. Das Problem bei diesem Zugang ist, dass äußerst restriktive Annahmen an die Endlichkeit höherer Momente gestellt werden müssen, um die Konsistenz bzw. die asymptotische Normalität etwa des naheliegenden Kleinste-Quadrate Schätzers (für $ARMA$ -Modelle) übertragen auf $GARCH$ -Zeitreihen beweisen zu können. So benötigen Giraitis/Robinson (2001) für den Nachweis der Konsistenz des zum LS-Schätzer asymptotisch äquivalenten Whittle-Schätzers die Existenz vierter Momente für den $GARCH$ -Prozess. In Aufgabe 14.4 wird deutlich wie stark diese Einschränkung für den $GARCH(1, 1)$ -Prozess ist. Um \sqrt{n} -Konsistenz

und asymptotische Normalität zu erhalten, setzen Giraitis/Robinson (2001) darüberhinaus $EX_t^8 < \infty$ voraus. Dass diese strengen Annahmen an die Existenz höherer Momente tatsächlich notwendig sind zeigen Mikosch/Straumann (2002) für stationäre $GARCH(1, 1)$ -Zeitreihen. Sie zeigen, dass der oben genannte Whittle Schätzer im Fall $EX_t^4 = \infty$ nicht mehr konsistent ist und, dass die Tatsache $EX_t^4 < \infty$ aber $EX_t^8 = \infty$ zu einer langsameren Konvergenzrate als $1/\sqrt{n}$ für den Whittle-Schätzer führt. Aus den genannten Gründen ist die auf den ersten Blick attraktiv erscheinende Übertragung der Schätztechniken von $ARMA$ - auf $GARCH$ -Zeitreihen nicht empfehlenswert und wird deshalb hier auch nicht weiter betrachtet.

Weitaus günstigere Eigenschaften besitzt die sogenannte *Quasi Maximum Likelihood (QML)-Methode*. Da die tatsächliche Likelihoodfunktion – selbst bei normalverteilten Innovationen – für $GARCH$ -Modelle nicht zugänglich ist, betrachten wir wie bereits in Abschnitt 11.2 bedingte Verteilungen unter der Voraussetzung, dass die Innovationen e_t standard normalverteilt sind. Unter diesen Annahmen ergibt (14.4) zusammen mit der Rekursion (14.5), dass die bedingte Verteilung von X_t gegeben $X_{t-1}, \dots, X_{1-p}, \sigma_0^2, \dots, \sigma_{1-q}^2$ eine $\mathcal{N}(0, a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2)$ -Verteilung ist. Damit lautet die bedingte Dichte von (X_1, \dots, X_n) gegeben $(X_0, \dots, X_{1-p}, \sigma_0^2, \dots, \sigma_{1-q}^2)$

$$\begin{aligned} & f(X_1, \dots, X_n | X_0, \dots, X_{1-p}, \sigma_0^2, \dots, \sigma_{1-q}^2) \\ &= \prod_{t=1}^n f(X_t | X_{t-1}, \dots, X_{t-p}, \sigma_0^2, \dots, \sigma_{1-q}^2) \\ &= \prod_{t=1}^n \frac{1}{\sqrt{2\pi\sigma_t^2(\underline{a}, \underline{b})}} \exp\left(-\frac{1}{2} X_t^2 / \sigma_t^2(\underline{a}, \underline{b})\right), \end{aligned} \quad (14.32)$$

mit $\sigma_t^2(\underline{a}, \underline{b}) = \sigma_t^2$ für $t = 1 - q, \dots, 0$, sowie $\sigma_t^2(\underline{a}, \underline{b}) = a_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2(\underline{a}, \underline{b})$ für $t \geq 1$. Die sich hieraus ergebende (bedingte) Gaußsche Log-Likelihood Funktion kann ohne Kenntnis von X_{1-p}, \dots, X_0 und $\sigma_{1-q}^2, \dots, \sigma_0^2$ nicht berechnet werden. Naheliegend ist die Ersetzung $X_{1-p} = \dots = X_0 = 0$ bzw. die Verwendung eines Anfangsstücks der Länge p aus den vorliegenden Beobachtungen. Beachtet man (14.29) sowie (14.31) aus dem Beweis zu Satz 14.10 so führt ein Nullsetzen aller Beobachtungen X_s mit $s \leq 0$ auf die weitere Ersetzung $\sigma_t^2 = a_0 / (1 - b_1 - \dots - b_q)$, $t = 1 - q, \dots, 0$. Insgesamt erhalten wir die folgende approximative bedingte Log-Likelihood Funktion (bis auf die irrelevante additive Konstante $-n \ln(2\pi)/2$)

$$L_n(\underline{a}, \underline{b}) = -\frac{1}{2} \sum_{t=1}^n \left(\frac{X_t^2}{v_t^2(\underline{a}, \underline{b})} + \ln v_t^2(\underline{a}, \underline{b}) \right) \quad (14.33)$$

mit

$$v_t^2(\underline{a}, \underline{b}) = \begin{cases} a_0 + \sum_{i=1}^{\min(t-1, p)} a_i X_{t-i}^2 + \sum_{j=1}^q b_j v_{t-1}^2(\underline{a}, \underline{b}), & t \geq 1 \\ a_0 / (1 - b_1 - \dots - b_q), & t = 1 - q, \dots, 0. \end{cases} \quad (14.34)$$

$(\hat{\underline{a}}, \hat{\underline{b}})^T$ bezeichne die Maximalstelle von $L_n(\underline{a}, \underline{b})$ gemäß (14.33) über einer kompakten Teilmenge $K \subset (0, \infty) \times [0, \infty)^p \times B$ mit $B \subset \mathbb{R}^q$ gemäß $B = \{(b_1, \dots, b_q)^T : 0 \leq b_j < 1, j = 1, \dots, q \text{ und } \sum_{j=1}^q b_j < 1\}$. Dann heißt $(\hat{\underline{a}}, \hat{\underline{b}})^T$ *Quasi Maximum Likelihood (QML)–Schätzer für $(\underline{a}, \underline{b})^T$* .

Berkes et al. (2003) haben den ersten rigorosen Beweis für die Konsistenz und die asymptotische Normalität des QML–Schätzers im GARCH–Modell gegeben. Ihr Resultat lautet wie folgt.

Satz 14.12 (*Konsistenz und asymptotische Normalität des QML–Schätzers im GARCH–Modell*)

Sei $(X_t : t \in \mathbb{Z})$ eine streng stationäre GARCH(p, q)–Zeitreihe mit Parameter $(\underline{a}, \underline{b})^T = (a_0, a_1, \dots, a_p, b_1, \dots, b_q)^T$. Unter den Voraussetzungen (14.28)

$$(\underline{a}, \underline{b})^T \text{ liegt im Inneren von } K \quad (14.35)$$

$$\text{die Verteilung von } e_t^2 \text{ ist keine Einpunktverteilung,} \quad (14.36)$$

$$\text{es gilt } \lim_{x \rightarrow 0} x^{-\mu} P\{e_t^2 \leq x\} = 0 \text{ für ein beliebiges } \mu > 0, \quad (14.37)$$

sowie $E|e_t|^{2+\delta} < \infty$ für ein $\delta > 0$ gilt

$$(\hat{\underline{a}}, \hat{\underline{b}})^T \longrightarrow (\underline{a}, \underline{b})^T \text{ f.s für } n \rightarrow \infty \quad (14.38)$$

d.h. der QML–Schätzer ist konsistent. Unter (14.28), (14.35) – (14.38) und $E|e_t|^{4+\delta} < \infty$ für ein beliebiges $\delta > 0$ gilt außerdem die folgende asymptotische Normalität für den QML–Schätzer

$$\sqrt{n} \left((\hat{\underline{a}}, \hat{\underline{b}})^T - (\underline{a}, \underline{b})^T \right) \longrightarrow \mathcal{N}(\underline{0}, \Sigma(\underline{a}, \underline{b})), \quad (14.39)$$

für eine geeignete $(p + q + 1) \times (p + q + 1)$ –Matrix $\Sigma(\underline{a}, \underline{b})$.

Beweis: Siehe Berkes et al. (2003) Theorem 4.1 und Theorem 4.2

□

Bemerkung 14.13

- (i) (14.37) ist erfüllt, falls die Verteilung von e_t eine beschränkte Lebesgue–Dichte in einer Umgebung von Null besitzt.
- (ii) Straumann (2005) gibt mit seinem Theorem 4.2.1 ein zu Satz (14.10) vergleichbares Resultat unter etwas abgeschwächten Voraussetzungen an.
- (iii) Eine explizite Angabe der Kovarianzmatrix $\Sigma(\underline{a}, \underline{b})$ in (14.39) ist uns nicht bekannt.

Mit Hilfe des *QML*-Schätzers $(\widehat{a}, \widehat{b})^T$ können wir unter Verwendung von (14.34) die Volatilitäten $\sigma_t, t = 1, \dots, n$, des zugrunde liegenden *GARCH*-Prozesses approximativ rekonstruieren. Wir bezeichnen deshalb (man beachte (14.34))

$$\widehat{\sigma}_t := v_t(\widehat{a}, \widehat{b}), t = 1, \dots, n, \quad (14.40)$$

als *rekonstruierte GARCH-Volatilität*. \square

Wir wollen nun noch den nichtparametrischen Schätzer $\frac{1}{n} \sum_{t=1}^n X_t^2$ des Erwartungswertes $E\sigma_1^2 = EX_1^2 = a_0/(1-a-b)$ der quadrierten Volatilität im *GARCH*(1,1)-Modell betrachten und mit Hilfe des zentralen Grenzwertsatzes Satz 10.8 für schwach abhängige Zufallsvariable seine asymptotische Normalität nachweisen. Bemerkenswert ist dabei, dass im Gegensatz zu den früheren Grenzwertsätzen hier kein linearer Prozess zugrunde liegt.

Satz 14.14 (*Asymptotische Normalität in GARCH(1,1)-Modellen*)

Sei (X_t) eine streng stationäre *GARCH*(1,1)-Reihe, deren Innovationen e_t die Momentenbedingung

$$E(ae_1^2 + b)^2 = a^2\eta + 2ab + b^2 < 1 \text{ mit } \eta := Ee_1^4 \quad (14.41)$$

erfüllen. Dann gilt die Grenzwertaussage

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (X_t^2 - E\sigma_1^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, a_0^2 \tau^2), \quad (14.42)$$

mit

$$\tau^2 = \frac{1+a+b}{(1-a-b)^2} \left(\frac{\eta(1+a-b)+2b}{1-(a^2\eta+2ab+b^2)} - \frac{1}{1-a-b} \right). \quad (14.43)$$

Um zu sehen, dass τ^2 tatsächlich nichtnegativ ist, beachte man, dass die rechte Seite von (14.43) monoton fallend in η ist und η minimal den Wert 1 annehmen kann wegen $\eta = Ee_1^4 \geq (Ee_1^2)^2 = 1$. Für $\eta = 1$ ist die rechte Seite jedoch gleich Null. Den Beweis von Satz 14.14 haben wir in die Serie der Aufgaben 14.4, 14.6, 14.7 zerlegt. Von der Idee her folgt er dem Muster von Satz 10.15. Zum Ende dieses Abschnitts wollen wir noch eine Invertibilität für *GARCH*-Zeitreihen formulieren und beweisen. Es gilt

Satz 14.15 (*Invertibilität von GARCH-Zeitreihen*)

Sei (X_t) eine streng stationäre *GARCH*(p, q)-Zeitreihe mit der Bedingung $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$. Dann gilt die Darstellung

$$\sigma_t^2 = \frac{a_0}{1 - \sum_{j=1}^q b_j} + \sum_{k=0}^{\infty} c_k X_{t-k}^2, \quad \forall t \in \mathbb{Z}, \quad (14.44)$$

mit

$$\sum_{k=1}^{\infty} c_k z^k = \frac{a_1 z + \dots + a_p z^p}{1 - b_1 z - \dots - b_q z^q}, \quad \forall |z| \leq 1 + \delta, \quad (14.45)$$

für ein geeignetes $\delta > 0$. Für

$$\tilde{\sigma}_t^2 := \frac{a_0}{1 - \sum_{j=1}^q b_j} + \sum_{k=1}^{t-1} c_k X_{t-k}^2, \quad t = 1, 2, \dots, \quad (14.46)$$

gilt, dass ein $\varrho > 1$ existiert mit

$$\varrho^t |\tilde{\sigma}_t^2 - \sigma_t| = \mathcal{O}_P(1). \quad (14.47)$$

Schließlich gilt für $v_t^2(\underline{a}, \underline{b})$ gemäß (14.34) die Identität

$$v_t^2(\underline{a}, \underline{b}) = \tilde{\sigma}_t^2, \quad \forall t \geq 1. \quad (14.48)$$

Bevor wir Satz 14.15 beweisen sei angemerkt, dass (14.48) zusammen mit (14.47) im Nachhinein die Approximation der bedingten Varianz durch $v_t^2(\underline{a}, \underline{b})$ zur Konstruktion des *QML*-Schätzers rechtfertigt. Die Darstellung (14.44) ist eine Art *AR*(∞)-Darstellung für *GARCH*-Zeitreihen.

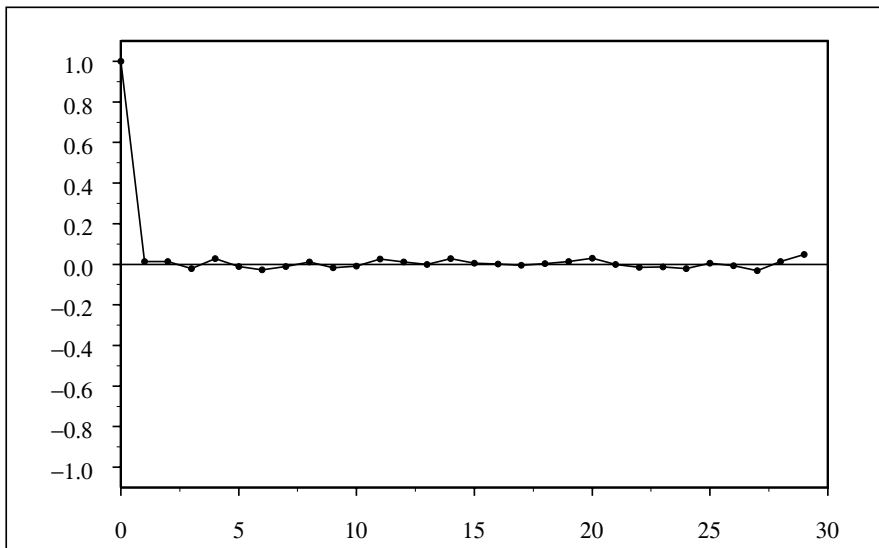


Abb. 14.5 Korrelogramm der geschätzten Innovationen aus einem an die log-returns des DAX angepassten *GARCH*(1,1)-Modells

Beweis: Da die Bedingung an die Parameter impliziert, dass das Polynom $B(z) = 1 - \sum_{j=1}^q b_j z^j$ keine Nullstellen im abgeschlossenen Einheitskreis der komplexen Zahlenebene besitzt (also auch keine Nullstellen für $|z| \leq 1 + \delta$, mit $\delta > 0$ geeignet hat), folgt ganz genauso wie im Falle von *ARMA*-Zeitreihen, dass die Koeffizienten $(c_k : k \geq 1)$ der Potenzreihe in (14.45) in geometrischer Rate abklingen. Damit und wegen $EX_t^2 < \infty$ (vgl. Satz 14.7) folgt, dass die rechte Seite in (14.44) wohldefiniert ist, da die Reihe absolut fast sicher konvergiert, siehe auch Abschnitt 7.3.

Durch direktes Einsetzen erhält man, dass die rechte Seite von (14.44) die *GARCH*-Rekursion (14.21) erfüllt. Dabei beachte man, dass für die Koeffizienten c_k die Gleichung

$$c_k - \sum_{j=1}^{\min(k-1, q)} b_j c_{k-j} = \begin{cases} a_k, & k = 1, \dots, p \\ 0, & \text{sonst} \end{cases}$$

gilt, die man durch Koeffizientenvergleich in (14.45) bestätigt. Die Eindeutigkeit der streng stationären Lösung der *GARCH*-Rekursion (vgl. Bemerkung 14.8) ergibt dann schließlich (14.44). (14.47) verifiziert man leicht mit Hilfe von $EX_t^2 < \infty$ aus dem geometrischen Abklingen der Koeffizienten c_k . (14.48) beweist man induktiv (vgl. Aufgabe 14.3). Damit ist Satz 14.15 vollständig bewiesen. \square

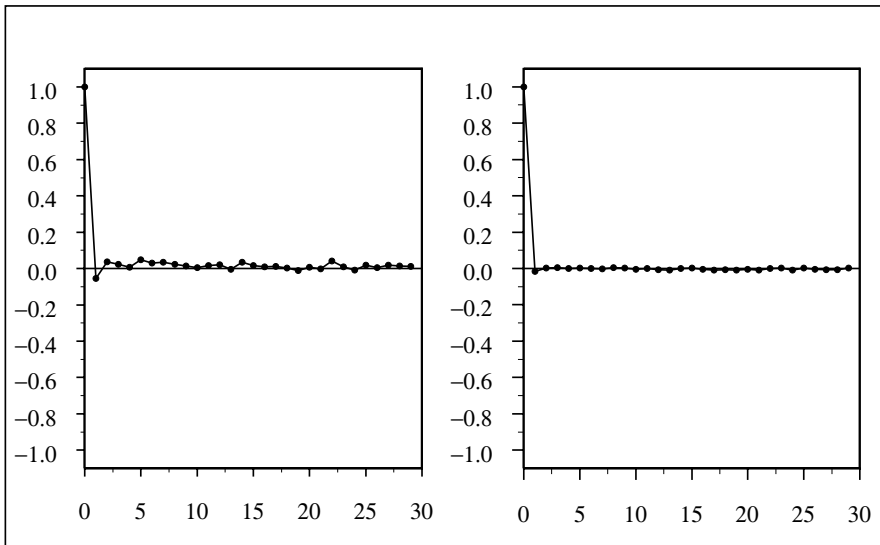


Abb. 14.6 Korrelogramm der absoluten (links) und quadrierten geschätzten Innovationen (rechts) aus einem an die log-returns des DAX angepassten *GARCH*(1, 1)-Modells

Wir schließen das Kapitel über *GARCH*-Zeitreihen nun mit einem Abschnitt über die Anwendung dieser Modelle auf einen realen Datensatz ab.

14.3 Anwendung der *GARCH*-Methodik

Bereits in Kapitel 1 wie zu Beginn dieses Kapitels haben wir als ein Beispiel für Finanzzeitreihen die Renditen des DAX für den Zeitraum Januar 1990 bis September 2005 betrachtet. Jetzt wollen wir ganz konkret an die sogenannten log-returns des DAX (vgl. Abbildung 14.1) ein *GARCH*(1, 1)-Modell anpassen. Hierzu verwenden wir die Routine `garch` des Statistiksoftwarepaketes *S + FinMetrics*[®], die eine Umsetzung des *QML*-Schätzers (vgl. (14.33)) beinhaltet. Nach Zentrierung der DAX log-returns erhalten wir aufgrund dieser

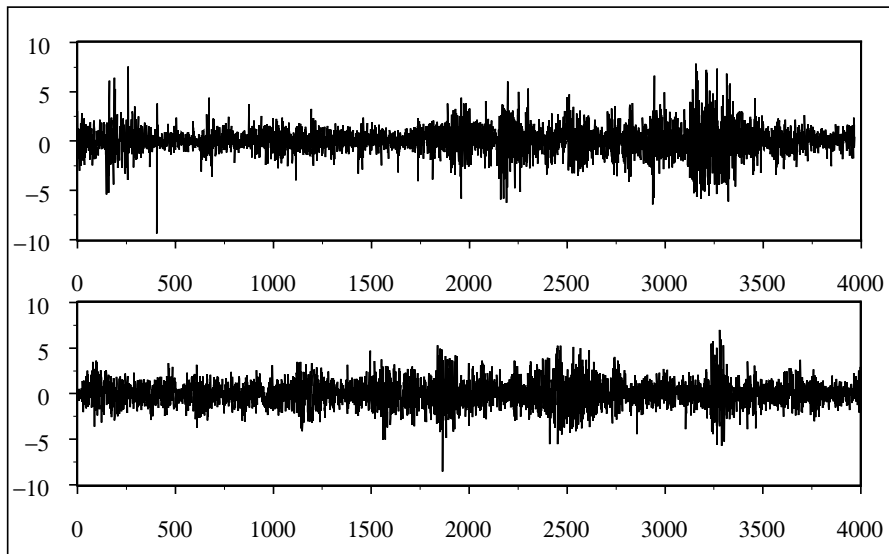


Abb. 14.7 Tägliche log-returns des DAX für den Zeitraum Januar 1990 bis September 2005 (oben) im Vergleich zu einem simulierten *GARCH*(1, 1)-Prozess gleicher Länge (unten) mit den aus den DAX-Daten geschätzten Parametern (14.49) und normalverteilten Innovationen

$n = 3967$ Datenpunkte die folgenden Schätzwerte

$$\hat{a}_0 = 0.03947, \hat{a} = 0.07755, \hat{b} = 0.90129. \quad (14.49)$$

Man erkennt, dass die Summe $\hat{a} + \hat{b}$ mit 0.97884 recht nahe bei der für streng stationäre Lösungen der *GARCH*(1, 1)-Gleichung mit endlichen zweiten Momenten zulässigen oberen Schranke eins liegt. Gemäß $v_t(\hat{a}_0, \hat{a}, \hat{b}) =$

$(v_t^2(\hat{a}_0, \hat{a}, \hat{b}))^{1/2}$ (vgl. (14.34)) können wir nun den Volatilitätsprozess σ_t unter der $GARCH(1, 1)$ -Modellannahme approximativ rekonstruieren. Direkt unter den log-returns des DAX in Abbildung 14.1 ist der so geschätzte Volatilitätsprozess zu sehen. Man erkennt, dass die Phasen höherer und niedrigerer Schwankungsbreite recht ordentlich im geschätzten Volatilitätsprozess wiedergegeben werden. Vergleicht man die Struktur des quadrierten Volatilitätsschätzer \hat{v}_t^2 aus (14.34), d.h.

$$\hat{v}_t^2 = \hat{a}_0 + \hat{a}X_{t-1}^2 + \hat{b}\hat{v}_{t-1}^2, \quad t \geq 1, \quad (14.50)$$

mit dem Ausdruck $Y_t = (1 - \varrho)X_{t-1}^2 + \varrho Y_{t-1}$, $t \geq 1$, der sich für ein exponentielles Glätten mit Glättungsparameter $\varrho \in (0, 1)$ der quadrierten log-returns ergibt, so erkennt man unter Beachtung von $\hat{a} + \hat{b} \approx 1$ und bis auf den additiven Term \hat{a}_0 eine große Ähnlichkeit. In der Tat wird ein exponentielles Glätten quadrierter returns in der Praxis durchaus eingesetzt, um modellfrei die quadrierte Volatilität zu schätzen.

In den Abbildungen 14.2 und 14.3 haben wir gesehen, dass die log-returns des DAX unkorreliert erscheinen, deren Beträge und Quadrate aber eine

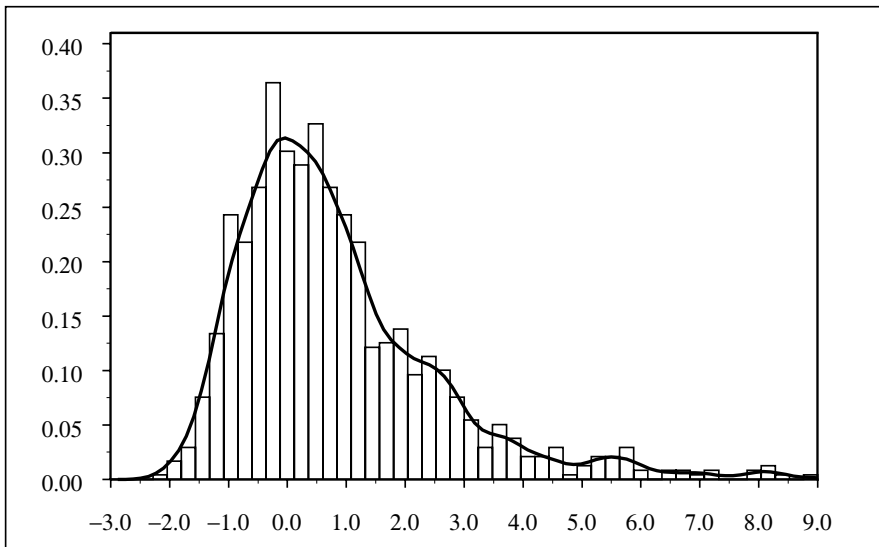


Abb. 14.8 Histogramm und geschätzte Dichte der Verteilung (zentriert am tatsächlichen Wert und mit \sqrt{n} gestreckt) des QML -Schätzers \hat{a}_0 im $GARCH(1, 1)$ -Modell mit Parametern gemäß (14.49). Zeitreihenlänge $n = 4000$ und 1000 Monte-Carlo-Wiederholungen

hohe und über längere Zeitdauern anhaltende Autokorrelationsstruktur aufweisen. Es liegt nahe, die gleichen empirischen Autokorrelationen auf Basis

der geschätzten Innovationen $\hat{e}_t := X_t/(\hat{v}_t^2(\hat{a}_0, \hat{a}, \hat{b}))^{1/2}$, $t = 1, \dots, n$, zu bestimmen (vgl. Abbildungen 14.5 und 14.6). In der Tat tauchen sowohl bei den geschätzten Innovationen, als auch bei deren Beträgen und Quadraten, keine nennenswerten Autokorrelationen (d.h. keine linearen Abhängigkeiten) mehr auf. Insofern war die Modellierung der DAX-Renditen mit einem $GARCH(1, 1)$ -Modell also durchaus erfolgreich.

Abbildung 14.7 enthält einen simulierten $GARCH(1, 1)$ -Prozess der Länge

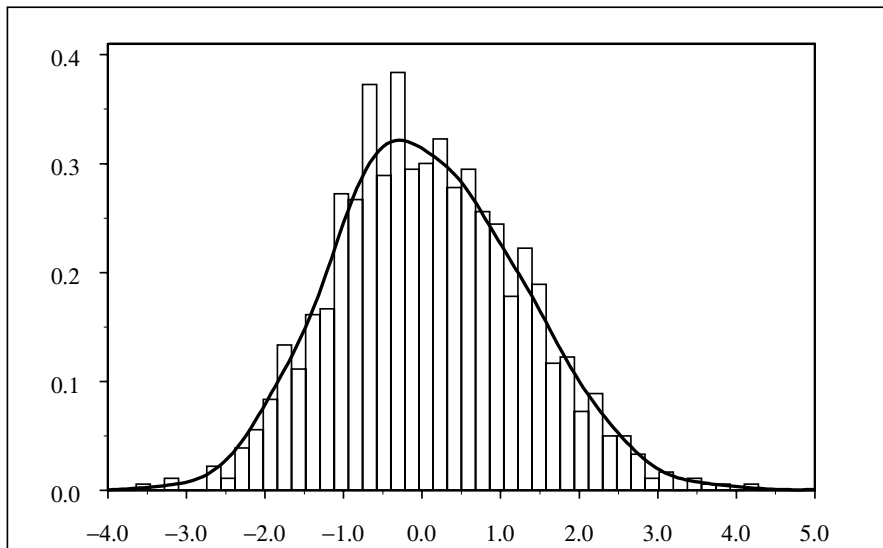


Abb. 14.9 Histogramm und geschätzte Dichte der Verteilung (zentriert am tatsächlichen Wert und mit \sqrt{n} gestreckt) des QML -Schätzers \hat{a} im $GARCH(1, 1)$ -Modell mit Parametern gemäß (14.49). Zeitreihenlänge $n = 4000$ und 1000 Monte-Carlo-Wiederholungen

$n = 3967$, mit den Parametern aus (14.49) und einem normalverteiltem weißen Rauschen im Vergleich zu den beobachteten log-returns des DAX. Beide graphischen Darstellungen in Abbildung 14.7 ähneln einander durchaus, auch wenn bei genauerem Hinsehen doch noch einige Unterschiede zu Tage treten. Besonders auffällig ist, dass die simulierte $GARCH(1, 1)$ -Zeitreihe den Übergang von Phasen höherer zu niedrigerer Volatilität jeweils weniger abrupt vollzieht und dass isolierte kurzfristige Gewinne und Verluste, wie etwa der fast zehnprozentige Verlust des DAX zu Beginn des Beobachtungszeitraumes, im simulierten $GARCH(1, 1)$ -Prozess mit normalverteilten Innovationen weit weniger häufig auftreten. Dies wird in gewisser Weise auch in der durch das $GARCH(1, 1)$ -Modell geschätzten Volatilität in Abbildung 14.1 deutlich. Dort erkennt man, dass die geschätzte Volatilität ebenfalls in der Regel nicht ab-

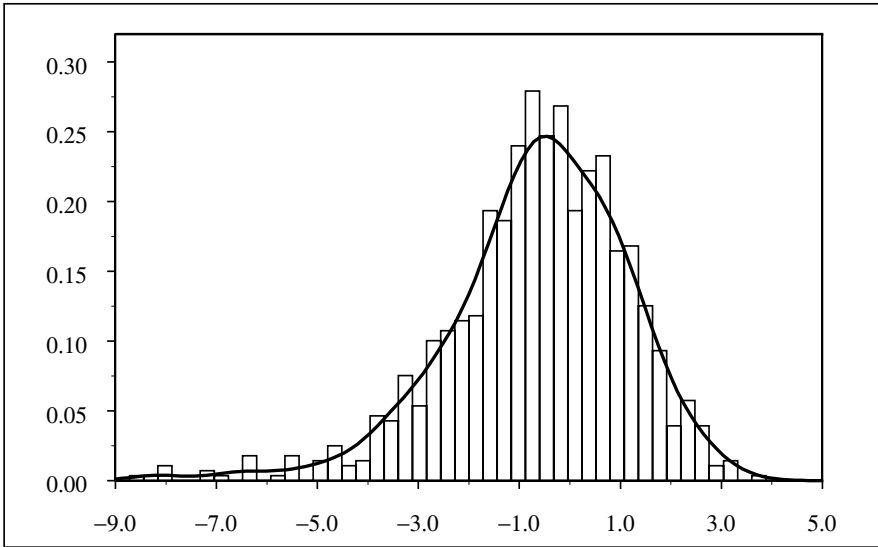


Abb. 14.10 Histogramm und geschätzte Dichte der Verteilung (zentriert am tatsächlichen Wert und mit \sqrt{n} gestreckt) des *QML*-Schätzers \hat{b} im *GARCH*(1, 1)-Modell mit Parametern gemäß (14.49). Zeitreihenlänge $n = 4000$ und 1000 Monte-Carlo-Wiederholungen

rupt, sondern eher über einem gewissen Zeitfenster abklingt. Natürlich könnte man alternative Verteilungsannahmen für die Innovationen ausprobieren. Etwa solche, die stärkeres Gewicht auf die Tailbereiche legen (wie beispielsweise eine Studentsche *t*-Verteilung), um zu sehen, ob dadurch eine weitere Verbesserung bei der Beschreibung der log-returns des DAX zu erreichen ist. Ein weiterer Vorschlag der sich in der Literatur findet, ist, zusätzliche additive Sprunganteile (sogenannte *jump*-Komponenten) im Rahmen der Modellierung von Finanzzeitreihen zu verwenden. Wir wollen allerdings beides nicht weiter verfolgen und unsere Untersuchung an dieser Stelle beenden.

Wir beschließen den Abschnitt mit einer Simulationsstudie über das Verhalten der *QML*-Schätzer \hat{a}_0 , \hat{a} und \hat{b} im *GARCH*(1, 1)-Modell. Dazu simulieren wir 1000 *GARCH*(1, 1)-Zeitreihen jeweils der Länge $n = 4000$ mit tatsächlichen Parametern gemäß (14.49). In jeder der simulierten Datenreihen werden die *QML*-Schätzer berechnet und ein Histogramm für die an dem tatsächlichen Wert zentrierten und mit \sqrt{n} gestreckten Größen aufgestellt (vgl. Abbildungen 14.8–14.10). Insgesamt zeigt sich ein recht stabiles Verhalten aller drei *QML*-Schätzer, wobei trotz des recht hohen Stichprobenumfangs von $n = 4000$ teilweise eine recht deutliche Schiefe der simulierten Verteilung und damit eine Abweichung von der asymptotischen Normalverteilung zu beobachten ist. Um einen Eindruck von der Korrelation der Schätzer \hat{a} und \hat{b} zu gewinnen, wird in Abbildung 14.11 ein Scatterplot der 1000 simulierten Schätzwerte (\hat{a}, \hat{b}) wie-

dergegeben. Dieser Scatterplot zeigt, dass der Zusammenhang der Schätzer für die beiden *GARCH*-Parameter recht groß ist.

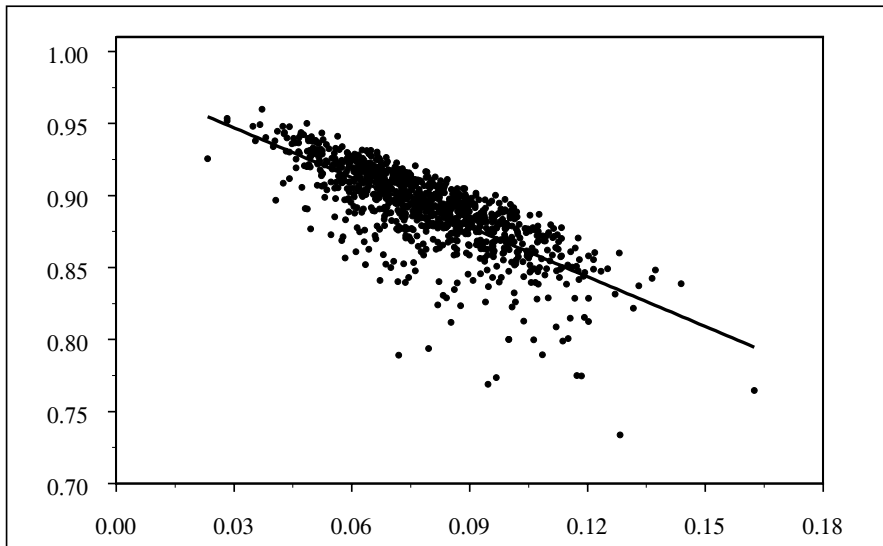


Abb. 14.11 Scatterplot und Regressionsgerade der simulierten *QML*-Schätzer (\hat{a}, \hat{b}) im *GARCH*(1,1)-Modell mit Parametern gemäß (14.49). Zeitreihenlänge $n = 4000$ und 1000 Monte-Carlo-Wiederholungen

Aufgaben

Aufgabe 14.1 Man zeige, dass unter der Voraussetzung $Ee_t^4 < \infty$ und $E\sigma_t^4 < \infty$, sowie der Nichtnegativität von a_1, \dots, a_p und b_1, \dots, b_q die Bedingung $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$ notwendig und hinreichend für die Existenz einer kausalen Lösung (X_t^2) der *ARMA*-Gleichung (14.7) ist.

Aufgabe 14.2 Man zeige, dass für stochastisch unabhängige Zufallsvariable X und Y aus $P\{X = Y\} = 1$ die fast sichere Konstanz von X und Y folgt.

Aufgabe 14.3 Man beweise die Gleichung (14.48).

Aufgabe 14.4 Man zeige, dass im streng stationären *GARCH*(1,1)-Modell (X_t) die Bedingungen $EX_1^4 < \infty$ und $E(ae_1^2 + b)^2 < 1$ äquivalent sind. Mit $\vartheta^2 := E(ae_1^2 + b)^2 = a^2\eta + 2ab + b^2$, $\eta := Ee_1^4$, zeige man in diesem Fall

$$EX_1^4 = \eta \frac{1}{1 - \vartheta^2} \frac{1 + a + b}{1 - a - b}.$$

Aufgabe 14.5 Man zeige, dass im streng stationären $GARCH(1, 1)$ -Modell (X_t) mit $EX_t^4 < \infty$ die Autokovarianzfunktion von (X_t^2) in geometrischer Rate abklingt, genauer $Cov(X_t^2, X_{t+h}^2) = \mathcal{O}((a+b)^h)$ für $h \in \mathbb{N}_0$.

Aufgabe 14.6 Sei (X_t) ein streng stationäres $GARCH(1, 1)$ -Modell mit $EX_t^4 < \infty$. Man zeige, dass für alle $s_1 < \dots < s_u < s_u + r = t$ und alle messbaren und quadratintegrierbaren Funktionen $f: \mathbb{R}^u \rightarrow \mathbb{R}$ die folgende Kovarianzungleichung gilt.

$$Cov(f(X_{s_1}^2, \dots, X_{s_u}^2), X_t^2) \leq \mathcal{O}(1) \sqrt{Ef(X_{s_1}^2, \dots, X_{s_u}^2)} \times \vartheta^r$$

mit $\vartheta := (E(ae_1^2 + b)^2)^{1/2} < 1$.

Hinweis: Man verwende die Darstellung (vgl. (14.15))

$$X_t^2 = a_0 e_t^2 \left(1 + \sum_{j=1}^{\infty} \prod_{i=1}^j (ae_{t-i}^2 + b) \right), \quad \forall t \in \mathbb{Z}, \quad (14.51)$$

und gehe analog zum Beweis von (10.88) für lineare Zeitreihen vor. Statt des dortigen \tilde{Z}_t verwende man hier $\tilde{X}_t^2 := a_0 e_t^2 \left(1 + \sum_{j=1}^{r-1} \prod_{i=1}^j (ae_{t-i}^2 + b) \right)$.

Aufgabe 14.7 Man beweise unter Verwendung der Aufgaben 14.4 und 14.6 den Satz 14.14. Insbesondere berechne man auch die asymptotische Varianz τ^2 aus (14.43).

Hinweis: Man gehe analog zum Beweis von Satz 10.15 vor, s. insbesondere (10.90)-(10.94).

Aufgabe 14.8 Sei (e_t) ein reelles weißes Rauschen bestehend aus i.i.d. Zufallsvariablen. Für $\mu \in \mathbb{R}$ betrachte man die Zeitreihen $X_{t,n} = e_t - \mu$ für $t = 1, \dots, n$ und $X_{t,n} = e_t + \mu$ für $t = n+1, \dots, 2n$.

Man zeige, dass einerseits $E \frac{1}{2n} \sum_{t=1}^{2n} X_{t,n} = 0$ gilt und andererseits für

$$\hat{\gamma}_n(h) = \frac{1}{2n} \sum_{t=1}^{2n-h} X_{t,n} X_{t+h,n}$$

die Aussage $\hat{\gamma}_n(h) \rightarrow \mu^2$ n.W. für $n \rightarrow \infty$ richtig ist $\forall h \geq 1$, obwohl die einzelnen $X_{t,n}$ für verschiedene t (bei festem n) sogar stochastisch unabhängig sind.

Dies einfache Beispiel zeigt, dass ein unbemerkter Strukturbruch im Erwartungswert einer Zeitreihe bei der Berechnung der empirischen Autokovarianzen ein „long memory-Verhalten“ vortäuschen kann.

Grundlagen multivariater Zeitreihen

In diesem Kapitel wollen wir in knapper Form Grundbegriffe mehrdimensionaler Zeitreihen einführen, ohne in dieser Einführung wirklich in die Tiefe gehen zu können. Jedoch wollen wir Beweise im Gegensatz zu bestehender Literatur wenn möglich nicht nur informell, sondern in mathematisch exakter Form darstellen. Für eine weiterführende Betrachtung verweisen wir z.B. auf die Monografien von Lütkepohl (1993) und Reinsel (2003), sowie auf die einschlägigen Kapitel in Brockwell/Davis (1991).

In der Praxis werden häufig zu einem Zeitpunkt $t \in T$ mehrere Messungen gleichzeitig durchgeführt, z.B. wird die Temperatur, der Luftdruck und die Luftfeuchtigkeit an einem bestimmten Ort gemessen, oder man betrachtet gleichzeitig für mehrere im DAX geführte Unternehmen tägliche Renditen bzw. log-returns (vgl. Abbildung 15.1). Man gelangt so zu einem Vektor von Beobachtungen

$$X_t = (X_{t1}, \dots, X_{tr})^T, \quad \forall t \in T, \quad (15.1)$$

und nennt dann $X = (X_t : t \in T)$ eine r -dimensionale oder r -variate Zeitreihe. Je nach Bedarf werden wir auch $X_t = X(t) = (X_1(t), \dots, X_r(t))^T$ schreiben. Im Folgenden sei stets $T = \mathbb{Z}$ und alle Komponenten $X_{tj}, j = 1, \dots, r, t \in \mathbb{Z}$, seien Elemente von $L_2^{\mathbb{C}}(P) = L_2^{\mathbb{C}}(\Omega, \mathcal{A}, P)$; wir sprechen dann kurz von X als einer L_2 -Reihe. Der Vektor der Erwartungswerte

$$EX_t := (EX_{t1}, \dots, EX_{tr})^T, \quad \forall t \in \mathbb{Z}, \quad (15.2)$$

heißt *Erwartungswert* von X_t , während die $r \times r$ -Matrix

$$\begin{aligned} \text{Cov}(X_t, X_s) &= (\text{Cov}(X_{tj}, X_{sk}))_{j,k=1,\dots,r} \\ &= (E(X_{tj} - EX_{tj})(\overline{X}_{sk} - \overline{EX}_{sk}))_{j,k=1,\dots,r} \\ &= E(X_t - EX_t)(\overline{X}_s - \overline{EX}_s)^T, \quad \forall s, t \in \mathbb{Z}, \end{aligned} \quad (15.3)$$

als *Autokovarianz(matrix)* von X_t und X_s bezeichnet wird. Bei multivariaten Zeitreihen spielt auch die Frage nach dem Zusammenhang zwischen den

Komponenten des Beobachtungsvektors eine wichtige Rolle.

In Verallgemeinerung des univariaten Stationaritätsbegriffs definieren wir

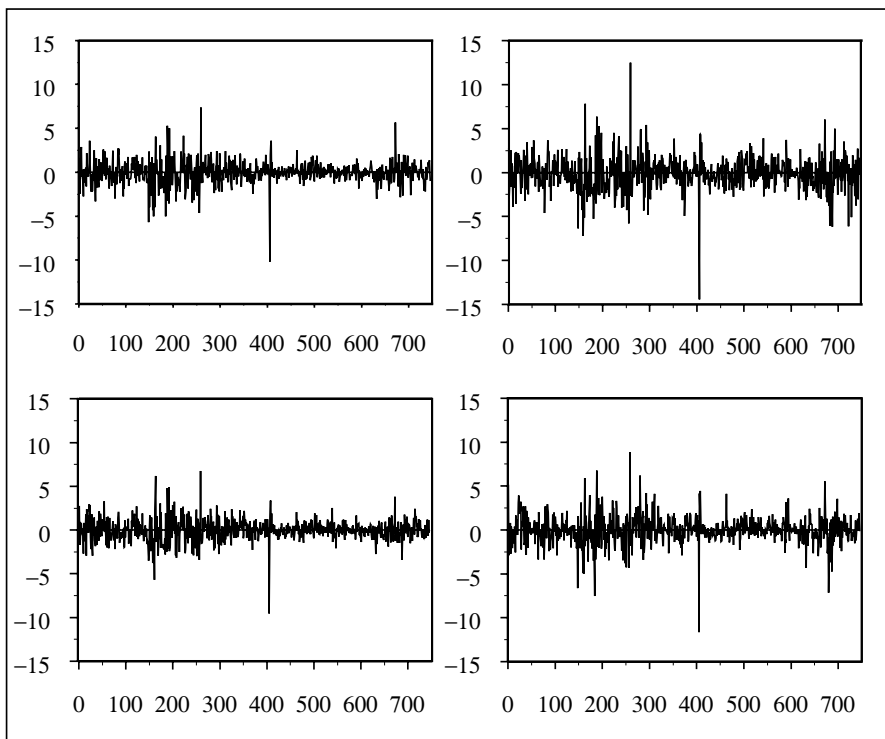


Abb. 15.1 Tägliche log-returns der Aktien der Deutschen Bank (oben links), der Volkswagen AG (oben rechts), der Siemens AG (unten links) und von BMW (unten rechts) für den Zeitraum 1990 – 1992

Definition 15.1 (Schwach stationäre multivariate Zeitreihe)

Eine r -dimensionale Zeitreihe $X = (X_t : t \in \mathbb{Z})$ heißt (schwach) stationär, wenn der Erwartungswert(vektor) $\mu := EX_t$ nicht von t abhängt und ebenfalls die (Auto-)Kovarianz(matrix)

$$\Gamma(h) = \text{Cov}(X_{t+h}, X_t) =: (\gamma_{ij}(h))_{i,j=1,\dots,r}, \quad \forall h, t \in \mathbb{Z}, \quad (15.4)$$

nicht von t abhängt. Offensichtlich gilt dann

$$\Gamma(-h) = \Gamma(h)^*, \quad \forall h \in \mathbb{Z}, \quad (15.5)$$

wobei für eine komplexe Matrix $A = (a_{ij})$ die komplex konjugierte Matrix A^* gemäß $A^* = \overline{A}^T$ definiert sei. Ist X reell, d.h. sind alle X_{tj} reell, so ist auch

$\Gamma(h)$ reell und es gilt $\Gamma(h)^* = \Gamma(h)^T$.

Die Korrelation(smatrix) $R(h)$ ist definiert durch

$$R(h) = (\gamma_{ij}(h)/(\gamma_{ii}(0)\gamma_{jj}(0))^{1/2})_{i,j=1,\dots,r} =: (\varrho_{ij}(h))_{i,j=1,\dots,r}. \quad (15.6)$$

Demgemäß ist also $\varrho_{ij}(h)$ die Korrelation zwischen $X_{t+h,i}$ und X_{tj} . Für $i = j$ ist $\varrho_{ii}(0) = 1$, jedoch gilt dies nicht notwendig für $i \neq j$.

Beispiel 15.2 Sei $e = (e_t : t \in \mathbb{Z})$ ein 1-dimensionales, reelles weißes Rauschen mit $e_t \sim (0, 1)$. Die 2-dimensionale Zeitreihe $X(t) = (X_1(t), X_2(t))^T, t \in \mathbb{Z}$, sei definiert gemäß

$$\begin{aligned} X_1(t) &= e_t \\ X_2(t) &= e_t + be_{t-1}, \quad b \in \mathbb{R}. \end{aligned}$$

Dann gilt $EX(t) = 0$, sowie $\Gamma(h) = R(h) = 0$ für $h \notin \{0, 1, -1\}$ und

$$\begin{aligned} \Gamma(0) &= \begin{pmatrix} 1 & 1 \\ 1 & 1+b^2 \end{pmatrix}, \Gamma(1) = \Gamma(-1)^T = \begin{pmatrix} 0 & 0 \\ b & b \end{pmatrix}, \\ R(0) &= \begin{pmatrix} 1 & 1/\sqrt{1+b^2} \\ 1/\sqrt{1+b^2} & 1 \end{pmatrix}, R(1) = R(-1)^T = \begin{pmatrix} 0 & 0 \\ \frac{b}{\sqrt{1+b^2}} & \frac{b}{1+b^2} \end{pmatrix}. \end{aligned}$$

□

In Verallgemeinerung des eindimensionalen weißen Rauschens definieren wir

Definition 15.3 (Multivariates weißes Rauschen und MA-Reihen)

Eine r -dimensionale Zeitreihe $e = (e_t : t \in \mathbb{Z})$ heißt weißes Rauschen, wenn sie zentriert ist, also für den Erwartungswertvektor $Ee_t = 0$ für alle t gilt und für die Kovarianzmatrizen der Vektoren e_t und e_s gilt

$$\text{Cov}(e_s, e_t) = \begin{cases} \Sigma & , \text{ für } s = t \\ 0 & , \text{ für } s \neq t \end{cases} \quad (15.7)$$

mit einer Kovarianzmatrix $\Sigma = (\sigma_{ij})_{i,j=1,\dots,r} \in \mathbb{C}^{r \times r}$. Wir schreiben dafür kurz $e \sim \text{WR}(0, \Sigma)$. Ist e reell, so auch Σ .

In Analogie zum univariaten Fall können wir neue stationäre Zeitreihen aus einem weißen Rauschen durch Reihendarstellung (sog. MA-Zeitreihen)

$$X_t = \sum_{j=-\infty}^{+\infty} c_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (15.8)$$

gewinnen. Dabei sind $c_j = (c_j(\varrho, \varrho'))_{\varrho, \varrho'=1,\dots,r}$ $r \times r$ -Matrizen aus $\mathbb{C}^{r \times r}$ mit der Eigenschaft

$$\sum_{j=-\infty}^{+\infty} |c_j(\varrho, \varrho')|^2 < \infty, \quad \forall \varrho, \varrho' = 1, \dots, r. \quad (15.9)$$

Die Konvergenz in (15.8) ist als $L_2^{\mathbb{C}}(P)$ -Konvergenz aller r Komponenten zu verstehen. Letztere folgt direkt aus dem univariaten Fall, da die Komponenten $(e_{t_{\varrho}} : t \in \mathbb{Z})$ von jeweils 1-dimensionale weiße Rauschen sind und deshalb die Reihen $\sum_{j=-\infty}^{+\infty} c_j(\varrho, \varrho') e_{t-j, \varrho'}$ für alle ϱ, ϱ' in $L_2^{\mathbb{C}}(P)$ konvergieren, vgl. Bemerkung 6.4 c) und Satz 6.3.

Man rechnet leicht nach (siehe Aufgabe 15.1), dass $X = (X_t : t \in \mathbb{Z})$ zentriert und stationär ist mit Kovarianz

$$\Gamma(h) = \sum_{j=-\infty}^{+\infty} c_{j+h} \Sigma c_j^*, \quad \forall h \in \mathbb{Z}. \quad (15.10)$$

15.1 Multivariate Spektraltheorie

In diesem Abschnitt werden wir eine multivariate Verallgemeinerung des Herglotz-Lemmas (Satz 3.4) und des Spektralsatzes (Satz 5.7) herleiten. Dazu seien im Folgenden alle Zeitreihen zentrierte L_2 -Prozesse.

Satz 15.4 (Multivariates Herglotz-Lemma)

Sei $X = (X(t) : t \in \mathbb{Z})$ schwach stationär mit Kovarianz $\Gamma(h), h \in \mathbb{Z}$. Dann gilt die folgende Matrixgleichung (wobei das Integral über eine Matrix von Maßen komponentenweise zu verstehen ist)

$$\Gamma(h) = \int_{(-\pi, \pi]} e^{ih\omega} \mu(d\omega) = \left(\int_{(-\pi, \pi]} e^{ih\omega} \mu_{j,k}(d\omega) \right)_{j,k=1,\dots,r}. \quad (15.11)$$

Dabei ist $\mu = (\mu_{jk})_{j,k=1,\dots,r}$ eine eindeutig bestimmte Hermitesche positiv semidefinite Matrix komplexwertiger, signierter Maße $\mu_{jk} = \nu_{jk} + i\varrho_{jk}$. Letzteres heißt, dass ν_{jk} und ϱ_{jk} Differenzen endlicher Maße sind, während Hermitesch bedeutet, dass für jedes $B \in \mathbb{B} \cap (-\pi, \pi]$ die Matrix $\mu(B) = (\mu_{jk}(B))_{j,k}$ Hermitesch ist, also $(\mu(B))^* = \mu(B)$ gilt. Entsprechend bedeutet μ positiv semidefinit, dass alle Matrizen $\mu(B)$ positiv semidefinit sind. Wir nennen μ Spektralmatrix von X oder von Γ .

Beweis: Sei $\alpha \in \mathbb{C}^r$ und $X_\alpha(t) := \alpha^T X(t)$. Wegen

$$\gamma_\alpha(t-s) := EX_\alpha(t) \overline{X_\alpha(s)} = \alpha^T EX(t) \overline{X(s)}^T \overline{\alpha} = \alpha^T \Gamma(t-s) \overline{\alpha},$$

ist X_α eine stationäre (i.A. komplexe) eindimensionale L_2 -Reihe mit Autokovarianzfunktion γ_α . Nach dem Herglotz-Lemma (Satz 3.4) gibt es ein eindeutig bestimmtes endliches Maß μ_α auf $(-\pi, \pi]$ mit

$$\gamma_\alpha(h) = \int_{(-\pi, \pi]} e^{ih\omega} \mu_\alpha(d\omega), \quad \forall h \in \mathbb{Z}.$$

Wir wählen nun verschiedene Werte für α . Ist α der j -te Einheitsvektor $e_j = (0, \dots, 0, 1, 0, \dots, 0)^T$, so schreiben wir γ_j und μ_j . Für $\alpha_1 = e_j + e_k, j \neq k$, d.h. $X_{\alpha_1} = X_j + X_k$, bzw. $\alpha_2 = e_k - ie_j$, also $X_{\alpha_2} = X_k - iX_j$, schreiben wir $\gamma_{\tau;jk}$ und $\mu_{\tau;jk}$, $\tau = 1, 2$. Mit dem Skalarprodukt $\langle \cdot, \cdot \rangle$ in $L_2^{\mathbb{C}}(P)$ gilt

$$\begin{aligned} & \frac{1}{2}[\gamma_{1;jk}(t-s) - \gamma_j(t-s) - \gamma_k(t-s)] \\ & + \frac{i}{2}[\gamma_{2;jk}(t-s) - \gamma_j(t-s) - \gamma_k(t-s)] \\ & = \frac{1}{2}[\langle X_j(t) + X_k(t), X_j(s) + X_k(s) \rangle - \langle X_j(t), X_j(s) \rangle - \langle X_k(t), X_k(s) \rangle] \\ & + \frac{i}{2}[\langle X_k(t) - iX_j(t), X_k(s) - iX_j(s) \rangle - \langle X_j(t), X_j(s) \rangle - \langle X_k(t), X_k(s) \rangle] \\ & = \langle X_j(t), X_k(s) \rangle = \gamma_{jk}(t-s). \end{aligned} \quad (15.12)$$

Wir definieren nun signierte Maße $\nu_{jk} := \frac{1}{2}(\mu_{1;jk} - \mu_j - \mu_k)$ und $\varrho_{jk} := \frac{1}{2}(\mu_{2;jk} - \mu_j - \mu_k)$ sowie $\mu_{jk} := \nu_{jk} + i\varrho_{jk}$.

Einsetzen der eindimensionalen Spektraldarstellungen in obige Gleichungskette liefert unmittelbar

$$\gamma_{jk}(h) = \int_{(-\pi, \pi]} e^{ih\omega} \mu_{jk}(d\omega), \quad \forall j, k = 1, \dots, r, \quad h \in \mathbb{Z}. \quad (15.13)$$

Wie im Beweis des univariaten Herglotz-Lemmas zeigt man, dass das komplexe, signierte Maß μ_{jk} eindeutig bestimmt ist, indem man beliebige Indikatoren im Raum $L_2^{\mathbb{C}}(\mu_j + \mu_k + \mu_{1;jk} + \mu_{2;jk})$ durch endliche trigonometrische Polynome approximiert.

Weiter gilt mit $\mu := (\mu_{jk})_{j,k=1,\dots,r}$ (dabei sei $\mu_{jj} := \mu_j$) und $h \in \mathbb{Z}$

$$\begin{aligned} \int_{(-\pi, \pi]} e^{ih\omega} \mu^*(d\omega) &= \left(\int_{(-\pi, \pi]} e^{ih\omega} (\nu_{kj} - i\varrho_{kj})(d\omega) \right)_{j,k} \\ &= \left(\int_{(-\pi, \pi]} e^{-ih\omega} \nu_{kj}(d\omega) - i \int_{(-\pi, \pi]} e^{-ih\omega} \varrho_{kj}(d\omega) \right)_{j,k} \\ &= \left(\int_{(-\pi, \pi]} e^{-ih\omega} (\nu_{kj} + i\varrho_{kj})(d\omega) \right)_{j,k} \\ &= \left(\int_{(-\pi, \pi]} e^{-ih\omega} \mu_{kj}(d\omega) \right)_{j,k} \\ &= \left(\overline{\gamma_{k,j}(-h)} \right)_{j,k} = (\gamma_{jk}(h))_{j,k} \quad (\text{vgl. (15.5)}) \\ &= \int_{(-\pi, \pi]} e^{ih\omega} \mu(d\omega). \end{aligned}$$

Aus der Eindeutigkeit der Darstellung folgt $\mu^* = \mu$. Außerdem gilt

$$\gamma_\alpha(t) = \alpha^T \Gamma(h) \bar{\alpha} = \alpha^T \int_{(-\pi, \pi]} e^{it\omega} \mu(d\omega) \bar{\alpha} = \int_{(-\pi, \pi]} e^{it\omega} (\alpha^T \mu \bar{\alpha})(d\omega),$$

insbesondere ist das reelle Maß $\alpha^T \mu \bar{\alpha}$ nichtnegativ und erfüllt somit die Ungleichung $\alpha^T (\mu_{jk}(B))_{j,k} \bar{\alpha} \geq 0$, $\forall \alpha \forall B$, d.h. μ ist positiv semidefinit. \square

Für die weiteren Überlegungen sei m ein reelles, endliches Maß, das alle in der Definition von μ_{jk} vorkommenden Maße dominiert. Man kann beispielsweise $m = \sum_{j,k=1}^r (\mu_j + \mu_k + \mu_{1;jk} + \mu_{2;jk})$ wählen und nennt $f_{jk} = d\mu_{jk}/dm$ die entsprechenden Radon–Nikodym Dichten. Dann gilt wie oben für die Matrix $f = (f_{jk})_{j,k}$. Man beachte im folgenden Ausdruck, dass im Gegensatz zu (15.11) nun der *Integrand* eine Matrix von Funktionen ist, während das Maß m eindimensional ist. Wiederum ist das Integral komponentenweise zu verstehen, jetzt steht die Matrix allerdings *unter* dem Integralzeichen.

$$\gamma_\alpha(t) = \alpha^T \int e^{it\cdot} f \, dm \cdot \bar{\alpha} = \int e^{it\cdot} \alpha^T f \bar{\alpha} \, dm, \quad \forall t \in \mathbb{Z}.$$

Also ist $\alpha^T f \bar{\alpha}$ eine m -Dichte von $\alpha^T \mu \bar{\alpha}$, insbesondere gilt $\alpha^T f \bar{\alpha} \geq 0$ m -f.s. Da dies für jedes $\alpha \in \mathbb{C}^r$ gilt, folgt, dass für m -fast jedes ω die Matrix $f(\omega)$ positiv semidefinit ist. Wegen $\mu^* = \mu$ folgt außerdem

$$f_{jk} = \frac{d\mu_{jk}}{dm} = \frac{d\bar{\mu}_{kj}}{dm} = \overline{\left(\frac{d\mu_{kj}}{dm} \right)} = \bar{f}_{kj} \quad m - f.s., \quad \forall j \neq k, \quad (15.14)$$

und

$$f_{jj} = \frac{d\mu_{jj}}{dm} \geq 0 \quad m - f.s.. \quad (15.15)$$

Aus der positiven Semidefinitheit von

$$\begin{pmatrix} f_{jj}(\omega) & f_{jk}(\omega) \\ f_{kj}(\omega) & f_{kk}(\omega) \end{pmatrix} \quad (15.16)$$

und (15.14), sowie (15.15) folgt, wie wir gleich nachrechnen werden, die wichtige Ungleichung

$$|f_{jk}(\omega)|^2 \leq f_{jj}(\omega) \cdot f_{kk}(\omega) \quad \text{für } m\text{-fast alle } \omega. \quad (15.17)$$

Dies impliziert für $g \in L_2^{\mathbb{C}}(\mu_j)$, $h \in L_2^{\mathbb{C}}(\mu_k)$

$$\left| \int gh \, d\mu_{jk} \right|^2 \leq \int |g|^2 d\mu_j \cdot \int |h|^2 d\mu_k, \quad (15.18)$$

denn

$$\begin{aligned} \left| \int gh \, d\mu_{jk} \right|^2 &= \left| \int gh f_{jk} \, dm \right|^2 \leq \left(\int |g| |h| f_{jj}^{1/2} f_{kk}^{1/2} \, dm \right)^2 \\ &\leq \int |g|^2 f_{jj} \, dm \cdot \int |h|^2 f_{kk} \, dm = \int |g|^2 d\mu_j \cdot \int |h|^2 d\mu_k. \end{aligned}$$

Zum Nachweis von (15.17) führt man eine Links- und Rechtsmultiplikation der Matrix (15.16) mit $(1, x \cdot f_{jk})$ bzw. $(1, x \cdot f_{kj})^T$ und einem beliebigen $x \in \mathbb{R}$ aus. Dies ergibt wegen der positiven Semidefinitheit und wegen (15.14)

$$f_{jj} + 2x|f_{jk}|^2 + x^2|f_{jk}|^2 \cdot f_{kk} \geq 0, \quad \forall x \in \mathbb{R}.$$

Folglich ist die Diskriminante der linken Seite nichtnegativ, also

$$f_{kk} \cdot f_{jj} \cdot |f_{jk}|^2 \geq |f_{jk}|^4.$$

Hieraus folgt (15.17).

In Verallgemeinerung der univariaten Inversionsformel (Satz 3.8) beweisen wir nun

Satz 15.5 (*Multivariate Inversionsformel*)

Sei X eine schwach stationäre, zentrierte r -dimensionale Zeitreihe mit absolut summierbaren Komponenten der Kovarianz, d.h. mit

$$\sum_{h=-\infty}^{+\infty} |\gamma_{jk}(h)| < \infty, \quad \forall j, k = 1, \dots, r. \quad (15.19)$$

Dann konvergiert

$$f_{jk}(\omega) := \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\omega} \gamma_{jk}(h) \quad (15.20)$$

gleichmäßig auf $[-\pi, \pi]$ gegen eine stetige, komplexwertige Funktion f_{jk} auf $[-\pi, \pi]$. f_{jk} ist die bzgl. des Lebesgue-Maßes auf $[-\pi, \pi]$ f.s. eindeutig bestimmte Funktion mit

$$\mu_{jk}(B) = \int_B f_{jk}(\omega) d\omega, \quad \forall B \in \mathbb{B} \cap (-\pi, \pi]. \quad (15.21)$$

Es gilt weiter

$$\gamma_{jk}(h) = \int_{-\pi}^{\pi} e^{ih\omega} f_{jk}(\omega) d\omega, \quad \forall h \in \mathbb{Z}. \quad (15.22)$$

Man bezeichnet die f_{jk} generell als Spektraldichten, im Fall $j \neq k$ genauer als Kreuzspektraldichten. Die zugehörige Matrix (f_{jk}) nennt man Spektraldichte-Matrix.

Beweis: Offenbar impliziert (15.19) die gleichmäßige Konvergenz in (15.20). Wie im Beweis von Satz 15.4 betrachten wir $\mu_{\tau;jk}$ und $\gamma_{\tau;jk}$, $\tau = 1, 2$. Nach der univariaten Inversionsformel ist

$$f_{\tau;jk}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} e^{-ih\omega} \gamma_{\tau;jk}(h) \quad (15.23)$$

eine Lebesgue-Dichte des Maßes $\mu_{\tau;jk}$ und es gilt

$$\gamma_{\tau;jk}(h) = \int_{-\pi}^{\pi} e^{ih\omega} f_{\tau;jk}(\omega) d\omega. \quad (15.24)$$

Entsprechendes gilt für γ_{jj} , f_{jj} und γ_{kk} , f_{kk} . Eine Zerlegung von $\gamma_{jk}(h)$ gemäß (15.12) liefert sodann die Gleichheit

$$f_{jk} = \frac{1}{2}(f_{1;jk} - f_{jj} - f_{kk}) + \frac{1}{2}(f_{2;jk} - f_{jj} - f_{kk}) \quad (15.25)$$

sowie (15.21). Ebenfalls mit (15.12) ergibt sich aus (15.24)

$$\gamma_{jk}(h) = \int_{-\pi}^{\pi} e^{ih\omega} \{\text{rechte Seite von (15.25)}\} d\omega = \int_{-\pi}^{\pi} e^{ih\omega} f_{jk} d\lambda,$$

also (15.22). \square

Beispiel 15.6 (Spektraldichtematrix des weißen Rauschens)

Sei $e \sim WR(0, \Sigma)$ eine weißes Rauschen. Nach der Inversionsformel (15.20) ist die Spektraldichtematrix f_e von e gleich

$$f_e(\omega) = \frac{1}{2\pi} \Sigma, \quad \forall \omega \in [-\pi, \pi]. \quad (15.26)$$

\square

Im reellen Fall entnimmt man aus (15.20) unmittelbar den folgenden Zusammenhang zwischen den Spektraldichten f_{jk} und f_{kj} .

$$f_{jk}(\omega) = \overline{f_{jk}(-\omega)} = f_{kj}(-\omega) = \overline{f_{kj}(\omega)}, \quad \forall j, k \quad \forall \omega \in [-\pi, \pi]. \quad (15.27)$$

Schließlich besitzt auch der univariate Spektralsatz eine multivariate Verallgemeinerung.

Satz 15.7 (Multivariater Spektralsatz)

Sei X eine r -variante schwach stationäre, zentrierte Zeitreihe. Dann gibt es eine eindeutig bestimmte Spektraldarstellung

$$\begin{aligned} X(t) &= \left(\int e^{it\omega} dZ_1(\omega), \dots, \int e^{it\omega} dZ_r(\omega) \right)^T \\ &= \int e^{it\omega} dZ(\omega), \quad \forall t \in \mathbb{Z}, \end{aligned} \quad (15.28)$$

mit $Z = (Z_1, \dots, Z_r)^T$, wobei Z_j das zur j -ten Komponente $X_j = (X_j(t) : t \in \mathbb{Z})$ nach dem univariaten Spektralsatz gehörige MOW ist. Z heißt multivariates MOW.

Für alle $j, k = 1, \dots, r$, sowie $g \in L_2^{\mathbb{C}}(\mu_j)$ und $h \in L_2^{\mathbb{C}}(\mu_k)$ gilt

$$\left\langle \int g dZ_j, \int h dZ_k \right\rangle = \int g \bar{h} d\mu_{jk}, \quad (15.29)$$

insbesondere

$$\langle Z_j(A), Z_k(B) \rangle = \mu_{jk}(A \cap B), \quad \forall A, B \in \mathbb{B} \cap (-\pi, \pi]. \quad (15.30)$$

Beweis: Nur (15.29) ist zu beweisen. Dazu habe g zunächst die Gestalt

$$g(\omega) = \sum_{\alpha=1}^m a_{\alpha} e^{it_{\alpha}\omega} \text{ und } h \text{ die Gestalt } h(\omega) = \sum_{\beta=1}^n b_{\beta} e^{it_{\beta}\omega}. \text{ Dann ist}$$

$$\begin{aligned} \left\langle \int g dZ_j, \int h dZ_k \right\rangle &= \left\langle \sum_{\alpha} a_{\alpha} X_j(t_{\alpha}), \sum_{\beta} b_{\beta} X_k(t_{\beta}) \right\rangle \\ &= \sum_{\alpha} \sum_{\beta} a_{\alpha} \bar{b}_{\beta} \gamma_{jk}(t_{\alpha} - t_{\beta}) = \sum_{\alpha} \sum_{\beta} a_{\alpha} \bar{b}_{\beta} \int e^{i(t_{\alpha} - t_{\beta}) \cdot} d\mu_{jk} \\ &= \int \sum_{\alpha} \sum_{\beta} a_{\alpha} \bar{b}_{\beta} e^{i(t_{\alpha} - t_{\beta}) \cdot} d\mu_{jk} = \int \left(\sum_{\alpha} a_{\alpha} e^{it_{\alpha} \cdot} \right) \overline{\left(\sum_{\beta} b_{\beta} e^{it_{\beta} \cdot} \right)} d\mu_{jk} \\ &= \int g \bar{h} d\mu_{jk}. \end{aligned}$$

Nun seien g, h allgemein und g_n, h_n von obiger Form mit $g_n \rightarrow g$ in $L_2^{\mathbb{C}}(\mu_j)$, sowie $h_n \rightarrow h$ in $L_2^{\mathbb{C}}(\mu_k)$ (zur Existenz derartiger approximierender Folgen (g_n) und (h_n) beachte man Lemma 5.5). Wegen $\int g_n dZ_j \rightarrow \int g dZ_j$ in $L_2^{\mathbb{C}}(P)$ und entsprechend für h_n, h gilt dann einerseits unter Verwendung von Satz A.9

$$A_n := \left\langle \int g_n dZ_j, \int h_n dZ_k \right\rangle \rightarrow \left\langle \int g dZ_j, \int h dZ_k \right\rangle.$$

Andererseits gilt $A_n = \int g_n \bar{h}_n d\mu_{jk} \rightarrow \int g \bar{h} d\mu_{jk}$, denn

$$\begin{aligned} \left| \int (g_n \bar{h}_n - g \bar{h}) d\mu_{jk} \right| &\leq \left| \int (g_n - g) \bar{h}_n d\mu_{jk} \right| + \left| \int g (\bar{h}_n - \bar{h}) d\mu_{jk} \right| \\ &\leq \left(\int |g_n - g|^2 d\mu_j \int |h_n|^2 d\mu_k \right)^{1/2} + \left(\int |g|^2 d\mu_j \int |h_n - h|^2 d\mu_k \right)^{1/2} \\ &\rightarrow 0 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Für das letzte Ungleichheitszeichen beachte man (15.18).

□

Eine unmittelbare Folge von (15.30) ist

$$|\mu_{jk}(A)| = |\langle Z_j(A), Z_k(A) \rangle| \leq \|Z_j(A)\| \cdot \|Z_k(A)\| \quad (15.31)$$

$$= (\mu_j(A) \cdot \mu_k(A))^{1/2}, \quad \forall A \in \mathbb{B} \cap (-\pi, \pi],$$

wobei $\|\cdot\|$ die Norm in $L_2^{\mathbb{C}}(P)$ ist. Nach (15.31) ist $\mu_{jk}(A)$ höchstens dann von Null verschieden, wenn dies sowohl für $\mu_j(A)$ als auch für $\mu_k(A)$ gilt.

Beispiel 15.8 (Zwei zyklische Zeitreihen)

Für festes $\omega_0 \in (0, \pi)$ seien zwei zyklische Zeitreihen

$$X_j(t) = e_{j1} \cos(\omega_0 t) + e_{j2} \sin(\omega_0 t), \quad t \in \mathbb{Z}, \quad j = 1, 2, \quad (15.32)$$

gegeben mit zentrierten, reellen Zufallsvariable e_{j1} und e_{j2} , die für festes j unkorreliert sind und dieselbe Varianz $\sigma_j^2 > 0$ haben. Weiter gelte für $j \neq k$

$$\text{Cov}(e_{j1}, e_{k1}) = \text{Cov}(e_{j2}, e_{k2}) =: \sigma_{jk} \quad (15.33)$$

und

$$\text{Cov}(e_{j1}, e_{k2}) = -\text{Cov}(e_{k1}, e_{j2}) =: \tau_{jk}. \quad (15.34)$$

Also ist σ_{12} die Kovarianz der Amplituden der „in Phase“ befindlichen Anteile von X_1 und X_2 (d.h. der Kosinus-Anteil von X_1 verglichen mit dem Kosinus-Anteil von X_2 und entsprechend mit den Sinus-Anteilen), während τ_{12} die Kovarianz der Amplituden der „phasenverschobenen“ Anteile von X_1 und X_2 ist (d.h. der Kosinus-Anteil von X_1 verglichen mit dem Sinus-Anteil von X_2 , bzw. mit negativem Vorzeichen bei Vertauschung von X_1 und X_2).

Dann ist leicht zu sehen (s. Aufgabe 15.4), dass die zweidimensionale Zeitreihe $X(t) = (X_1(t), X_2(t))^T, t \in \mathbb{Z}$, schwach stationär ist mit den Autokovarianzen

$$\begin{aligned} \gamma_{11}(h) &= \sigma_1^2 \cos(\omega_0 h) \\ \gamma_{22}(h) &= \sigma_2^2 \cos(\omega_0 h), \\ \gamma_{12}(h) &= \sigma_{12} \cos(\omega_0 h) + \tau_{12} \sin(\omega_0 h). \end{aligned} \quad (15.35)$$

Die zugehörige Spektralmaße μ_{11} , μ_{22} und μ_{12} sind diskret mit den beiden Trägerpunkten ω_0 und $-\omega_0$. Es gilt für die zugehörigen Zähldichten f_{11} , f_{22} und f_{12}

$$\begin{aligned} f_{11}(\omega) &= \frac{1}{2}(1(\omega = -\omega_0) + 1(\omega = \omega_0))\sigma_1^2, \\ f_{22}(\omega) &= \frac{1}{2}(1(\omega = -\omega_0) + 1(\omega = \omega_0))\sigma_2^2, \\ f_{12}(\omega) &= \frac{1}{2}(1(\omega = -\omega_0) + 1(\omega = \omega_0))\sigma_{12} + i\frac{1}{2}(1(\omega = -\omega_0) - 1(\omega = \omega_0))\tau_{12}. \end{aligned} \quad (15.36)$$

Den Realteil von $f_{12}(\omega)$ (also im Wesentlichen $\sigma_{12}/2$) nennt man Kospektrum und den Imaginärteil von f_{12} (also im Wesentlichen $\pm\tau_{12}/2$) nennt man Quadratspektrum. Wir haben diese Begriffe schon jetzt eingeführt, um später in Abschnitt 15.3, in dem es um stetige Spektraldichten f_{jk} geht, eine Interpretation des Ko- und Quadratspektrums zur Verfügung zu haben. \square

15.2 Multivariate Filter

Sei $Y = (Y_t : t \in \mathbb{Z})$ eine r -variate, stationäre, zentrierte komplexe Zeitreihe mit Spektralmatrix $\mu^Y = (\mu_{ij}^Y)_{i,j=1,\dots,r}$. Sei nun $\varphi = (\varphi_{\kappa j})_{\kappa=1,\dots,k,j=1,\dots,r}$ eine Matrix von Funktionen $\varphi_{\kappa j} \in L_2^{\mathbb{C}}(\mu_j^Y) \forall \kappa, j$, wobei $\mu_j^Y = \mu_{jj}^Y$ das Spektralmaß der j -ten Komponente von Y ist. Wir definieren nun eine k -variate Zeitreihe $X = (X(t) : t \in \mathbb{Z})$ durch

$$\begin{aligned} X(t) &= \left(\sum_{j=1}^r \int e^{it \cdot} \varphi_{\kappa j} dZ_j^Y \right)_{\kappa=1,\dots,k} \\ &=: \int e^{it \cdot} \varphi dZ^Y, \quad \forall t \in \mathbb{Z}. \end{aligned} \quad (15.37)$$

Offenbar ist $X(t) \in (L_2^{\mathbb{C}}(P))^k$ mit $EX(t) = 0 \in \mathbb{C}^k$. Unter Verwendung von (15.29) erhält man leicht

$$EX(t+h)\overline{X}(t)^T = \left(\sum_{j=1}^r \sum_{j'=1}^r \int_{(-\pi,\pi]} \varphi_{\kappa j} \overline{\varphi}_{\kappa' j'} e^{ih \cdot} d\mu_{jj'}^Y \right)_{\kappa,\kappa'=1,\dots,k} \quad (15.38)$$

woraus sich die schwache Stationarität von X ergibt. Sind $f_{jj'}^Y$, m -Dichten von $\mu_{jj'}^Y$, wie im Beweis von Satz 15.4 und $f^Y = (f_{jj'}^Y)$, so schreibt sich (15.38) in Matrixform als

$$\begin{aligned} \Gamma_X(h) &= EX(t+h)\overline{X}(t)^T = \int e^{ih \cdot} \varphi f \varphi^* dm \\ &=: \int e^{ih \cdot} \varphi d\mu^Y \varphi^*, \quad \forall h \in \mathbb{Z}. \end{aligned} \quad (15.39)$$

Da in (15.39) das Integral nicht vom dominierenden Maß m abhängt, ist die symbolische Schreibweise im letzten Ausdruck gerechtfertigt. Die zu X gehörige Spektralmatrix μ^X ist wegen (15.39)

$$\mu^X(B) = \int_B \varphi f \varphi^* dm =: \int_B \varphi d\mu^Y \varphi^*, \quad \forall B \in \mathbb{B} \cap (-\pi, \pi], \quad (15.40)$$

und das MOW zu X ist für $B \in \mathbb{B} \cap (-\pi, \pi]$

$$Z^X(B) = \left(\sum_{j=1}^r \int 1_B \varphi_{\kappa j} dZ_j^Y \right)_{\kappa=1,\dots,k} =: \int 1_B \varphi dZ^Y. \quad (15.41)$$

Es gilt nämlich unter Verwendung von (15.29) offenbar für $\kappa = 1, \dots, k$ und $A, B \in \mathbb{B} \cap (-\pi, \pi]$

$$\langle Z_\kappa^X(A), Z_\kappa^X(B) \rangle = \mu_\kappa^X(A \cap B). \quad (15.42)$$

Weiter gilt für jede komplexwertige Funktion g auf $(-\pi, \pi]$ mit $g \cdot \varphi_{\kappa j} \in L_2^{\mathbb{C}}(\mu_j^Y) \forall \kappa, j$ die Gleichung

$$\int g dZ_\kappa^X = \sum_{j=1}^r \int g \varphi_{\kappa j} dZ_j^Y, \quad \kappa = 1, \dots, k, \quad (15.43)$$

wie sich aus (15.41) über den Aufbau messbarer Funktionen ergibt. Hieraus ergibt sich die multivariate Substitutions- oder Kettenregel für das Hintereinanderschalten von Filtern.

Satz 15.9 (*Multivariate Substitutionsregel*)

Sei $Y = (Y(t) : t \in \mathbb{Z})$ eine r -variante stationäre, zentrierte komplexe Zeitreihe mit Spektralmatrix μ^Y und $\varphi = (\varphi_{\kappa j})_{\kappa=1, \dots, k, j=1, \dots, r}$ eine Filtermatrix mit $\varphi_{\kappa j} \in L_2^{\mathbb{C}}(\mu_j^Y) \forall \kappa, j$. Sei weiter eine Filtermatrix $\psi = (\psi_{\lambda \kappa})_{\lambda=1, \dots, l, \kappa=1, \dots, k}$ gegeben mit $\psi_{\lambda \kappa} \varphi_{\kappa j} \in L_2^{\mathbb{C}}(\mu_j^Y) \forall \lambda, \kappa, j$. Dann gilt mit der Zeitreihe X aus (15.37) die Gleichheit

$$\int \psi dZ^X = \int \psi \cdot \varphi dZ^Y. \quad (15.44)$$

Der Beweis folgt aus (15.43), indem $g = \psi_{\lambda \kappa}$ gesetzt und über κ summiert wird.

Schließlich erhalten wir auch ein Analogon zu Satz 6.3.

Satz 15.10 (*Multivariate Filter in Reihendarstellung*)

Falls Matrizen $c_s \in \mathbb{C}^{k \times r}$, $s \in \mathbb{Z}$, existieren derart, dass für die $k \times r$ -Matrix $\varphi = (\varphi_{\kappa j})_{\kappa=1, \dots, k, j=1, \dots, r}$ die Reihenentwicklungen

$$\varphi_{\kappa j}(\omega) = \sum_{s=-\infty}^{+\infty} c_s(\kappa, j) e^{-is\omega} \quad (15.45)$$

in $L_2^{\mathbb{C}}(\mu_j^Y)$ gelten $\forall \kappa, j$, so folgt die Darstellung

$$X_t = \sum_{s=-\infty}^{+\infty} c_s Y_{t-s}, \quad \forall t \in \mathbb{Z}, \quad (15.46)$$

wobei die Konvergenz komponentenweise in $L_2^{\mathbb{C}}(P)$ zu verstehen ist.

Der Beweis folgt aus der univariaten Version (Satz 6.3) gemäß $\int \varphi_{\kappa j} e^{it\cdot} dZ_j^Y = \sum_{s=-\infty}^{+\infty} c_s(\kappa, j) Y_j(t-s)$ durch Summation über j .

15.3 Der quadratische Kohärenzkoeffizient und verwandte Größen

Wird eine r -dimensionale Zeitreihe $X = (X_t : t \in \mathbb{Z})$ betrachtet, so interessiert natürlich die Frage, wie sich die Abhängigkeit zwischen zwei Komponenten $X_j(t)$ und $X_k(t)$, $j \neq k$, der Zeitreihe messen läßt. Hierzu stellen wir die folgenden Überlegungen an.

Unter der Voraussetzung, dass X eine r -variate, stationäre, zentrierte und komplexe Zeitreihe mit $r \geq 2$ ist, stellen wir uns die Frage, wie man durch Anwendung eines univariaten Filters $\varphi \in L_2^{\mathbb{C}}(\mu_j)$ auf X_j die Reihe $(X_k(t) : t \in \mathbb{Z})$ möglichst gut approximieren kann. Wir suchen also ein φ derart, dass der Anpassungsfehler

$$\|X_k(t) - \int e^{it\cdot} \varphi dZ_j\|^2 =: \Delta(\varphi) \quad (15.47)$$

minimal wird. Dabei bezeichnet $(Z_1, \dots, Z_r)^T$ das zu X gehörige MOW mit Spektralmatrix $\mu = (\mu_{jk})$, $\mu_j = \mu_{jj}$, $\mu_k = \mu_{kk}$. Ist m ein dominierendes Maß wie in den Ausführungen im Anschluss an Satz 15.4 und sei $f_{jk} = d\mu_{jk}/dm$, so erhält man unter Verwendung von (15.29) und (15.14)

$$\begin{aligned} \Delta(\varphi) &= \int f_{kk} dm + \int |\varphi|^2 f_{jj} dm - \int \bar{\varphi} f_{kj} dm - \int \varphi \bar{f}_{kj} dm \\ &= \int f_{kk} dm + \int |\varphi - \frac{f_{kj}}{f_{jj}}|^2 f_{jj} dm - \int \frac{|f_{kj}|^2}{f_{jj}} dm \\ &= \int \left(1 - \frac{|f_{kj}|^2}{f_{kk} f_{jj}}\right) d\mu_k + \int |\varphi - \frac{f_{kj}}{f_{jj}}|^2 d\mu_j. \end{aligned} \quad (15.48)$$

Dabei erstrecken sich alle Integrale über $(-\pi, \pi]$. Man verwende außerdem die Konvention $0/0 = 0$, sowie die Ungleichung (15.17). Offenbar wird $\Delta(\varphi)$ minimal durch Wahl von $\varphi_{opt} = \frac{f_{kj}}{f_{jj}}$. Dabei gehört f_{kj}/f_{jj} wegen (15.17) zu $L_2^{\mathbb{C}}(\mu_j)$.

Der sich ergebende optimal approximierende Prozess $\int e^{it\cdot} (f_{kj}/f_{jj}) dZ_j$ hat die Spektraldichte $|f_{kj}|^2/f_{jj}$. Das Verhältnis dieser Spektraldichte zur Spektraldichte von (X_k) , also

$$K_{jk}^2(\omega) := \frac{|f_{kj}(\omega)|^2}{f_{kk}(\omega) f_{jj}(\omega)} = \frac{|f_{jk}(\omega)|^2}{f_{jj}(\omega) f_{kk}(\omega)}, \quad \omega \in (-\pi, \pi), \quad (15.49)$$

nennt man *quadratischen Kohärenzkoeffizienten* (oder auch quadratische Kohärenz), wobei in (15.49) $K_{jk}^2(\omega)$ zu Null gesetzt wird, wenn der Nenner verschwindet.

Da $E|X_k(t)|^2 = \int_{(-\pi, \pi]} f_{kk}(\omega) m(d\omega)$ die Varianz des Prozesses X_k ist und $\int_{(-\pi, \pi]} |f_{kj}(\omega)|^2 / f_{jj}(\omega) m(d\omega)$ die Varianz des optimal approximierenden Prozesses (auf Basis von (X_j)), kann man $K_{jk}^2(\omega)$ als den Varianzanteil von X_k

zur Frequenz ω interpretieren, der durch X_j (linear) erklärt werden kann.

Offenbar ist K_{jk}^2 von der Wahl des dominierenden Maßes m unabhängig. Im Grunde wird (15.49) nur verwendet, wenn m das Lebesgue-Maß ist, also wenn Spektraldichten existieren, z.B. wie im Fall der multivariaten Inversionsformel (Satz 15.5). Analog zum univariaten Fall wählt man dann aus Symmetriegründen meist das abgeschlossene Intervall $[-\pi, \pi]$ als Definitionsbereich der f_{jk} . Nach (15.17) gilt für m -fast alle ω

$$0 \leq K_{jk}^2(\omega) \leq 1 \quad (15.50)$$

und aus (15.47) ergibt sich für den Anpassungsfehler bei Verwendung des optimalen Filters φ_{opt}

$$\Delta(\varphi_{opt}) = \int_{(-\pi, \pi]} (1 - K_{jk}^2) d\mu_k, \quad (15.51)$$

wobei der Integrand wegen (15.50) nichtnegativ ist. Mit der obigen Bezeichnung liest man nun aus der Formel (15.48) leicht das folgende Resultat ab.

Satz 15.11

Genau dann existiert ein Filter $\varphi \in L_2^{\mathbb{C}}(\mu_j)$ mit

$$X_k(t) = \int e^{it \cdot} \varphi dZ_j, \quad \forall t \in \mathbb{Z}, \quad (15.52)$$

wenn gilt

$$K_{jk}^2(\omega) = 1 \text{ für } \mu_k\text{-fast alle } \omega \in (-\pi, \pi]. \quad (15.53)$$

Beweis: Die Gleichung (15.52) ist äquivalent mit $\Delta(\varphi) = 0$. Nach (15.48) ergibt sich daraus $\int_{(-\pi, \pi]} (1 - K_{jk}^2) d\mu_k = 0$; wegen $K_{jk}^2 \leq 1$ m -f.s., folgt (15.53). Umgekehrt folgt aus (15.53) mit $\varphi = \varphi_{opt}$ aus (15.51) sofort $\Delta(\varphi_{opt}) = 0$. \square

Der Maximalwert 1 der quadratischen Kohärenz für alle Frequenzen geht also einher mit der Möglichkeit, X_k mit Hilfe eines Filters aus X_j zu rekonstruieren. Das andere Extrem, nämlich $K_{jk}^2(\omega) = 0$ für m -fast alle ω , bedeutet gemäß Satz 15.4 gerade die Unkorreliertheit von $X_k(t)$ und $X_j(s) \quad \forall s, t \in \mathbb{Z}$.

Die quadratische Kohärenz zwischen X_j und X_k ändert sich nicht, wenn beide Komponenten nicht auslöschend gefiltert werden. Genauer gilt mit obigen Bezeichnungen

Satz 15.12

Seien $\varphi_j \in L_2^{\mathbb{C}}(\mu_j)$ und $\varphi_k \in L_2^{\mathbb{C}}(\mu_k)$. Dann ist die quadratische Kohärenz der gefilterten Zeitreihen $Y_j(t) = \int e^{it \cdot} \varphi_j dZ_j$ und $Y_k(t) = \int e^{it \cdot} \varphi_k dZ_k$ dieselbe wie die der ursprünglichen Zeitreihen X_j und X_k an allen Stellen ω mit $\varphi_j(\omega) \cdot \varphi_k(\omega) \neq 0$.

Beweis: Aufgabe 15.2 . □

Wir wollen nun noch einige weitere Größen, die im Zusammenhang mit Kreuzspektraldichten gebräuchlich sind, vorstellen. Dazu setzen wir die Situation der multivariaten Inversionsformel (Satz 15.5) voraus. Darüberhinaus sei X reell, sodass alle $\gamma_{jk}(h)$ ebenfalls reell sind. In Fragen von praktischem Interesse sind diese Voraussetzungen in der Regel erfüllt. Ein dominierendes Maß m ist in diesem Fall das Lebesgue-Maß auf $(-\pi, \pi]$ und Radon-Nikodym Ableitungen f_{jk} können nach Formel (15.20) berechnet werden und sind dann stetig auf $(-\pi, \pi]$.

Da f_{jk} (im Folgenden sei stets $j \neq k$) im Allgemeinen komplexwertig ist, geht man meist äquivalent zur Angabe des Real- und Imaginärteils über. Schreiben wir dazu in (15.20) $e^{-ih\omega} = \cos h\omega - i \sin h\omega$ und zerlegen γ_{jk} in die Differenz einer geraden und ungeraden Funktion gemäß

$$\gamma_{jk}(h) = \frac{\gamma_{jk}(h) + \gamma_{jk}(-h)}{2} - \frac{\gamma_{jk}(-h) - \gamma_{jk}(h)}{2}, \quad (15.54)$$

so ergibt sich die gewünschte Zerlegung von f_{jk} in Realteil c_{jk} und Imaginärteil q_{jk} gemäß

$$f_{jk}(\omega) = c_{jk}(\omega) + i q_{jk}(\omega) \quad (15.55)$$

mit

$$c_{jk}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \cos(\omega h) \left\{ \frac{\gamma_{jk}(h) + \gamma_{jk}(-h)}{2} \right\} \quad (15.56)$$

und

$$q_{jk}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{+\infty} \sin(\omega h) \left\{ \frac{\gamma_{jk}(-h) - \gamma_{jk}(h)}{2} \right\}. \quad (15.57)$$

Man nennt c_{jk} das *Kospektrum* und q_{jk} das *Quadraturspektrum* von f_{jk} . In einigen Lehrbüchern, z.B. Brockwell/Davis (1991), wird der Negativwert von q_{jk} als Quadraturspektrum bezeichnet; wir folgen in der Bezeichnung aber Hamilton (1994). Man liest aus (15.56) und (15.57) unmittelbar ab, dass das Kospektrum eine gerade Funktion und das Quadraturspektrum eine ungerade Funktion ist, also dass

$$c_{jk}(\omega) = c_{jk}(-\omega) \text{ und } q_{jk}(\omega) = -q_{jk}(-\omega), \quad \forall \omega \in [-\pi, \pi], \quad (15.58)$$

gilt. Hieraus folgt unmittelbar

$$\text{Cov}(X_j(t), X_k(t)) = \int_{-\pi}^{\pi} c_{jk}(\omega) d\omega, \quad (15.59)$$

sodass das Kospektrum $c_{jk}(\omega)$ als der Kovarianzanteil von $X_j(t)$ und $X_k(t)$ zur Frequenz ω interpretiert werden kann.

Beispiel 15.13 In Fortsetzung von Beispiel 15.2 erhält man durch Berechnung von $f_{12} = c_{12} + i q_{12}$ mit Hilfe der Inversionsformel (15.20) oder auch mit Hilfe von (15.56) und (15.57)

$$\begin{aligned} c_{12}(\omega) &= \frac{1}{2\pi}(1 + b \cos \omega) \\ q_{12}(\omega) &= \frac{b}{2\pi} \sin \omega, \quad \forall \omega \in [-\pi, \pi]. \end{aligned}$$

□

Eine gewisse Interpretation von Ko- und Quadratspektrum wird durch die folgenden Überlegungen gegeben

Bemerkung 15.14 (Ko- und Quadratspektrum)

Seien für $j = 1, 2$ reelle, zentrierte, stationäre Zeitreihen $(X_j(t) : t \in \mathbb{Z})$ gegeben, für die die Voraussetzungen der Inversionsformel vorliegen. Nach Aufgabe 6.6 gibt es für $j = 1, 2$ reelle Spektraldarstellungen

$$X_j(t) = 2 \int_{(0,\pi)} \cos(t\omega) dU_j(\omega) + 2 \int_{(0,\pi)} \sin(t\omega) dV_j(\omega), \quad \forall t \in \mathbb{Z}.$$

Dabei sind U_j und V_j , reellwertige, orthogonale MOW's auf $(0, \pi)$. Die Kovarianz der Amplituden „der in Phase schwingenden Komponenten“ $\cos(\omega t) dU_1(\omega)$ und $\cos(\omega t) dU_2(\omega)$ sowie $\sin(\omega t) dV_1(\omega)$ und $\sin(\omega t) dU_2(\omega)$ ist (zunächst in informeller Schreibweise; eine mathematisch exakte Formulierung findet sich in Aufgabe 15.3. Man vergleiche auch die Überlegungen in Beispiel 15.8, in dem die f_{jk} Zähldichten sind.)

$$\langle dU_1(\omega), dU_2(\omega) \rangle = \langle dV_1(\omega), dV_2(\omega) \rangle = \frac{1}{2} c_{12}(\omega) d\omega,$$

während die Kovarianz der Amplituden der „phasenverschobenen Komponenten“ $\cos(\omega t) dU_1(\omega)$ und $\sin(\omega t) dV_2(\omega)$ sowie $\cos(\omega t) dU_2(\omega)$ und $\sin(\omega t) dV_1(\omega)$ gleich

$$\langle dU_1(\omega), dV_2(\omega) \rangle = -\langle dU_2(\omega), dV_1(\omega) \rangle = \frac{1}{2} q_{12}(\omega) d\omega,$$

ist. Damit ist

$$c_{12}(\omega) d\omega = \langle dU_1(\omega), dU_2(\omega) \rangle + \langle dV_1(\omega), dV_2(\omega) \rangle,$$

die Summe der Gewichte der Kovarianzen der in Phase schwingenden Anteile der Reihen X_1 und X_2 zur Frequenz ω und $q_{12}(\omega)$ entsprechend die Differenz der Gewichte der Kovarianzen für die phasenverschobenen Anteile, also

$$q_{12}(\omega) d\omega = \langle dU_1(\omega), dV_2(\omega) \rangle - \langle dU_2(\omega), dV_1(\omega) \rangle.$$

□

Eine Zerlegung von f_{jk} in Polarkoordinaten liefert

$$f_{jk}(\omega) = \alpha_{jk}(\omega) e^{i\vartheta_{jk}(\omega)} \quad (15.60)$$

mit dem sogenannten *Amplitudenspektrum*

$$\alpha_{jk}(\omega) = (c_{jk}^2(\omega) + q_{jk}^2(\omega))^{1/2} = |f_{jk}(\omega)|$$

und dem sogenannten *Phasenspektrum*

$$\vartheta_{jk}(\omega) = \arg(f_{jk}(\omega)) \in (-\pi, \pi]$$

von X_j und X_k , wobei $\arg(z)$ die Argumentfunktion ist, die einer komplexen Zahl in Polarkoordinatenform $z = re^{i\varphi}$ mit $r > 0$, $\varphi \in (-\pi, \pi]$, den Wert φ zuordnet.

Im folgenden Beispiel werden wir sehen, dass sich eine Zeitverzögerung der Komponente $X_k(t)$ gegenüber $X_j(t)$ in einer positiven Steigung von $\vartheta_{jk}(\omega)$ widerspiegelt.

Beispiel 15.15 (Zeitverzögerung und Phasenspektrum)

Wir nehmen an, dass $(X_k(t))$ eine zeitverzögerte Version von $(X_j(t))$ ist, d.h. für $d > 0$ gelte $X_k(t) = X_j(t - d)$. Wir erhalten dann die Darstellung

$$X_k(t) = \int e^{i(t-d)\omega} dZ_j(\omega) \quad (15.61)$$

und damit

$$\gamma_{jk}(h) = \int_{-\pi}^{\pi} e^{i(h+d)\omega} f_{jj}(\omega) d\omega. \quad (15.62)$$

Also ist $f_{jk}(\omega) = e^{id\omega} f_{jj}(\omega)$. Während das Amplituden-Spektrum $|f_{jk}(\omega)|$ unabhängig von d ist, erhält man als Phasen-Spektrum den Ausdruck $\vartheta_{jk}(\omega) = \arg(e^{id\omega}) = d\omega - j2\pi$, wobei $j \in \mathbb{Z}$ in Abhängigkeit von ω so bestimmt ist, dass $d\omega - j2\pi \in (-\pi, \pi]$ gilt. Das Phasenspektrum ist also stückweise linear mit einer Steigung, die gleich der Zeitverschiebung d ist. \square

Beispiel 15.16 (Ein ökonomisches Modell von Preis und Angebot)

$Y_1(t)$ beschreibe den vom Mittelwert abweichenden Preis einer Ware und $Y_2(t)$ das vom Mittelwert abweichende Angebot dieser Ware zur Zeit t . Dabei sollen folgende Beziehungen gelten.

$$\begin{aligned} Y_1(t) &= -a_1 Y_2(t) + e_1(t) \\ Y_2(t) &= a_2 Y_1(t-1) + e_2(t) \end{aligned} \quad (15.63)$$

mit Konstanten $0 < a_1, a_2 < 1$ und zwei untereinander unkorrelierten weißen Rauschen $e_j(t) \sim (0, \sigma_j^2)$, $j = 1, 2$. Die Gleichungen (15.63) können als

zweidimensionale Filtergleichungen geschrieben werden. Man definiere dazu die Filtermatrix $\varphi = (\varphi_{\kappa j})_{\kappa, j=1,2}$ gemäß $\varphi(\omega) = A(e^{-i\omega})$ und

$$A(z) = \begin{pmatrix} 1 & a_1 \\ -a_2 z & 1 \end{pmatrix}.$$

Mit $Y(t) = (Y_1(t), Y_2(t))^T$, $e(t) = (e_1(t), e_2(t))^T$ und den zugehörigen MOW's $Z^Y = (Z_1^Y, Z_2^Y)^T$, $Z^e = (Z_1^e, Z_2^e)^T$ ist (15.63) gleichbedeutend mit

$$\int e^{it \cdot} \varphi dZ^Y = e(t), \quad \forall t \in \mathbb{Z}. \quad (15.64)$$

Da die Determinante $\det A(z) = 1 + a_1 a_2 z$ für $|z| < (a_1 a_2)^{-1}$ nicht verschwindet, insbesondere auch nicht für alle $|z| = 1$, existiert $\varphi^{-1}(\omega) \quad \forall \omega \in [-\pi, \pi]$ und alle Komponenten von φ^{-1} sind beschränkte Funktionen, so dass $Y(t) := \int e^{it \cdot} \varphi^{-1} dZ^e$ existiert. Eine Anwendung der multivariaten Substitutionsregel (vgl. Satz 15.9) zeigt, dass

$$\int e^{it \cdot} \varphi dZ^Y = \int e^{it \cdot} \varphi \cdot \varphi^{-1} dZ^e = \int e^{it \cdot} dZ^e = e(t)$$

gilt. Somit löst die oben definierte Reihe $Y = (Y(t) : t \in \mathbb{Z})$ die Gleichungen (15.63) bzw. (15.64). Die Lösung ist auch eindeutig, wie eine Anwendung des Filters φ^{-1} auf (15.64) zeigt.

Mit der Diagonalmatrix $f^e(\omega) = \frac{1}{2\pi} \text{diag}(\sigma_1^2, \sigma_2^2)$ der Spektraldichten des weißen Rauschens e ergibt (15.39) als Matrix der Spektraldichten $f^Y(\omega)$

$$\begin{aligned} f^Y(\omega) &= \varphi^{-1}(\omega) f^e(\omega) (\varphi^{-1}(\omega))^* \\ &= \frac{1}{2\pi} \frac{1}{|1 + a_1 a_2 e^{-i\omega}|^2} \begin{pmatrix} \sigma_1^2 + \sigma_2^2 a_1^2 & \sigma_1^2 a_2 e^{i\omega} - \sigma_2^2 a_1 \\ \sigma_1^2 a_2 e^{-i\omega} - \sigma_2^2 a_1 & \sigma_1^2 a_2^2 + \sigma_2^2 \end{pmatrix}. \end{aligned} \quad (15.65)$$

Der quadratische Kohärenzkoeffizient ergibt sich demnach zu

$$\begin{aligned} K_{12}^2(\omega) &= \frac{|\sigma_1^2 a_2 e^{i\omega} - \sigma_2^2 a_1|^2}{(\sigma_1^2 + \sigma_2^2 a_1^2)(\sigma_1^2 a_2^2 + \sigma_2^2)} \\ &= \frac{\sigma_1^4 a_2^2 + \sigma_2^4 a_1^2 - 2\sigma_1^2 \sigma_2^2 a_1 a_2 \cos \omega}{(\sigma_1^2 + \sigma_2^2 a_1^2)(\sigma_1^2 a_2^2 + \sigma_2^2)} \end{aligned} \quad (15.66)$$

und das Phasenspektrum lautet

$$\vartheta_{12}(\omega) = \arg(\sigma_1^2 a_2 e^{i\omega} - \sigma_2^2 a_1). \quad (15.67)$$

Die quadratische Kohärenz wird maximal bei $\omega = \pi$, d.h. dass der lineare Zusammenhang zwischen Preis und Angebot bei hohen Frequenzen am stärksten ist. Man überlegt sich leicht, dass $\vartheta_{12}(\omega)$ in der Nähe von π monoton steigend ist, die Ableitung ϑ'_{12} in diesem Bereich also positiv ist. Dies kann in Analogie zu Beispiel 15.15 als eine zeitliche Verzögerung der Warenmenge gegenüber dem Preis für ω in der Nähe von π gedeutet werden.

Um festzustellen, ob dieses Modell in realen Situationen brauchbar ist, benötigt man auf Beobachtungen basierende Schätzwerte der Spektraldichten. Hierzu dienen die Überlegungen im folgenden Abschnitt 15.4 \square

15.4 Schätzer der Spektraldichtematrix

In diesem Abschnitt wollen wir kurz vorstellen, wie die Spektraldichten und Kreuzspektraldichten multivariater Zeitreihen geschätzt werden können. Dabei gehen wir im Rahmen dieser Einführung jedoch nicht auf Beweise und technische Einzelheiten ein. Tatsächlich kann man vollständig analog zum univariaten Fall vorgehen und entsprechend Gleichung (12.46) Lag-Window-Schätzer für f_{jk} definieren gemäß

$$\hat{f}_{n,jk}(\omega) = \frac{1}{2\pi} \sum_{|h| \leq M(n)} k\left(\frac{h}{M(n)}\right) \bar{\gamma}_{jk}(h) e^{-ih\omega}, \quad 0 \leq \omega \leq \pi, \quad (15.68)$$

mit $\bar{\gamma}_{jk}(h) = 1/n \sum_{t=1}^{n-|h|} X_j(t) X_k(t+|h|)$. Da f_{jk} für $j \neq k$ im Allgemeinen komplex ist, schätzt man meist den Realteil c_{jk} und Imaginärteil q_{jk} , also reelle Größen gemäß $\hat{c}_{jk} = \mathcal{Re} \hat{f}_{jk}$ und $\hat{q}_{jk} = \mathcal{Im} \hat{f}_{jk}$. Ausgeschrieben bedeutet dies für $0 \leq \omega \leq \pi$

$$\hat{c}_{n,jk}(\omega) = \frac{1}{2\pi} \sum_{|h| \leq M(n)} k\left(\frac{h}{M(n)}\right) \cos(\omega h) \frac{\bar{\gamma}_{jk}(h) + \bar{\gamma}_{jk}(-h)}{2} \quad (15.69)$$

$$\hat{q}_{n,jk}(\omega) = \frac{1}{2\pi} \sum_{|h| \leq M(n)} k\left(\frac{h}{M(n)}\right) \sin(\omega h) \frac{\bar{\gamma}_{jk}(-h) - \bar{\gamma}_{jk}(h)}{2}. \quad (15.70)$$

Unter entsprechenden Voraussetzungen wie im univariaten Fall kann gezeigt werden (siehe Hannan (1970), S. 289), dass für $j \neq k$ der vierdimensionale reelle Vektor $\hat{H}_n(\omega) = (\hat{f}_{jj}(\omega), \hat{f}_{kk}(\omega), \hat{c}_{jk}(\omega), \hat{q}_{jk}(\omega))^T$ für $\omega \in [0, \pi]$ eine vierdimensionale asymptotische Normalverteilung besitzt mit Mittelwertvektor $H(\omega) := (f_{jj}(\omega), f_{kk}(\omega), c_{jk}(\omega), q_{jk}(\omega))^T$ und Kovarianzmatrix $(M(n)/n) \tau^2(\omega) \cdot V(\omega)$ mit $\tau^2(\omega)$ aus (12.60) und V (unter Auslassung des Arguments ω) gemäß

$$V = \begin{pmatrix} f_{jj}^2 & |f_{jk}|^2 & f_{jj}c_{jk} & f_{jj}q_{jk} \\ & f_{kk}^2 & f_{kk}c_{kj} & f_{kk}q_{jk} \\ & & \frac{1}{2}(f_{jj}f_{kk} + c_{jk}^2 - q_{jk}^2) & c_{jk}q_{jk} \\ \text{symmetrisch} & & & \frac{1}{2}(f_{jj}f_{kk} - c_{jk}^2 + q_{jk}^2) \end{pmatrix}. \quad (15.71)$$

Genauer gilt für $n \rightarrow \infty$

$$\sqrt{\frac{n}{M(n)}} (\hat{H}_n(\omega) - H(\omega)) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \tau^2(\omega) V(\omega)). \quad (15.72)$$

Wie in Abschnitt 12.6 können punktweise Konfidenzintervalle für das Ko-Spektrum, das Quadratur-Spektrum, den quadratischen Kohärenzkoeffizienten, sowie Amplituden- und Phasenspektrum (mit Hilfe der Delta-Methode)

hergeleitet werden. Wir verzichten hier auf eine explizite Darstellung und verweisen z.B. auf Brockwell/Davis (1991). Vielmehr wollen wir uns in dem folgenden Abschnitt mit der multivariaten Verallgemeinerung der ARMA-Modelle befassen.

15.5 Multivariate ARMA-Reihen

Eine unmittelbare Verallgemeinerung der univariaten ARMA-Reihen kann wie folgt definiert werden.

Definition 15.17 (*Multivariate ARMA-Reihen*)

Eine multivariate ARMA(p, q)-Reihe ist eine komplexe, zentrierte, stationäre r -dimensionale Zeitreihe, die eine Gleichung

$$\sum_{j=0}^p a_j X_{t-j} = \sum_{j=0}^q b_j e_{t-j}, \quad \forall t \in \mathbb{Z}, \quad (15.73)$$

erfüllt mit gegebenen Koeffizientenmatrizen $a_j, b_j \in \mathbb{C}^{r \times r}$, $p, q \in \mathbb{N}$, $a_0 = b_0 = I_r$ die $r \times r$ -Einheitsmatrix, $a_p \neq \mathbf{0}$, $b_q \neq \mathbf{0}$ und einem geeigneten reellen oder komplexen r -dimensionalen weißen Rauschen $e \sim \text{WR}(0, \Sigma)$.

Genau wie im univariaten Fall können wir die ARMA-Gleichung (15.73) mit Hilfe von Filtern oder dem Lag-Operator darstellen.

Bemerkung 15.18 (Die ARMA-Gleichung in Filterdarstellung)

Seien $A(z) = \sum_{j=0}^p a_j z^j$, $B(z) = \sum_{j=0}^q b_j z^j$, $z \in \mathbb{C}$, die (matrixwertigen) z -Transformationen und $\varphi(\omega) = A(e^{i\omega})$, $\psi(\omega) = B(e^{-i\omega})$ damit gebildete Filter, so schreibt sich (15.73) mit Hilfe dieser Filter und der zugehörigen MOW's Z_X und Z_e wie folgt.

$$\int e^{it \cdot} \varphi dZ_X = \int e^{it \cdot} \psi dZ_e, \quad \forall t \in \mathbb{Z}. \quad (15.74)$$

Auch die kompakte Darstellung

$$A(L)X_t = B(L)e_t, \quad \forall t \in \mathbb{Z}, \quad (15.75)$$

mit dem Lag-Operator L ist wie im univariaten Fall zu verstehen (siehe Bemerkung 7.3). \square

In weitgehender Analogie zum univariaten Fall gilt das folgende Resultat zur Existenz und Eindeutigkeit von Lösungen.

Satz 15.19 (*Existenz und Eindeutigkeit von Lösungen der ARMA-Gleichg.*)

Sei $\det A(z) \neq 0$ für alle z mit $|z| \leq 1$ (es existiere also die Inverse von $A(z)$)

für derartige z). Dann hat die ARMA-Gleichung (15.73) genau eine Lösung. Diese ist gegeben durch

$$X(t) = \int e^{it \cdot} \varphi^{-1} \psi dZ_e, \quad \forall t \in \mathbb{Z}. \quad (15.76)$$

Dabei kann $X(t)$ in einer einseitigen MA-Darstellung

$$X(t) = \sum_{s=0}^{\infty} c_s e_{t-s}, \quad \forall t \in \mathbb{Z} \quad (15.77)$$

dargestellt werden, wobei c_s die Entwicklungskoeffizienten(-Matrizen) der Potenzreihenentwicklung von $A^{-1}(z)B(z)$ sind, also

$$A^{-1}(z)B(z) = \sum_{s=0}^{\infty} c_s z^s, \quad (15.78)$$

die für $|z| \leq 1 + \varepsilon, \varepsilon > 0$ geeignet, (elementweise) konvergiert.

Beweis: Wegen $\det A(z) \neq 0$ sogar für alle z mit $|z| \leq 1 + \varepsilon, \varepsilon > 0$ geeignet, existiert $A^{-1}(z)$ in diesem Bereich und bildet elementweise eine für $|z| \leq 1 + \varepsilon$ holomorphe Funktion. Dann ist auch $A^{-1}(z)B(z)$ für $|z| \leq 1 + \varepsilon$ holomorph und lässt sich (elementweise) in eine konvergente Potenzreihe entwickeln, d.h. es gilt (15.78), wobei übrigens wie im univariaten Fall (siehe Bemerkung 7.8) gilt

$$|c_s(x, j)| \leq c \varrho^{-s}, \quad \forall s = 0, 1, 2, \dots, \quad (15.79)$$

mit geeigneter Konstante $c > 0$ und $\varrho = 1 + \varepsilon$. Da die Komponenten von $\varphi^{-1}\psi$ beschränkte Funktionen auf $[-\pi, \pi]$ sind, ist $\varphi^{-1}\psi$ ein Filter und damit $X = (X(t) : t \in \mathbb{Z})$ gemäß (15.76) eine zentrierte, stationäre Zeitreihe. Anwenden des Filter φ auf X liefert mit der Substitutionsregel (Satz 15.9)

$$\int e^{it \cdot} \varphi dZ^X = \int e^{it \cdot} \varphi(\varphi^{-1}\psi) dZ_e = \int e^{it \cdot} \psi dZ_e,$$

d.h. X löst (15.73). Nach Satz 15.10 folgt die Darstellung (15.77). Anwendung des Filters φ^{-1} auf (15.74) zeigt außerdem, dass die Lösung eindeutig ist. \square

Bemerkung 15.20 (Spektraldichtematrix einer ARMA-Reihe)

In der Situation von Satz 15.19 ist nach (15.76), (15.26) und (15.39) die Spektraldichtematrix f_X der ARMA-Reihe X gleich

$$f_X(\omega) = \frac{1}{2\pi} (\varphi^{-1}\psi) \Sigma (\varphi^{-1}\psi)^*(\omega), \quad \forall \omega \in [-\pi, \pi]. \quad (15.80)$$

Bemerkung 15.21 (Berechnung der Koeffizienten-Matrizen)

Die Berechnung der Koeffizienten-Matrizen c_j in *ARMA*-Modellen erfolgt (erneut analog zum univariaten Fall) am einfachsten rekursiv mit Hilfe der folgenden Formeln, die man durch Koeffizientenvergleich aus der Beziehung $B(z) = A(z) \cdot \sum_{s=0}^{\infty} c_s z^s$ erhält:

$$\begin{aligned} c_0 &= a_0 = I_r \in \mathbb{R}^{r \times r} \\ c_j &= b_j - \sum_{i=1}^j a_i c_{j-i}, \quad j \geq 1, \end{aligned} \quad (15.81)$$

mit $a_i = 0$ für $i > p$ und $b_j = 0$ für $j > q$. Wie im univariaten Fall nennt man wegen des Vorliegens der *MA*-Darstellung (15.77) mit (elementweise) absolut konvergenten Koeffizientenmatrizen c_j die Reihe $X = (X(t) : t \in \mathbb{Z})$ *kausal*.

Auch der dazu duale Begriff der *Invertibilität* wird in völliger Analogie zum univariaten Fall (vgl. Definition 7.11) benutzt. In Formel (7.16) ist dabei $e = (e_t : t \in \mathbb{Z})$ ein r -variates weißes Rauschen und d_j sind Matrizen aus $\mathbb{C}^{r \times r}$, deren Elemente absolut summierbar sind. Mit einer analogen Überlegung wie in Satz 15.19 zeigt man, dass Invertibilität gegeben ist, wenn $\det B(z) \neq 0$, $\forall |z| \leq 1$ gilt.

Die Autokovarianz $\Gamma(h) = EX(t+h)X(t)^*$ berechnet sich z.B. aus (15.10) oder aus den multivariaten *Yule-Walker-Gleichungen*.

$$\Gamma(j) + \sum_{\kappa=1}^p a_{\kappa} \Gamma(j - \kappa) = \sum_{j \leq \kappa \leq q} b_{\kappa} \Sigma c_{\kappa-j}^* \quad (15.82)$$

für $j = 0, 1, 2, \dots$, die man aus der *ARMA*-Gleichung (15.73) durch Multiplikation mit X_{t-j}^* und anschließender Erwartungswertbildung erhält. Dabei löst man die ersten $p+1$ Gleichungen, um zunächst $\Gamma(0), \dots, \Gamma(p)$ zu erhalten. Danach kann man $\Gamma(j)$ für $j > p$ rekursiv aus (15.82) bestimmen. \square

Beispiel 15.22 (Multivariates *AR*(1)-Modell)

Wir spezialisieren (15.82) nun für das reelle, r -variante *AR*(1)-Modell. Die *Yule-Walker-Gleichungen* ergeben sich dann mit $p = 1$ und $q = 0$ zu

$$\begin{cases} \Gamma(0) + a_1 \Gamma(-1) = \Sigma & (j = 0) \\ \Gamma(1) + a_1 \Gamma(0) = 0 & (j = 1) \end{cases} \quad (15.83)$$

und

$$\Gamma(j) + a_1 \Gamma(j-1) = 0 \quad (j \geq 2). \quad (15.84)$$

Aus (15.83) folgt mit $\Gamma(-1) = \Gamma(1)^T$ sowie $\Gamma(0) = \Gamma(0)^T$

$$\Gamma(0) - a_1 \Gamma(0) a_1^T = \Sigma. \quad (15.85)$$

Mit Hilfe des vec -Operators, der aus einer Matrix durch Untereinanderstellen der Spalten einen Spaltenvektor macht, und des Kronecker-Produkts von Matrizen (s. etwa Lütkepohl (1993)), kann man eine explizite Lösung von (15.85) angeben. (15.85) ist nämlich äquivalent zu

$$\text{vec } \Gamma(0) - \text{vec } a_1 \Gamma(0) a_1^T = \text{vec } \Sigma.$$

Wegen $\text{vec } a_1 \Gamma(0) a_1^T = (a_1 \otimes a_1) \text{vec } \Gamma(0)$ ergibt sich hieraus

$$(I_{r^2 \times r^2} - a_1 \otimes a_1) \cdot \text{vec } \Gamma(0) = \text{vec } \Sigma, \quad (15.86)$$

also

$$\text{vec } \Gamma(0) = (I_{r^2 \times r^2} - a_1 \otimes a_1)^{-1} \text{vec } \Sigma. \quad \square \quad (15.87)$$

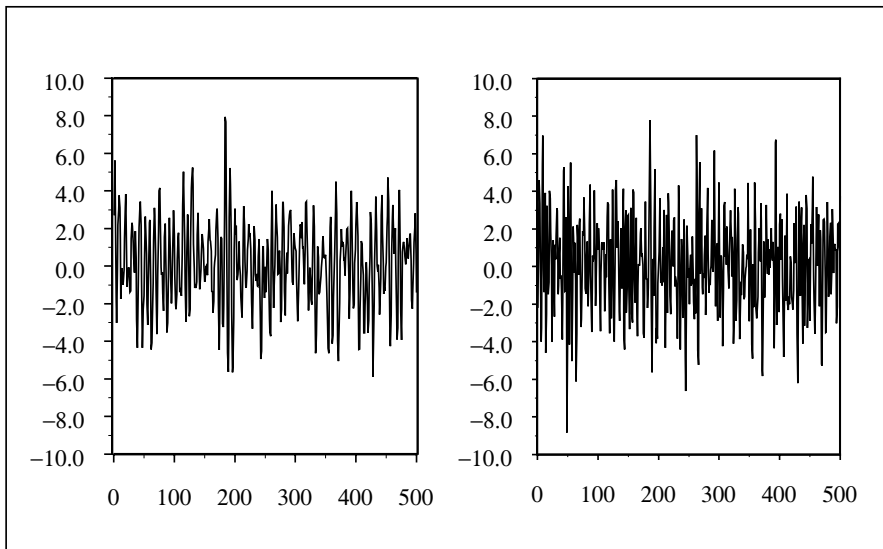


Abb. 15.2 Simulation der Länge $n = 500$ der zweidimensionalen $AR(2)$ -Zeitreihe aus Beispiel 15.23

Beispiel 15.23 (Multivariates $AR(2)$ -Modell)

Wie man sich leicht klar macht, treten bei multivariaten $ARMA$ -Modellen selbst für kleinere Ordnungen schnell eine Vielzahl von Parametern auf. Um die Dimension der Parameter möglichst gering zu halten, versucht man deshalb in der praktischen Anwendung durch inhaltliche Überlegungen möglichst eine gewisse Strukturierung des Modells vorzunehmen. Besonders günstig ist

es natürlich, wenn gerechtfertigt werden kann, dass eine Reihe von Modellparametern gleich einem festen Wert (etwa Null) gesetzt werden können und nicht geschätzt werden müssen. Als Beispiel betrachten wir das folgende zweidimensionale $AR(2)$ -Modell

$$X_t + \begin{pmatrix} a_1 & 0 \\ 0 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} a_2 & 0 \\ b_1 & b_2 \end{pmatrix} X_{t-2} + e_t, \quad t \in \mathbb{Z}. \quad (15.88)$$

Wegen $\det A(z) = (1+a_1z+a_2z^2)(1+b_2z^2)$ erfüllen wir die hinreichende Bedingung für die Existenz einer eindeutigen stationären Lösung von (15.88), wenn die Parameter a_1 und a_2 die übliche Stationaritäts- und Kausalitätsbedingung für univariate $AR(2)$ -Prozesse erfüllen und wenn $|b_2| < 1$ (vgl. Satz 15.19). b_1 kann offenbar beliebig gewählt werden. Man erkennt aus (15.88), dass die erste Komponente $(X_1(t))$ eine univariate $AR(2)$ -Zeitreihe darstellt. Nur die zweite Komponente hängt von zurückliegenden Beobachtungen beider Komponenten ab. b_1 ist dabei der Faktor für den Einfluss von $X_1(t-2)$ auf $X_2(t)$.

In Abbildung 15.2 ist eine simulierte Zeitreihe der Länge $n = 500$ mit den Parametern $a_1 = -1.27, a_2 = 0.81, b_1 = 1.00$ und $b_2 = 0.80$ zu sehen. Das weiße Rauschen ist zweidimensional normalverteilt jeweils mit Varianz 1 und Korrelationskoeffizient 0.5. Basierend auf dieser simulierten Zeitreihe haben wir mit Hilfe der multivariaten Yule-Walker-Methode die Parametermatrizen geschätzt und gemäß (15.80) einen Schätzer für die Spektraldichtematrix gewonnen. In Abbildung 15.3 sind die Diagonalelemente dieses Schätzers zu sehen, die den Spektraldichten der beiden Komponenten entsprechen und in Abbildung 15.4 ist der Kohärenzkoeffizient und das Phasenspektrum abgebildet. Man erkennt, dass die erste Komponente, deren Spektraldichte einen peak bei der Frequenz $\omega_0 = \pi/4$ besitzt, einen deutlichen Einfluss auf das Spektrum der zweiten Komponente nimmt, der auch durch den Kohärenzkoeffizienten wiedergegeben wird. Für den Fall $b_1 = 0$ wäre nämlich auch die zweite Komponente eine univariate $AR(2)$ -Zeitreihe mit einzigem peak der Spektraldichte bei $\omega_1 = \pi/2$. \square

Bemerkung 15.24 (Identifizierbarkeitsprobleme)

Sei $X(t) = (X_1(t), X_2(t))^T$, $t \in \mathbb{Z}$, eine 2-dimensionale $AR(1)$ -Reihe mit

$$X(t) - \begin{pmatrix} 0 & \varrho \\ 0 & 0 \end{pmatrix} X(t-1) = e(t), \quad \forall t \in \mathbb{Z}. \quad (15.89)$$

Die z -Transformation ist $A(z) = \begin{pmatrix} 1 & -\varrho z \\ 0 & 1 \end{pmatrix}$ und $\det A(z) = 1$ hat überhaupt keine Nullstellen. Mit $A^{-1} = \begin{pmatrix} 1 & \varrho z \\ 0 & 1 \end{pmatrix} = I + \varrho \begin{pmatrix} 0 & z \\ 0 & 0 \end{pmatrix}$ ergibt sich nach Satz 15.5 die MA -Darstellung

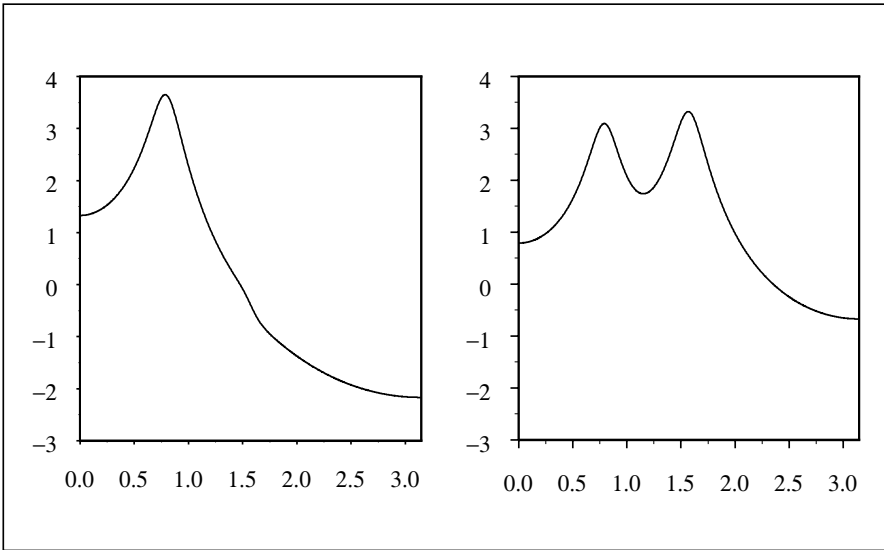


Abb. 15.3 Logarithmierte Diagonalelemente der über ein $AR(2)$ -Modell aus den Daten aus Abbildung 15.2 geschätzten Spektraldichtematrix

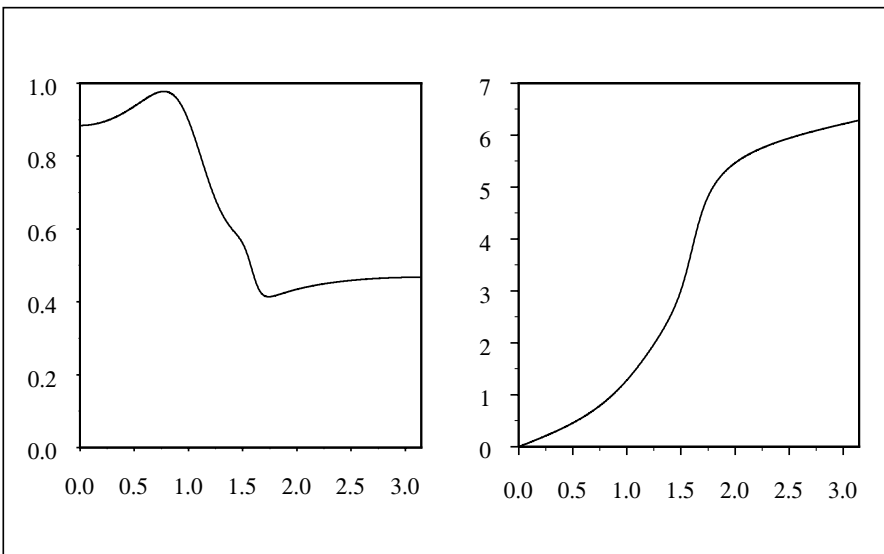


Abb. 15.4 Kohärenzkoeffizient (links) und Phasenspektrum (rechts) der über ein $AR(2)$ -Modell aus den Daten aus Abbildung 15.2 geschätzten Spektraldichtematrix

$$X(t) = e(t) + \begin{pmatrix} 0 & \varrho \\ 0 & 0 \end{pmatrix} e(t-1), \quad \forall t \in \mathbb{Z}. \quad (15.90)$$

Damit gehorcht $X = (X(t) : t \in \mathbb{Z})$ sowohl einer $AR(1)$ -Darstellung als auch einer $MA(1)$ -Darstellung. In der Klasse der $ARMA$ -Modelle sind die Koeffizientenmatrizen also nicht eindeutig bestimmt. Eine ausführliche Diskussion der Identifizierbarkeitsproblematik für multivariate $ARMA$ -Zeitreihen findet man z.B. bei Lütkepohl (1993), Chap.7.

In Anwendungen spielen glücklicherweise AR -Modelle die Hauptrolle und bei diesen treten keine Identifizierbarkeitsprobleme auf. \square

15.6 Schätzung des Mittelwertvektors und der Autokovarianzmatrix einer multivariaten Zeitreihe

Für eine reelle, r -variante Zeitreihe $X = (X_t : t \in \mathbb{Z})$ mit Mittelwertvektor $\mu = EX_t$ und Autokovarianz $\Gamma(h) = Cov(X_{t+h}, X_t)$ gibt es in völliger Analogie zum univariaten Fall auf der Basis von Beobachtungsvektoren X_1, \dots, X_n den *Mittelwertschätzer (empirischer Mittelwert)*

$$\bar{X} = \bar{X}_n = \frac{1}{n} \sum_{t=1}^n X_t, \quad (15.91)$$

den *Autokovarianzschätzer (empirische Autokovarianz)*

$$\hat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{t+h} - \bar{X}_n)(X_t - \bar{X}_n)^T = (\hat{\gamma}_{ij}(h))_{i,j=1,\dots,r}, \quad (15.92)$$

für $h \geq 0$ und $\hat{\Gamma}(-h) := \hat{\Gamma}(h)^T$, und den *Autokorrelationsschätzer (empirische Autokorrelation)*

$$\hat{R}(h) = (\hat{\varrho}_{ij}(h))_{i,j=1,\dots,r} \quad (15.93)$$

mit den *empirischen Kreuzkorrelationen*

$$\hat{\varrho}_{ij}(h) = \frac{\hat{\gamma}_{ij}(h)}{\sqrt{\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)}}, \quad h \in \mathbb{Z}. \quad (15.94)$$

Im Fall $i = j$ ist $\hat{\varrho}_{ii}(h)$ die übliche univariate Autokorrelationsfunktion der i -ten Komponente $(X_i(t) : t \in \mathbb{Z})$.

In Verallgemeinerung des univariaten Falles gilt für den empirischen Mittelwert \bar{X}_n

Satz 15.25 (*Konsistenz des Mittelwertschätzers*)

Für eine reelle, stationäre r -variante Zeitreihe $X = (X_t : t \in \mathbb{Z})$ ist $\bar{X}_n = n^{-1} \sum_{t=1}^n X_t$ ein komponentenweise erwartungstreuer Schätzer für $\mu = EX_t$. Falls für alle $i = 1, \dots, r$ die Konvergenz $\gamma_{ii}(h) \rightarrow 0$ für $h \rightarrow \infty$ gilt, so folgt für $n \rightarrow \infty$

$$\bar{X}_n \rightarrow \mu \quad \text{n.W.} \quad \text{und} \quad (15.95)$$

$$E((\bar{X}_n - \mu)^T (\bar{X}_n - \mu)) \rightarrow 0. \quad (15.96)$$

Gilt darüberhinaus $\sum_{i=1}^r \sum_{h=-\infty}^{+\infty} |\gamma_{ii}(h)| < \infty$, so folgt sogar

$$nE(\bar{X}_n - \mu)^T (\bar{X}_n - \mu) \rightarrow \sum_{i=1}^r \sum_{h=-\infty}^{\infty} \gamma_{ii}(h) = 2\pi \sum_{i=1}^r f_{ii}(0), \quad (15.97)$$

wobei jeweils f_{ii} die Spektraldichte der i -ten Komponente $(X_i(t) : t \in \mathbb{Z})$ von (X_t) bezeichnet.

Der Beweis ergibt sich unmittelbar durch Anwendung der entsprechenden univariaten Bemerkung 10.2 auf die Komponenten $(X_i(t) : t \in \mathbb{Z})$. \square

Ohne Beweis führen wir noch ein zu Satz 10.12 multivariates Analogon zur asymptotischen Normalverteilung von \bar{X}_n an.

Satz 15.26 (*Asymptotische Normalverteilung des Mittelwertschätzers*)

Sei $e = (e_t : t \in \mathbb{Z})$ ein r -variates weißes Rauschen mit unabhängigen und identisch verteilten r -Vektoren $e_t \sim (0, \Sigma)$. Weiter seien α_t , $t \geq 0$, reelle $r \times r$ Matrizen mit

$$\sum_{t=0}^{\infty} t |\alpha_t(j, k)| < \infty, \quad \forall j, k = 1, \dots, r, \quad (15.98)$$

und $\mu \in \mathbb{R}^r$. Dann gilt für die streng stationäre $MA(\infty)$ -Zeitreihe

$$X_t = \mu + \sum_{s=0}^{\infty} \alpha_s e_{t-s}, \quad \forall t \in \mathbb{Z}, \quad (15.99)$$

der multivariate zentrale Grenzwertsatz ($n \rightarrow \infty$)

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sum_{h=-\infty}^{\infty} \Gamma(h)\right), \quad (15.100)$$

wobei $\Gamma(h)$ die Autokovarianzmatrix von (X_t) zum Lag h bezeichnet (vgl. (15.10)). Die asymptotische Kovarianzmatrix $\Sigma_{\bar{X}} := \sum_{h=-\infty}^{\infty} \Gamma(h)$ ist nach Satz 15.5 gleich

$$\Sigma_{\bar{X}} = f(0) \cdot 2\pi, \quad (15.101)$$

wobei f die Spektraldichtematrix von X ist.

Man kann (15.100) dazu verwenden, ein asymptotisches Konfidenzellipse für μ zu bestimmen. Es folgt nämlich aus (15.100) unmittelbar

$$n(\bar{X}_n - \mu)^T \Sigma_{\bar{X}}^{-1} (\bar{X}_n - \mu) \xrightarrow{\mathcal{D}} \chi_r^2, \quad (15.102)$$

wobei χ_r^2 eine Chi-Quadrat-Verteilung mit r Freiheitsgraden bezeichnet. Ist $\chi_{r,1-\alpha}^2$ das $(1 - \alpha)$ -Quantil ($0 < \alpha < 1$) dieser Verteilung, so impliziert (15.102), dass

$$\{\mu \in \mathbb{R}^r : n(\bar{X}_n - \mu)^T \Sigma_X^{-1} (\bar{X}_n - \mu) \leq \chi_{r,1-\alpha}^2\} \quad (15.103)$$

ein Konfidenzellipse mit asymptotischer Überdeckungswahrscheinlichkeit $1 - \alpha$ für $EX_t = \mu$ ist.

Zur Konsistenz der empirischen Kovarianz und Korrelation zitieren wir die folgenden Ergebnisse. Da wir explizite und leicht zugängliche Literaturangaben machen wollen, beschränken wir uns bei den folgenden Aussagen auf den Fall $r = 2$. Grundsätzlich sind vergleichbare Resultate auch für höherdimensionale Zeitreihen zu erwarten.

Satz 15.27 (*Konsistenz der empirischen Autokovarianz*)

Sei $X_t = \sum_{s=0}^{\infty} \alpha_s e_{t-s}$, $t \in \mathbb{Z}$, eine bivariate Zeitreihe ($r = 2$) mit unabhängigen und identisch verteilten $e_t \sim (0, \Sigma)$ und 2×2 -Koeffizientenmatrizen α_s , die komponentenweise absolut summierbar sind. Dann gilt für $n \rightarrow \infty$ und $i, j = 1, 2, h \in \mathbb{Z}$,

$$\hat{\gamma}_{ij}(h) \rightarrow \gamma_{ij}(h) \quad n.W. \quad (15.104)$$

sowie

$$\hat{\varrho}_{ij}(h) \rightarrow \varrho_{ij}(h) \quad n.W. \quad (15.105)$$

Beweis: Siehe Brockwell/Davis (1991), Chap. 11, Theorem 11.2.1. \square

Satz 15.28 (*Asymptotische Normalität der empirischen Autokorrelation*)

Sind im vorangehenden Satz die beiden Komponenten-Prozesse $(e_{1t} : t \in \mathbb{Z})$ und $(e_{2t} : t \in \mathbb{Z})$ zusätzlich stochastisch unabhängig, so gilt für $h \neq k$

$$\sqrt{n} \begin{pmatrix} \hat{\varrho}_{12}(h) \\ \hat{\varrho}_{12}(k) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \tilde{\Sigma}) \quad \text{für } n \rightarrow \infty, \quad (15.106)$$

mit $\tilde{\Sigma} = (\tilde{\sigma}_{ij})$ und

$$\tilde{\sigma}_{11} = \tilde{\sigma}_{22} = \sum_{t=-\infty}^{+\infty} \varrho_{11}(t) \varrho_{22}(t) \quad (15.107)$$

sowie

$$\tilde{\sigma}_{12} = \tilde{\sigma}_{21} = \sum_{t=-\infty}^{+\infty} \varrho_{11}(t) \varrho_{22}(t+k-h). \quad (15.108)$$

Beweis: Siehe Brockwell/Davis (1991), Chap. 11, Theorem 11.2.2 □

Will man bei einer zentrierten, bivariaten Zeitreihe $X(t) = (X_1(t), X_2(t))^T$, $t \in \mathbb{Z}$, testen, ob die Komponenten X_1 und X_2 unkorreliert sind, d.h. ob $\varrho_{12}^X(h) = EX_1(t+h)X_2(t) = 0$ ist $\forall h \in \mathbb{Z}$, so ist nach obigem Satz und unter den dort gemachten Voraussetzungen

$$\hat{\varrho}_{12}^X(h) \pm \frac{1}{\sqrt{n}} z_{1-\alpha/2} \cdot \tilde{\sigma}_{11} \quad (15.109)$$

ein Konfidenzintervall, das für festes h unter der Hypothese der Unkorreliertheit von X_1 und X_2 den Nullpunkt mit asymptotischer Wahrscheinlichkeit $1 - \alpha$ enthält. Lehnt man also die Hypothese ab, wenn der Nullpunkt nicht im Konfidenzintervall liegt, so erhält man einen Test zum Niveau α .

In der Praxis ist $\tilde{\sigma}_{11}$ allerdings fast immer unbekannt, da zur Berechnung nach (15.107) die Korrelationen $\varrho_{ii}(t)$ für alle Zeitpunkte $t \in \mathbb{Z}$ bekannt sein müssten. Als Ausweg verwendet man die sogenannte *prewhitening*-Methode, die wir schon im Abschnitt 12.4 kennengelernt haben und die auf den folgenden Überlegungen basiert.

Wendet man auf jede der Komponenten X_1 und X_2 jeweils Filter φ_1 und φ_2 an, d.h. nichts anderes als dass man auf $(X(t) : t \in \mathbb{Z})$ den bivariaten Filter

$$\varphi = \begin{pmatrix} \varphi_1 & 0 \\ 0 & \varphi_2 \end{pmatrix}$$

anwendet, so ergibt sich für $\int e^{it\cdot} \varphi dZ^X = Y(t)$ offenbar nach (15.29) für die Kovarianz der gefilterten Reihe $\Gamma_Y(h) = \text{diag} \left(\int e^{ih\cdot} |\varphi_1|^2 d\mu_1, \int e^{ih\cdot} |\varphi_2|^2 d\mu_2 \right)$. Die Reihe $(Y(t) : t \in \mathbb{Z})$ hat also wieder unkorrelierte Komponenten, wenn dies für die Ausgangsreihe $(X(t) : t \in \mathbb{Z})$ gilt. Man testet deshalb die Hypothese der Unkorreliertheit an $(Y(t) : t \in \mathbb{Z})$ anstelle von $(X(t) : t \in \mathbb{Z})$.

Gelingt es nun, Filter φ_j , $j = 1, 2$, zu finden derart, dass $(Y_j(t) : t \in \mathbb{Z})$ ein weißes Rauschen ist (prewhitening), so gilt in (15.107) $\tilde{\sigma}_{11}^Y = \tilde{\sigma}_{22}^Y = 1$, so dass entsprechend zu (15.109) das Konfidenzintervall $\hat{\varrho}_{12}^Y(h) \pm \frac{1}{\sqrt{n}} z_{1-\alpha/2}$ unter der Hypothese der Unkorreliertheit das approximative Konfidenzniveau $1 - \alpha$ hat.

In der Praxis passt man meist an die beiden Komponentenreihen univariate *AR*-Reihen an, berechnet sodann die jeweiligen Residuen $Y_1(t)$ und $Y_2(t)$ (dies ist die Filterung oder auch das prewhitening) und führt dann mit diesen Residuen den beschriebenen Test durch.

15.7 Lineare Vorhersage bei multivariaten Zeitreihen

In Verallgemeinerung zum univariaten Fall wollen wir in diesem Abschnitt die Frage behandeln, wie man für r -variante (Spalten-)Vektoren X_1, X_2, \dots, X_n , deren Komponenten zu einem reellen Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ gehören, den Wert X_{n+1} komponentenweise durch eine Linearkombination der Komponenten von X_1, \dots, X_n möglichst gut approximiert.

Im Rahmen der Zeitreihenanalyse sind X_1, X_2, \dots zentrierte Elemente einer r -variaten Zeitreihe mit $H = L_2^{\mathbb{R}}$ und dem üblichen Skalarprodukt. Im Folgenden darf der Hilbertraum aber allgemein sein.

Nach den Überlegungen im univariaten Fall ist für den von allen Komponenten der Vektoren X_1, \dots, X_n aufgespannten Teilraum H_n von H , also

$$H_n = \text{span}\{X_{tj} : 1 \leq t \leq n, \quad 1 \leq j \leq r\}, \quad (15.110)$$

der optimale lineare Vorhersagevektor $\hat{X}_{n+1} = (\hat{X}_{n+1,1}, \dots, \hat{X}_{n+1,r})^T$ gegeben durch die komponentenweise Projektion von X_{n+1} auf H_n . Mit dem Projektionsoperator $P_n : H \rightarrow H$ auf H_n gilt also

$$\hat{X}_{n+1} := P_n X_{n+1} := (P_n X_{n+1,1}, \dots, P_n X_{n+1,r})^T. \quad (15.111)$$

Wir nennen \hat{X}_{n+1} die *Einschrittvorhersage*.

Die Berechnung von \hat{X}_{n+1} kann in weitgehender Analogie zum univariaten Fall erfolgen. Dazu dehnen wir die rekursive Gram-Schmidt Orthogonalisierung auf den r -variaten Fall aus. Für Vektoren $X = (X_1, \dots, X_r)^T$ und $Y = (Y_1, \dots, Y_r)^T \in H^r$ bezeichne $\langle X, Y \rangle$ die $r \times r$ Matrix

$$\langle X, Y \rangle = (\langle X_i, Y_j \rangle)_{i,j=1,\dots,r}. \quad (15.112)$$

Für reelle $r \times r$ -Matrizen A und B gilt dann offenbar

$$\langle AX, BY \rangle = A \langle X, Y \rangle B^T \quad (15.113)$$

sowie

$$\langle X, Y \rangle = \langle Y, X \rangle^T. \quad (15.114)$$

Ist $\langle X, Y \rangle = \mathbf{0}$ (die Nullmatrix), so nennen wir X und Y orthogonal, was natürlich gleichbedeutend mit $X_i \perp Y_j \quad \forall i, j = 1, \dots, r$ ist.

Mit diesen Bezeichnungen ist \hat{X}_{n+1} das eindeutig bestimmte Element aus H_n^r mit $(X_{n+1} - \hat{X}_{n+1}) \perp X_j$ für $1 \leq j \leq r$.

Die Vektoren $Y_1 := X_1, Y_j := X_j - \hat{X}_j, j > 1$, sind nach Konstruktion paarweise orthogonal, und es gilt wegen $\hat{X}_j \in H_j, \forall j > 1$,

$$H_j = \text{span}\{X_1, \dots, X_j\} = \text{span}\{Y_1, \dots, Y_j\}. \quad (15.115)$$

Damit gibt es Matrizen $c_{ni} \in \mathbb{R}^{r \times r}$ mit

$$\hat{X}_{n+1} = \sum_{i=1}^n c_{ni} Y_{n+1-i}. \quad (15.116)$$

Rechtsmultiplikation von (15.116) mit Y_{n+1-j} liefert

$$c_{nj} \cdot \langle Y_{n+1-j}, Y_{n+1-j} \rangle = \langle \hat{X}_{n+1}, Y_{n+1-j} \rangle,$$

also wegen $\langle X_{n+1} - \hat{X}_{n+1}, Y_j \rangle = 0, \quad j = 1, \dots, n,$

$$c_{nj} \cdot \langle Y_{n+1-j}, Y_{n+1-j} \rangle = \langle X_{n+1}, Y_{n+1-j} \rangle \quad (15.117)$$

für alle $j = 1, \dots, n$.

Es gibt immer eine Lösung c_{nj} von (15.117), da die Projektion \hat{X}_{n+1} stets existiert.

Bezeichnet A^+ die Moore–Penrose–Inverse einer Matrix A , so ist

$$c_{nj} = \langle X_{n+1}, Y_{n+1-j} \rangle \langle Y_{n+1-j}, Y_{n+1-j} \rangle^+ \quad (15.118)$$

eine Lösung von (15.117). Die obige schrittweise Konstruktion der Vektoren Y_1, Y_2, \dots ist das multivariate Gram–Schmidt–Orthogonalisierungsverfahren.

Die $r \times r$ -Matrix v_n der quadratischen Vorhersagefehler sei in Verallgemeinerung von (4.9)

$$v_n = \langle X_{n+1} - \hat{X}_{n+1}, X_{n+1} - \hat{X}_{n+1} \rangle = \langle Y_{n+1}, Y_{n+1} \rangle, \quad n \geq 1. \quad (15.119)$$

Wir wollen voraussetzen, dass alle $X_{tj}, t \geq 1, j = 1, \dots, r$, linear unabhängig sind in H . Hieraus folgt, dass alle Matrizen v_n nicht ausgeartet sind und somit die Moore–Penrose–Inverse in (15.118) die übliche Inverse ist.

Nach den in (4.10) und (4.11) angegebenen Rekursionsformeln (nun als Matrixgleichungen zu lesen) kann man rekursiv nacheinander die Matrizen $v_0; c_{11}, v_1; c_{22}, c_{21}, v_2; c_{33}, c_{32}, c_{31}, v_3; \dots$ berechnen.

Es sei daran erinnert, dass der Beweis im univariaten Fall so geführt wurde, dass er auch im multivariaten Fall seine Gültigkeit behält.

Bemerkung 15.29

Whittle (1963) hat eine multivariate Verallgemeinerung der Levinson–Rekursion (vgl. Kapitel 4) angegeben. Wir wollen hier auf eine Darstellung verzichten. \square

15.8 Zustandsraummodelle

Viele der bisher betrachteten Modelle für Zeitreihen, z.B. die ARMA-Reihen, lassen sich als sogenannte Zustandsraummodelle darstellen (state-space-models). Diese stammen ursprünglich aus der Kontrolltheorie linearer Systeme

(siehe z.B. Hannan/Deistler (1988)). Eine ihrer besonderen Eigenschaften ist die Möglichkeit, lineare Vorhersagen rekursiv berechnen zu können. Dies geschieht mit dem Kalman-Filter, den wir im nächsten Abschnitt kennen lernen werden.

Bei der Darstellung auch univariater Zeitreihen als Zustandsraummodell wird man in natürlicher Weise auf Zufallsvektoren und sogar Zufallsmatrizen geführt. Deren Erwartungswerte sind im Folgenden, wie auch bisher schon, komponentenweise zu verstehen.

Zu einem *Zustandsraummodell* gehören zwei Folgen von Zufallsvektoren

$$X_1, X_2, \dots \text{ aus } \mathbb{R}^r, \quad \text{die Zustände zur Zeit } t = 1, 2, \dots \quad (15.120)$$

und

$$Y_1, Y_2, \dots \text{ aus } \mathbb{R}^v, \quad \text{die Beobachtungen zur Zeit } t = 1, 2, \dots \quad (15.121)$$

Dabei sollen für die Zustandsvektoren die folgenden *Zustandsgleichungen*

$$X_{n+1} = F_n X_n + e_n, \quad n = 1, 2, \dots, \quad (15.122)$$

gelten mit bekannten Matrizen $F_n \in \mathbb{R}^{r \times r}$ und Zufallsvektoren (Störungen) $e_n \in \mathbb{R}^r, n \geq 1$, mit

$$E e_n = \mathbf{0} \in \mathbb{R}^r \quad (15.123)$$

und bekannten Kovarianzmatrizen $Q_n \in \mathbb{R}^{r \times r}$ gemäß

$$E e_m e_n^T = \begin{cases} Q_n & , \text{ falls } m = n \\ \mathbf{0} \in \mathbb{R}^{r \times r} & , \text{ sonst } . \end{cases} \quad (15.124)$$

Die Beobachtungsvektoren Y_n sollen die folgenden *Beobachtungsgleichungen*

$$Y_n = G_n X_n + \eta_n, \quad n = 1, 2, \dots, \quad (15.125)$$

erfüllen mit ebenfalls gegebenen Matrizen $G_n \in \mathbb{R}^{v \times r}$ und Störungsvektoren $\eta_n, n \geq 1$, mit

$$E \eta_n = \mathbf{0} \in \mathbb{R}^v, \quad \forall n \geq 1, \quad (15.126)$$

sowie bekannten Kovarianzmatrizen $R_n \in \mathbb{R}^{v \times v}$ gemäß

$$E \eta_m \eta_n^T = \begin{cases} R_n & , \text{ falls } m = n \\ \mathbf{0} & , \text{ sonst } . \end{cases} \quad (15.127)$$

Weiter seien alle Komponenten von e_1, e_2, \dots mit denen von η_1, η_2, \dots unkorreliert, d.h. $E e_m \eta_n^T = \mathbf{0} \quad \forall m, n \geq 1$. Außerdem sei der Anfangszustand X_1 mit den Störungen $\{e_n\}$ und $\{\eta_n\}$ im selben Sinne unkorreliert.

Bemerkung 15.30

Aus (15.122) folgt durch sukzessives Einsetzen

$$\begin{aligned}
 X_n &= F_{n-1}X_{n-1} + e_{n-1} \\
 &= F_{n-1}(F_{n-2}X_{n-2} + e_{n-2}) + e_{n-1} \\
 &= F_{n-1}F_{n-2}X_{n-2} + F_{n-1}e_{n-2} + e_{n-1} \\
 &\vdots \\
 &= (F_{n-1} \cdots F_1)X_1 + (F_{n-2} \cdots F_2)e_1 + \dots + F_{n-1}e_{n-2} + e_{n-1}
 \end{aligned} \tag{15.128}$$

und entsprechend

$$\begin{aligned}
 Y_n &= (G_n F_{n-1} \cdots F_1)X_1 + (G_n F_{n-1} \cdots F_2)e_1 + \dots \\
 &\quad + G_n F_{n-1}e_{n-2} + G_n e_{n-1} + \eta_n.
 \end{aligned} \tag{15.129}$$

Als Folgerung ergibt sich unmittelbar für $1 \leq n \leq m$

$$Ee_m X_n^T = \mathbf{0}, \quad Ee_m Y_n^T = \mathbf{0}, \tag{15.130}$$

$$E\eta_m X_n^T = \mathbf{0}, \tag{15.131}$$

und für $1 \leq n < m$

$$E\eta_m Y_n^T = \mathbf{0}. \tag{15.132}$$

□

Wir wollen die bisher gemachten Annahmen in der folgenden Definition festhalten.

Definition 15.31

Eine Zeitreihe Y_1, Y_2, \dots hat eine Zustandsraum-Darstellung, wenn es ein Zustandsraum-Modell (15.122) –(15.127) (einschließlich aller Unkorreliertheitsforderungen) für $Y_n, n \geq 1$, gibt.

Beispiel 15.32 (Zustandsraum-Darstellung einer 1-dim. ARMA-Zeitreihe)

Sei ein univariater ARMA(p, q)-Prozess $Y = (Y_t : t \in \mathbb{Z})$ durch

$$A(L)Y_t = B(L)\varepsilon_t, \quad \forall t \in \mathbb{Z}, \tag{15.133}$$

gegeben mit weißem Rauschen $(\varepsilon_t : t \in \mathbb{Z})$, $\varepsilon_t \sim (0, \sigma^2)$, und $A(z) \neq 0$ für $|z| \leq 1$. Für $j > p$ sei $a_j = 0$ und für $j > p$ entsprechend $b_j = 0$. Bezeichne nun (Z_t) den kausalen AR(p)-Prozess, der die Gleichung

$$A(L)Z_t = \varepsilon_t, \quad \forall t \in \mathbb{Z}, \tag{15.134}$$

löst. Dann löst $Y_t = B(L)Z_t$ die Gleichung (15.133), denn es gilt

$$A(L)(B(L)Z_t) = B(L)A(L)Z_t = B(L)\varepsilon_t.$$

Für $r = \max(p, q + 1)$ ergibt sich also mit $X_t = (Z_{t-r+1}, \dots, Z_t)^T$

$$Y_t = (b_{r-1}, \dots, b_0)X_t. \quad (15.135)$$

Dies ist eine störungsfreie *Beobachtungsgleichung* im Sinne von (15.125) mit $v = 1$, $G = G_t = (b_{r-1}, \dots, b_0)$ und $\eta_t = \mathbf{0} \forall t$. Eine (15.122) entsprechende *Zustandsgleichung* kann durch

$$X_{t+1} = - \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & & \dots & 1 \\ a_r & a_{r-1} & a_{r-2} & \dots & a_1 \end{pmatrix} X_t + \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{pmatrix} \varepsilon_{t+1}, \quad (15.136)$$

gegeben werden mit nicht von n abhängiger Matrix $F = F_t \in \mathbb{R}^{r \times r}$ und $e_t = (0, \dots, 0, 1)^T \varepsilon_{t+1}$. \square

Im nächsten Abschnitt geht es nun um das grundsätzliche Problem der Approximation der Zustände X_t aus den Beobachtungen Y_s . Wir werden gewissermaßen exemplarisch das sogenannte Vorhersageproblem behandeln.

15.9 Der Kalman-Filter zur linearen Vorhersage

Das Vorhersageproblem in einem Zustandsraum-Modell besteht darin, den Wert X_{n+1} komponentenweise durch eine Linearkombination der Komponenten von Y_1, \dots, Y_n und eines Vektors Y_0 , der mit allen anderen vorkommenden Zufallsvariablen unkorreliert sei, zu approximieren. Häufig wählt man z.B. $Y_0 = (1, \dots, 1)^T$, um einen nichtverschwindenden Erwartungswert berücksichtigen zu können. Dann ist der optimale lineare Vorhersagevektor $\hat{X}_{n+1} = (\hat{X}_{n+1,1}, \dots, \hat{X}_{n+1,r})^T$ gegeben durch die komponentenweise Projektion von X_{n+1} auf den Teilraum H_n des Hilbertraumes $H := L_2^{\mathbb{R}}(\Omega, \mathcal{A}, P)$, der von allen Komponenten von Y_1, \dots, Y_n und Y_0 aufgespannt wird, also

$$H_n = \text{span}\{Y_{tj}, t = 0, \dots, n, j = 1, \dots, v\}, n \geq 0. \quad (15.137)$$

Mit dem Projektionsoperator $P_n : H \rightarrow H$ auf H_n , komponentenweise auf X_{n+1} angewandt, gilt also

$$\hat{X}_{n+1} = P_n X_{n+1} = (P_n X_{n+1,1}, \dots, P_n X_{n+1,r})^T. \quad (15.138)$$

Die Einschrittvorhersage \hat{X}_{n+1} ist charakterisiert als ein Element aus H_n mit $X_{n+1} - \hat{X}_{n+1} \perp H_n$, wobei $X \perp H_n$ wie schon früher bedeutet, dass alle

Komponenten von X auf H_n senkrecht stehen.

Generell verstehen wir im Folgenden für einen Vektor $X = (X_1, \dots, X_r)^T$ aus H^r und einen abgeschlossenen linearen Teilraum $V \subset H$ unter der Projektion $P_V X = (P_V X_1, \dots, P_V X_r)^T$ den Vektor der Projektionen $P_V X_j$ auf V . $P_V X$ ist der eindeutig bestimmte Vektor aus V^r mit $X - P_V X \perp V$.

Leicht zeigt man die folgenden *Rechenregeln*:

Für $X, Y \in H^r$ und $k \times r$ -Matrizen A, B gilt

$$P_V(AX + BY) = AP_V X + BP_V X \quad (\text{Linearität}). \quad (15.139)$$

Ist $Y \in H^v$ und $V = \text{span}\{Y_1, \dots, Y_v\}$ und $X \in H^r$ so schreiben wir für $P_V X$ auch $P(X|Y)$. Es gilt

$$P(X|Y) = EXY^T \cdot (EYY^T)^+, \quad (15.140)$$

wobei $(EYY^T)^+$ die Moore-Penrose-Inverse von EYY^T bezeichne.

Sind V_1 und V_2 zwei orthogonale Unterräume von H , so gilt für $X \in H^r$ und den Summenraum $V_1 \oplus V_2$

$$P_{V_1 \oplus V_2} X = P_{V_1} X + P_{V_2} X. \quad (15.141)$$

Satz 15.33 (Die Kalman Vorhersage)

Sei ein Zustandsraum-Modell (15.122)–(15.127) gegeben und $\hat{X}_n = P_{n-1} X_n$ die Einschnitt-Vorhersage sowie $\sigma_n = E(X_n - \hat{X}_n)(X_n - \hat{X}_n)^T$ die Kovarianzmatrix der Vorhersagefehler $X_n - \hat{X}_n$.

Mit den Anfangsbedingungen

$$\hat{X}_1 = P_0 X_1, \quad \sigma_1 = E(X_1 - \hat{X}_1)(X_1 - \hat{X}_1)^T$$

ergeben sich dann die Werte von \hat{X}_{n+1} , $n \geq 1$, rekursiv aus

$$\hat{X}_{n+1} = F_n \hat{X}_n + A_n B_n^+ (Y_n - G_n \hat{X}_n) \quad (15.142)$$

$$\sigma_{n+1} = F_n \sigma_n F_n^T + Q_n - A_n B_n^+ A_n^T, \quad (15.143)$$

wobei zur Abkürzung

$$\begin{aligned} A_n &= F_n \sigma_n G_n^T \\ B_n &= G_n \sigma_n G_n^T + R_n \end{aligned}$$

gesetzt wird und B_n^+ die Moore-Penrose-Inverse von B_n sei.

Beweis: Wir definieren nach Gram-Schmidt orthogonale Vektoren ξ_n durch $\xi_0 = Y_0$ und

$$\begin{aligned}\xi_n &= Y_n - P_{n-1}Y_n = G_nX_n + \eta_n + P_{n-1}(G_nX_n + \eta_n) \\ &= G_n(X_n - \hat{X}_n) + \eta_n, \quad n \geq 1.\end{aligned}\quad (15.144)$$

Es folgt aus den Rechenregeln (15.139) – (15.141)

$$\begin{aligned}\hat{X}_{n+1} &= P_nX_{n+1} = P_{n-1}X_{n+1} + P(X_{n+1}|\xi_n) \\ &= P_{n-1}(F_nX_n + e_n) + E(X_{n+1}\xi_n^T)(E\xi_n\xi_n^T)^+\xi_n \\ &= F_n\hat{X}_n + A_nB_n^+\xi_n\end{aligned}\quad (15.145)$$

wobei benutzt wurde, dass

$$\begin{aligned}E(X_{n+1}\xi_n^T) &= E(F_nX_n + e_n)((X_n - \hat{X}_n)^TG_n^T + \eta_n^T) \\ &= F_n\sigma_nG_n^T = A_n\end{aligned}$$

sowie

$$\begin{aligned}E\xi_n\xi_n^T &= E(G_n(X_n - \hat{X}_n) + \eta_n)((X_n - \hat{X}_n)^TG_n^T + \eta_n^T) \\ &= G_n\sigma_nG_n^T + R_n = B_n\end{aligned}$$

gilt.

Mit $\xi_n = Y_n - P_{n-1}(G_nX_n + \eta_n) = Y_n - G_n\hat{X}_n$ folgt also (15.142). Es bleibt der Nachweis von (15.143). Hierzu beachte man

$$\begin{aligned}\sigma_{n+1} &= EX_{n+1}X_{n+1}^T - E\hat{X}_{n+1}\hat{X}_{n+1}^T \\ &= E(F_nX_n + e_n)(X_n^TF_n^T + e_n^T) \\ &\quad - E(F_n\hat{X}_n + A_nB_n^+\xi_n)(\hat{X}_n^TF_n^T + \xi_n^T(A_nB_n^+)^T) \\ &= F_nEX_nX_n^TF_n^T + Q_n - F_nE\hat{X}_n\hat{X}_n^TF_n^T - A_nB_n^+E\xi_n\xi_n^T(A_nB_n^+)^T \\ &= F_n\sigma_nF_n^T + Q_n - A_nB_n^+A_n^T.\end{aligned}$$

□

In der Zeitreihenanalyse ist man meist an der Vorhersage künftiger Beobachtungswerte, also z.B. der Einschnitt-Vorhersage P_nY_{n+1} interessiert, vgl. Beispiel 15.32. Deren Berechnung kann ebenfalls mit Hilfe der Kalman-Rekursion erfolgen, wie die folgenden Überlegungen für die allgemeinere Frage der Berechnung der h -Schritt-Vorhersage P_nY_{n+h} , $h = 1, 2, \dots$, zeigen. Nach (15.125) gilt

$$P_nY_{n+h} = G_{n+h}P_nX_{n+h}, \quad h = 1, 2, \dots, \quad (15.146)$$

wodurch die h -Schrittvorhersage der Y_n auf diejenige der X_n zurückgeführt wird, welche wiederum durch Anwendung von (15.122) auf die Einschnitt-Vorhersage P_nX_{n+1} zurückgeführt werden kann gemäß

$$\begin{aligned}P_nX_{n+h} &= P_n(F_{n+h-1}X_{n+h-1} + e_{n+h-1}) \\ &= F_{n+h-1}P_nX_{n+h-1} \\ &\vdots \\ &= (F_{n+h-1}F_{n+h-2} \cdots F_{n+1})P_nX_{n+1}.\end{aligned}\quad (15.147)$$

Dabei kann $P_n X_{n+1} = \hat{X}_{n+1}$ mit Hilfe der Rekursion in Satz 15.33 berechnet werden. Für den h -Schritt-Vorhersagefehler (eine $r \times r$ -Matrix)

$$\sigma_n^{(h)} = E(X_{n+h} - P_n X_{n+h})(X_{n+h} - P_n X_{n+h})^T \quad (15.148)$$

ergibt sich aus (15.122) und (15.147) die Gleichung

$$X_{n+h} - P_n X_{n+h} = F_{n+h-1}(X_{n+h-1} - P_n X_{n+h-1}) + e_{n+h-1},$$

und damit die Rekursionsformel

$$\sigma_n^{(h)} = F_{n+h-1} \sigma_n^{(h-1)} F_{n+h-1}^T + Q_{n+h-1}. \quad (15.149)$$

Aus (15.125) und (15.146) folgt nun für den h -Schritt-Vorhersagefehler $v_n^{(h)}$ der Y (eine $v \times v$ -Matrix), d.h. für

$$v_n^{(h)} = E(Y_{n+h} - P_n Y_{n+h})(Y_{n+h} - P_n Y_{n+h})^T$$

die Beziehung

$$v_n^{(h)} = G_{n+h} \sigma_n^{(h)} G_{n+h}^T + R_{n+h}. \quad (15.150)$$

Aufgaben

Aufgabe 15.1 Man zeige, dass $X = (X_t : t \in \mathbb{Z})$ aus Definition 15.3 zentriert und stationär ist mit Kovarianz (15.10).

Aufgabe 15.2 Man beweise Satz 15.12.

Aufgabe 15.3 Seien für $j = 1, 2$ reelle stationäre, zentrierte Zeitreihen $X_j = (X_j(t) : t \in \mathbb{Z})$ gegeben, die die Voraussetzungen (15.19) für die Inversionsformel (15.20) erfüllen. U_j und V_j seien die MOW's in der reellen Spektralzerlegung (vgl. Aufgabe 6.6)

$$X_j(t) = 2 \int_{(0, \pi)} \cos(t\omega) dU_j(\omega) + 2 \int_{(0, \pi)} \sin(t\omega) dV_j(\omega), \quad \forall t \in \mathbb{Z}.$$

Man zeige für $B \in (0, \pi) \cap \mathbb{B}$

a) $2\langle U_1, U_2 \rangle(B) = 2\langle V_1, V_2 \rangle(B) = \int_B c_{12}(\omega) d\omega.$

b) $2\langle U_1, V_2 \rangle(B) = -2\langle U_2, V_1 \rangle(B) = \int_B q_{12}(\omega) d\omega,$

wobei $d\omega$ Integration bzgl. des Lebesgue-Maßes bedeute.

Hinweis: Man verwende die Darstellung der U_j und V_j aus Aufgabe 6.6.

Aufgabe 15.4 a) Man beweise die Aussagen (15.35) und (15.36).

b) Man bestimme in Beispiel 15.8 die MOW's U_j und V_j aus Aufgabe 15.3.

Aufgabe 15.5 Seien $e_1 = (e_1(t) : t \in \mathbb{Z})$ und $e_2 = (e_2(t) : t \in \mathbb{Z})$ zwei voneinander unabhängige, reelle weiße Rauschen mit Varianzen 0.30 bzw. 0.5. Man definiere eine zweidimensionale Zeitreihe $X(t) = (X_1(t), X_2(t))^T, t \in \mathbb{Z}$, durch

$$X_1(t) = 0.7 X_1(t-1) + e_1(t) \text{ und } X_2(t) = X_1(t) + e_2(t).$$

Man berechne für $j, k = 1, 2$ die zugehörigen Autokovarianzen γ_{jk} und die Kreuzspektraldichten f_{jk} .

Aufgabe 15.6 Sei $(X_2(t))$ eine reelle, zentrierte, stationäre Zeitreihe mit absolut summierbarer Autokovarianzfunktion und (e_t) eine weißes Rauschen $e_t \sim (0, \sigma_e^2)$, das unkorreliert mit $(X_2(t))$ ist. Ein Prozess $(X_1(t))$ werde nun durch die „Regressionsgleichung“

$$X_1(t) = aX_2(t) + e_t, \quad t \in \mathbb{Z},$$

mit $a \in \mathbb{R}$, definiert. Man zeige für die Spektraldichten f_{jk} von $(X_1(t), X_2(t))$

$$\begin{aligned} f_{12}(\omega) &= a f_{22}(\omega) \text{ (reell !)}, \\ f_{11}(\omega) &= a^2 f_{22}(\omega) + \sigma_e^2 / (2\pi), \\ K_{12}^2(\omega) &= \left(1 + \frac{\sigma_e^2}{2\pi a^2 f_{22}(\omega)}\right)^{-1}. \end{aligned}$$

Aufgabe 15.7 Man zeige mit den Bezeichnungen aus Bemerkung 15.18, dass mit der $rp \times rp$ -Matrix

$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ -I_r & 0 & \dots & 0 & 0 \\ 0 & -I_r & \dots & 0 & 0 \\ \vdots & & & \vdots & \vdots \\ 0 & 0 & \dots & -I_r & 0 \end{pmatrix} \quad (15.151)$$

die Gleichung

$$\det(I_{r \cdot p} + \mathbf{A}z) = \det(A(z))$$

gilt, wobei $A(z)$ die z -Transformation aus Bemerkung 15.18 ist.

Aufgabe 15.8 Man betrachte ein bivariates (also $r = 2$) $MA(1)$ -Modell $X_t = 1 + b e_{t-1}$ mit der 2×2 -Matrix b und einem zweidimensionalen weißen Rauschen $e \sim WR(0, \Sigma)$ gemäß

$$b = - \begin{pmatrix} 0.5 & 0.4 \\ -0.2 & 0.3 \end{pmatrix} \text{ und } \Sigma = \begin{pmatrix} 8 & 2 \\ 2 & 4 \end{pmatrix}.$$

Man bestimme die Autokovarianz- und Autokorrelationsmatrizen $\Gamma(h)$ bzw. $R(h)$ für $h = 0, 1$, sowie $\det B(z)$ mit der z -Transformation $B(z) = I_2 + b \cdot z$.

A

Anhang

A.1 Einige nützliche Formeln

In diesem ersten Teil des Anhangs stellen wir eine Reihe von Formeln zusammen, die wir im Laufe des Textes an mehreren Stellen gebraucht haben. Da sie den laufenden Text nach unserer Einschätzung gestört hätten, haben wir sie herausgezogen und in diesen Anhang verwiesen.

$$e^{i\lambda} = \cos \lambda + i \sin \lambda \quad (\text{A.1})$$

$$\cos \lambda = (e^{i\lambda} + e^{-i\lambda})/2 \quad (\text{A.2})$$

$$\sin \lambda = (e^{i\lambda} - e^{-i\lambda})/(2i) \quad (\text{A.3})$$

$$\sin(\lambda + \mu) = \sin \lambda \cdot \cos \mu + \cos \lambda \cdot \sin \mu \quad (\text{A.4})$$

$$\cos(\lambda + \mu) = \cos \lambda \cdot \cos \mu - \sin \lambda \cdot \sin \mu. \quad (\text{A.5})$$

Für $r, s \in \mathbb{Z}$ mit $r \leq s$ und alle $\omega \in \mathbb{R}$ gilt

$$\sum_{k=r}^s e^{i\omega k} = \begin{cases} s - r + 1 & , \omega \in 2\pi\mathbb{Z} \\ \exp(i\omega \frac{r+s}{2}) \sin(\frac{(s-r+1)\omega}{2}) / \sin(\frac{\omega}{2}) & , \omega \notin 2\pi\mathbb{Z} . \end{cases} \quad (\text{A.6})$$

Speziell für $\omega = \omega_j = 2\pi j/n$ ergibt sich

$$\sum_{k=1}^n e^{i\omega_j k} = \begin{cases} n & , j \in n\mathbb{Z} \\ 0 & , j \notin n\mathbb{Z} . \end{cases} \quad (\text{A.7})$$

Beweis: Da (A.7) direkt aus (A.6) folgt, zeigen wir nur (A.6). Dies gelingt mit Hilfe der Formel für geometrische Summen, denn

$$\begin{aligned}
\sum_{k=r}^s e^{i\omega k} &= (\exp(i\omega(s+1)) - \exp(i\omega r)) / (\exp(i\omega) - 1) \\
&= \exp(i\omega \frac{r+s}{2}) \cdot \frac{\exp(i\omega(s-r+1)/2) - \exp(-i\omega(s-r+1)/2)}{\exp(i\omega/2) - \exp(-i\omega/2)}
\end{aligned} \tag{A.8}$$

und (A.3). \square

Es sei $f : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{C}$ und $g : \mathbb{Z} \rightarrow \mathbb{C}$. Dann gelten

$$\sum_{r,s=1}^n f(r,s) = \sum_{h=0}^{n-1} \sum_{t=1}^{n-|h|} f(t+|h|, t) + \sum_{h=-(n-1)}^{-1} \sum_{t=1}^{n-|h|} f(t, t+|h|) \tag{A.9}$$

Im Falle $f(r,s) = f(s,r)$ bedeutet dies

$$\sum_{r,s=1}^n f(r,s) = \sum_{h=-(n-1)}^{n-1} \sum_{t=1}^{n-|h|} f(t+|h|, t) \tag{A.10}$$

Für den weiteren Spezialfall $f(r,s) = g(r-s)$ erhalten wir schließlich

$$\sum_{r,s=1}^n g(r-s) = \sum_{h=-(n-1)}^{n-1} (n-|h|)g(h) \tag{A.11}$$

Beweis: (A.10) und (A.11) folgen direkt aus (A.9). (A.9) selbst ergibt sich aus einer Veränderung der Summationsreihenfolge. Die rechte Seite stellt die Summation des quadratischen Schemas entlang der Diagonalen $h = r - s = -(n-1), \dots, n-1$ dar. \square

A.2 Integration komplexer Funktionen

Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : \Omega \rightarrow \mathbb{C}$ eine Abbildung derart, dass der Realteil $u = \operatorname{Re} f$ und der Imaginärteil $v = \operatorname{Im} f$ beide \mathcal{A} - \mathbb{B} messbar sind, so heißt f messbar. f heißt μ -integrierbar, wenn darüber hinaus u und v μ -integrierbar sind. Die komplexe Zahl

$$\int f d\mu = \int u d\mu + i \int v d\mu \tag{A.12}$$

heißt das Integral von f bzgl. μ .

Ist $\mu = P$ ein Wahrscheinlichkeitsmaß so heißt eine messbare Funktion $X : \Omega \rightarrow \mathbb{C}$ eine komplexe Zufallsvariable. Ist X bzgl. des Wahrscheinlichkeitsmaßes P integrierbar, so heißt $EX := \int X dP$ der Erwartungswert von X .

In Erweiterung des reellen Falles gilt für μ -integrierbare Funktionen f

$$\left| \int f \, d\mu \right| \leq \int |f| \, d\mu. \quad (\text{A.13})$$

Den Beweis findet man bei Bauer (1990).

Es gilt auch eine komplexe Variante des Satzes von Lebesgue.

Satz A.1 (*Satz von Lebesgue, majorisierte Konvergenz*)

Sind $f, f_n, n \geq 1, g$ komplexe, messbare Funktionen mit den Eigenschaften

$$\lim_{n \rightarrow \infty} f_n = f \text{ } \mu\text{-f.s.} \quad (\mu \text{ fast sichere Konvergenz}) \quad (\text{A.14})$$

$$|f_n| \leq |g| \text{ } \mu\text{-f.s. } \forall n \in \mathbb{N} \text{ sowie } \int |g| \, d\mu < \infty, \quad (\text{A.15})$$

so folgt für $n \rightarrow \infty$

$$\int |f_n - f| \, d\mu \rightarrow 0 \quad (\text{A.16})$$

und damit wegen (A.13) auch

$$\int f_n \, d\mu \rightarrow \int f \, d\mu. \quad (\text{A.17})$$

Zum Beweis wende man auf $|f_n - f|, n \geq 1$, den reellen Satz von Lebesgue mit der Majorante $2|g|$ an.

Beispiel A.2 (Spezielles Kroneckerlemma)

Ist $c_k, k \in \mathbb{Z}$, eine absolut summierbare Folge komplexer Zahlen, so gilt für $n \rightarrow \infty$

$$\frac{1}{n} \sum_{|k| < n} |k c_k| \rightarrow 0. \quad (\text{A.18})$$

Der Beweis ist früher als Aufgabe 2.5 gestellt worden.

Definition A.3 (L_r -Räume)

Für $1 \leq r < \infty$ bezeichnet $L_r^{\mathbb{C}}(\Omega, \mathcal{A}, \mu) = L_r^{\mathbb{C}}(\mu)$ die Familie der Äquivalenzklassen μ -fast sicher übereinstimmender messbarer Funktionen f mit

$$\|f\|_r := \left(\int |f|^r \, d\mu \right)^{\frac{1}{r}} < \infty.$$

$L_r^{\mathbb{C}}(\mu)$ ist ein Banachraum, d.h. ein linearer, normierter und vollständiger Raum mit der Norm $\|f\|_r$.

$L_2^{\mathbb{C}}(\mu)$ ist ein Hilbert-Raum mit innerem Produkt $\langle f, g \rangle_2 = \int f \bar{g} \, d\mu$, siehe Anhang A.3. Entsprechendes gilt im reellen Fall in $L_r^{\mathbb{R}}(\mu)$.

Die Konvergenz in $L_r^{\mathbb{C}}(\mu)$ oder $L_r^{\mathbb{R}}(\mu)$ heißt L_r -Konvergenz und $\|\cdot\|_r$ heißt L_r -Norm. Im Fall $r = 2$ spricht man auch von der Konvergenz im quadratischen Mittel. Anstelle von Äquivalenzklassen arbeiten wir wie üblich mit entsprechenden Repräsentanten dieser Klassen.

A.3 Elementare Hilbertraum Theorie

In diesem Abschnitt des Anhangs stellen wir eine Reihe von grundlegenden Aussagen zusammen, die in Hilberträumen gelten. Da wir quadratintegrierbare Zufallsvariable, mit denen wir es ja in der Zeitreihenanalyse in der Regel zu tun haben, als Elemente eines (komplexen oder reellen) L_2 -Raumes auffassen können (vgl. Definition A.3) und diese L_2 -Räume selbst Hilberträume sind, können wir an vielen Stellen im Text die hier zusammengestellten Gleichungen und Ungleichungen Gewinn bringend einsetzen und sparen uns an vielen Stellen direkte Nachweise von Aussagen, die vom Blickpunkt der Hilbertraumtheorie bereits bekannt sind.

Definition A.4 (*Innenproduktraum und Skalarprodukt*)

Ein Vektorraum H über K ($K = \mathbb{R}$ oder $K = \mathbb{C}$) heißt Innenproduktraum, wenn es eine Abbildung $\langle \cdot, \cdot \rangle : H \times H \rightarrow K$ gibt (das innere Produkt oder Skalarprodukt) mit folgenden Eigenschaften

$$\langle x, y \rangle = \overline{\langle y, x \rangle} \quad (\text{A.19})$$

$$\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \quad (\text{A.20})$$

$$\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \quad (\text{A.21})$$

$$\langle x, x \rangle \geq 0 \quad \forall x \in H \quad (\text{A.22})$$

$$\langle x, x \rangle = 0 \Leftrightarrow x = 0, \quad (\text{A.23})$$

dabei seien x, y, z beliebig aus H und α beliebig aus K . Die Norm von x ist definiert als

$$\|x\| := +\sqrt{\langle x, x \rangle}.$$

Im Folgenden bezeichnet $(H, \langle \cdot, \cdot \rangle)$ stets einen Innenproduktraum.

Lemma A.5 (*Cauchy-Schwarz-Ungleichung*)

Aus (A.19)–(A.23) folgt

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|, \quad \forall x, y \in H. \quad (\text{A.24})$$

Beweis: Für $c = -\langle y, x \rangle / \|x\|^2$ mit $x \neq 0$ gilt

$$\begin{aligned} 0 &\leq \|x\|^2 \cdot \|cx + y\|^2 \\ &= \|x\|^2 \cdot (|c|^2 \|x\|^2 + \bar{c} \langle y, x \rangle + c \langle x, y \rangle + \|y\|^2) \\ &= \|x\|^2 \cdot (|\langle x, y \rangle|^2 / \|x\|^2 - 2 |\langle x, y \rangle|^2 / \|x\|^2 + \|y\|^2) \\ &= -|\langle x, y \rangle|^2 + \|x\|^2 \cdot \|y\|^2. \end{aligned}$$

□

Lemma A.6 (*Dreiecksungleichung*)

Für $x, y \in H$ gilt

$$\|x + y\| \leq \|x\| + \|y\|. \quad (\text{A.25})$$

Beweis: Unter Verwendung der Cauchy–Schwarz–Ungleichung ergibt sich

$$\begin{aligned} \|x + y\|^2 &= \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle \\ &\leq \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \end{aligned}$$

□

Aus der Dreiecksungleichung folgt

Bemerkung A.7

Für alle $x, y, z \in H$ gilt:

$$\|x - z\| \leq \|x - y\| + \|y - z\|. \quad (\text{A.26})$$

Damit ist $d(x, y) := \|x - y\|$ eine Metrik auf H . Falls H versehen mit dieser Metrik vollständig ist, d.h. jede Cauchy–Folge konvergiert, so heißt H ein *Hilbertraum*.

Beispiel A.8

a) Der Raum $\mathbb{C}^n = \{x = (x_1, \dots, x_n) : x_i \in \mathbb{C} \forall i\}$ ist ein Hilbertraum mit Skalarprodukt

$$\langle x, y \rangle := \sum_{i=1}^n x_i \overline{y_i},$$

wobei $x = (x_1, \dots, x_n)$ und $y = (y_1, \dots, y_n)$ seien.

b) Für jeden Maßraum $(\mathcal{X}, \mathcal{B}, \mu)$ definiert

$$\langle f, g \rangle := \int f \overline{g} d\mu$$

ein inneres Produkt auf dem Raum $L_2^{\mathbb{C}}(\mathcal{X}, \mathcal{B}, \mu)$ der Äquivalenzklassen der μ –fast sicher übereinstimmenden, quadratintegrierbaren komplexen Funktionen auf $(\mathcal{X}, \mathcal{B})$. Entsprechendes gilt für $L_2^{\mathbb{R}}(\mathcal{X}, \mathcal{B}, \mu)$.

Einen Beweis dieser Aussagen findet man z.B. bei Dudley (1989), Theorem 5.2.1. □

Satz A.9

Für einen Innenproduktraum $(H, \langle \cdot, \cdot \rangle)$ ist $\langle x, y \rangle$ gemeinsam in beiden Variablen stetig, d.h. aus $x_n \rightarrow x, y_n \rightarrow y$ folgt $\langle x_n, y_n \rangle \rightarrow \langle x, y \rangle$.

Beweis:

$$\begin{aligned} &|\langle x_n, y_n \rangle - \langle x, y \rangle| \\ &\leq |\langle x_n, y_n \rangle - \langle x, y_n \rangle| + |\langle x, y_n \rangle - \langle x, y \rangle| \\ &\leq \|x_n - x\|(\|y_n - y\| + \|y\|) + \|x\| \|y_n - y\| \rightarrow 0. \end{aligned}$$

□

Die sich ergebende Identität

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \quad (\text{A.27})$$

heißt *Parallelogrammregel*.

Für $\langle x, y \rangle = 0$ verwenden wir die Schreibweise $x \perp y$ und sagen, dass x und y zueinander *orthogonal* sind oder aufeinander *senkrecht stehen*. Offenbar gilt der Satz des Pythagoras:

$$x \perp y \Rightarrow \|x + y\|^2 = \|x\|^2 + \|y\|^2. \quad (\text{A.28})$$

Für $A \subset H$ sei $A^\perp := \{y : \langle x, y \rangle = 0 \quad \forall x \in A\}$.

Satz A.10 (*Orthogonale Zerlegung*)

Sei $(H, \langle \cdot, \cdot \rangle)$ ein Hilbertraum und F ein abgeschlossener linearer Unterraum. Jedes $x \in H$ kann dann eindeutig zerlegt werden gemäß

$$x = y + z \text{ mit } y \in F \text{ und } z \in F^\perp.$$

Beweis: Sei $c := \inf\{\|x - f\| : f \in F\}$. Wähle $f_n \in F$, $n \geq 1$, mit $\|x - f_n\| \downarrow c$. Die Parallelogrammregel angewandt auf $f_m - x$ und $f_n - x \quad \forall m, n$ liefert

$$\|f_m - f_n\|^2 = 2 \underbrace{\|f_m - x\|^2}_{\rightarrow c^2} + 2 \underbrace{\|f_n - x\|^2}_{\rightarrow c^2} - 4 \underbrace{\|(f_m + f_n)/2 - x\|^2}_{\geq c^2},$$

wobei das Ungleichheitszeichen wegen $(f_m + f_n)/2 \in F$ und der Definition von c gilt. Es folgt $\|f_m - f_n\| \rightarrow 0$ für $m, n \rightarrow \infty$. Es gibt also wegen der Abgeschlossenheit von F ein $y \in F$ mit $f_n \rightarrow y$ für $n \rightarrow \infty$.

Setzt man nun $z = x - y$ so folgt $x - f_n \rightarrow z$, also $\|x - f_n\| \rightarrow \|z\|$ und folglich $\|z\| = c$. Wir zeigen, dass $z \perp F$, also $z \in F^\perp$ gilt.

Unter der Annahme, dass $\langle z, f \rangle \neq 0$ für ein $f \in F$ gilt, definiere man $u = u(v) := \langle z, f \rangle v$ für $v \in \mathbb{R}$. Dann gilt für genügend kleines $v > 0$

$$\begin{aligned} \|x - (y + uf)\|^2 &= \|z - uf\|^2 \\ &= c^2 - 2v |\langle z, f \rangle|^2 + v^2 |\langle z, f \rangle|^2 \cdot \|f\|^2 \\ &< c^2. \end{aligned}$$

Wegen $y + uf \in F$ ergibt sich ein Widerspruch.

Zum Beweis der Eindeutigkeit von y sei $g \in F$ derart, dass $x = g + h$ gilt mit $h \in F^\perp$. Dann folgt $(y - g) + (z - h) = 0$, wobei die beiden Summanden senkrecht aufeinander stehen, und deshalb $\|y - g\|^2 + \|z - h\|^2 = 0$, d.h. $y = g, z = h$. \square

Die Abbildung $P : H \rightarrow F, x \mapsto y$, wobei $x = y + z$ mit $y \in F$ und $z \in F^\perp$ heißt *Orthogonalprojektion* oder auch einfach *Projektion* auf F . Wir schreiben dafür auch P_F .

Eigenschaften einer Orthogonalprojektion P sind

$$P \text{ ist linear} \quad (\text{A.29})$$

$$(x \in F \Leftrightarrow Px = x) \text{ und } (x \in F^\perp \Leftrightarrow Px = 0) \quad (\text{A.30})$$

$$P_V(P_W x) = P_V x \text{ für abgeschl. Unterräume } V \subset W \subset H \quad (\text{A.31})$$

$$P^2 = P \quad (P \text{ ist idempotent}) \quad (\text{A.32})$$

$$\langle Px, y \rangle = \langle x, Py \rangle \quad (P \text{ ist selbstadjungiert}) \quad (\text{A.33})$$

$$\|Px\| \leq \|x\| \quad (\text{A.34})$$

Die Aussage (A.31) heißt *Iterationseigenschaft* von P . Die Eigenschaften (A.32) und (A.33) eines linearen Operators P implizieren, dass P die Orthogonalprojektion auf $\text{im } P = \{Px : x \in H\}$ ist.

Die Eigenschaft (A.34) folgt schon aus (A.32) zusammen mit (A.33). Außerdem impliziert (A.34) die Stetigkeit von P (ist sogar äquivalent). Weiter impliziert (A.32) zusammen mit (A.34) die Abgeschlossenheit von $\text{im } P$.

Nach Konstruktion erfüllt $Px = y$ die Beziehung

$$\|x - Px\|^2 = \inf\{\|x - f\|^2 : f \in F\}. \quad (\text{A.35})$$

Auch hierdurch ist $Px \in F$ charakterisiert, denn sei $f_0 \in F$ derart, dass $\|x - Px\|^2 = \|x - f_0\|^2 =: c^2$ gilt, so folgt mit $x = Px + z = y + z$ und $\|z\|^2 = c^2$ wegen $y - f_0 \in F$

$$c^2 = \|x - f_0\|^2 = \|(y - f_0) + z\|^2 = \|y - f_0\|^2 + c^2,$$

d.h. $y = f_0$.

Orthogonalität

Definition A.11 (*Abgeschlossene Hülle*)

Sei $\{x_t : t \in T\} \subset H$ Teilmenge eines Hilbertraumes H . Dann bezeichnet $\overline{\text{span}}\{x_t : t \in T\}$ die abgeschlossene Hülle der Menge $\text{span}\{x_t : t \in T\}$ aller endlichen Linearkombinationen $\alpha_1 x_{t_1} + \dots + \alpha_n x_{t_n}$ mit $\alpha_i \in K, t_i \in T \forall i$.

Falls T endlich ist, ist die Menge der endlichen Linearkombinationen schon selbst abgeschlossen, wie sukzessive Anwendung des folgenden Lemmas zeigt. Dann bedeutet also span und $\overline{\text{span}}$ dasselbe.

Lemma A.12

Ist V ein abgeschlossener, linearer Teilraum von $(H, \langle \cdot, \cdot \rangle)$ und ist $y \in H, y \notin V$, sowie V^* der von V und y aufgespannte Raum, so ist V^* abgeschlossen.

Beweis: Wir folgen hier Rudin (1970), Abschnitt 4.15. Es gelte für $x_n \in V$ und $\lambda_n \in K$

$$x_n + \lambda_n y \rightarrow z, \quad n \rightarrow \infty. \quad (\text{A.36})$$

Wir zeigen, dass z die Darstellung $z = x + \lambda y$ mit $x \in V$ und $\lambda \in K$ besitzt. Zunächst gibt es ein $\eta < \infty$ mit $\|x_n + \lambda_n y\| < \eta \quad \forall n$. Falls eine Teilfolge von $(|\lambda_n|)_n$ gegen ∞ konvergiert (der Einfachheit halber die Folge selbst), so folgt

$$\left\| \frac{x_n}{\lambda_n} + y \right\| < \frac{\eta}{|\lambda_n|} \rightarrow 0,$$

also $-y \in V$ (da V abgeschlossen ist) im Widerspruch zur Annahme $y \notin V$. Also ist $\lambda_n, n \geq 1$, beschränkt und enthält somit eine konvergente Teilfolge $\lambda_{n'} \rightarrow \lambda$. Wegen $\lambda_{n'} y \rightarrow \lambda y$ folgt aus (A.36) auch $x_{n'} \rightarrow x := z - \lambda y$ mit $x \in V$, da V abgeschlossen ist. Insgesamt:

$$x_{n'} + \lambda_{n'} y \rightarrow x + \lambda y.$$

□

Bemerkung A.13

Ist $M = \text{span}\{x_1, \dots, x_n\}$, so ist die Orthogonalprojektion P_M von H auf M dadurch charakterisiert, dass $P_M x$ die Darstellung

$$P_M x = \alpha_1 x_1 + \dots + \alpha_n x_n$$

besitzt mit geeigneten Konstanten α_i und dass gilt

$$x - P_M x \perp M \quad (\Leftrightarrow x - P_M x \perp x_1, \dots, x_n). \quad (\text{A.37})$$

Die letzte Beziehung ist offenbar gleichbedeutend mit der Gültigkeit der sog. *Normalgleichungen*

$$\sum_{i=1}^n \alpha_i \langle x_i, x_j \rangle = \langle x, x_j \rangle, \quad \forall j = 1, \dots, n. \quad (\text{A.38})$$

Da die Projektion stets existiert, gibt es immer eine Lösung $\alpha = (\alpha_1, \dots, \alpha_n)^T$ der Normalgleichungen, die jedoch nicht eindeutig sein muss.

Eine Lösung ist genau dann eindeutig bestimmt, wenn x_1, \dots, x_n linear unabhängig sind, was wiederum äquivalent damit ist, dass die Matrix $A = (\langle x_i, x_j \rangle)_{i,j=1,\dots,n}$ nicht ausgeartet ist. Letzteres sieht man folgendermaßen ein:

Die Gleichungen (A.38) bedeuten in Matrixform $\alpha^T A = \mathbf{x}^T$ mit $\mathbf{x} = (\langle x, x_1 \rangle, \dots, \langle x, x_n \rangle)^T$. Ist A nicht ausgeartet, so ist also $\alpha = A^{-1} \mathbf{x}$ eindeutig bestimmt. Ist A jedoch ausgeartet, so gibt es einen Vektor $\mathbf{c} \neq 0$ mit $\mathbf{c}^T A = 0$. Dann besitzt (A.38) neben einer beliebigen Lösung α auch die davon verschiedene Lösung $\alpha + \mathbf{c}$.

Definition A.14 (*Orthonormale Menge*)

Eine Menge $\{e_t : t \in T\}$ heißt orthonormal, wenn $e_s \perp e_t \ \forall s \neq t$ und $\|e_t\| = 1 \ \forall t \in T$ gilt.

Satz A.15 (*Projektion auf orthonormale Mengen*)

Sei $\{e_1, \dots, e_k\}$ orthonormal. Für $M = \text{span}\{e_1, \dots, e_k\}$ gilt dann

$$P_M x = \sum_{i=1}^k \langle x, e_i \rangle e_i, \quad \forall x \in H, \quad (\text{A.39})$$

$$\|P_M x\|^2 = \sum_{i=1}^k |\langle x, e_i \rangle|^2, \quad \forall x \in H, \quad (\text{A.40})$$

$$\left\| x - \sum_{i=1}^k \langle x, e_i \rangle e_i \right\| \leq \left\| x - \sum_{i=1}^k c_i e_i \right\|, \quad \forall c_1, \dots, c_k \in K. \quad (\text{A.41})$$

Gleichheit gilt in (A.41) genau dann, wenn $c_i = \langle x, e_i \rangle \ \forall i$.

Beweis: Aus (A.38) folgt mit $x_i = e_i \ \forall i$ die Gleichheit $\alpha_j = \langle x, e_j \rangle \ \forall j$, also (A.39). (A.40) folgt unmittelbar aus (A.39). (A.41) folgt wegen (A.35). \square

Aus (A.40) und (A.34) folgt, falls $\{e_1, \dots, e_k\}$ orthonormal ist

$$\sum_{i=1}^k |\langle x, e_i \rangle|^2 \leq \|x\|^2 \quad (\text{Bessel-Ungleichung}). \quad (\text{A.42})$$

Definition A.16 (*Separabler Hilbertraum*)

Ein Hilbertraum $(H, \langle \cdot, \cdot \rangle)$ heißt separabel, wenn ein endliches oder abzählbar unendliches Orthonormalsystem $\{e_t : t \in T\}$ in H existiert mit $H = \overline{\text{span}}\{e_t : t \in T\}$.

Schließlich gilt

Satz A.17

Sei $H = \overline{\text{span}}\{e_1, e_2, \dots\}$ mit einem Orthonormalsystem $\{e_i : i \in \mathbb{N}\}$. Dann gilt:

a) Die Menge aller endlichen Linearkombinationen der e_i ist dicht in H .

b) $x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$ (d.h. $\lim_{n \rightarrow \infty} \sum_{i=1}^n \langle x, e_i \rangle e_i = x$.)

c) $\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2$

d) $\langle x, y \rangle = \sum_{i=1}^{\infty} \langle x, e_i \rangle \langle e_i, y \rangle$ (Parseval-Gleichung)

e) $x = 0 \Leftrightarrow x \perp e_i \ \forall i \in \mathbb{N}$.

Beweis: a) Für $H_n := \text{span}\{e_1, \dots, e_n\}$ ist offensichtlich $\overline{\bigcup_{n=1}^{\infty} H_n}$ ein abgeschlossener, linearer Teilraum von H , der $\{e_1, e_2, \dots\}$ enthält und muss deshalb gleich H sein.

b) Nach der Bessel-Ungleichung gilt

$\sum_{i=1}^k \langle x, e_i \rangle^2 \leq \|x\|^2 \forall k$, also $\sum_{i=1}^{\infty} \langle x, e_i \rangle^2 \leq \|x\|^2$. Nach (A.41) folgt wegen a), dass durch Wahl von k die Differenz $\|x - \sum_{i=1}^k \langle x, e_i \rangle e_i\| < \varepsilon$ (beliebig) gemacht werden kann. Wegen $\|x - P_{H_n} x\| \leq \|x - P_{H_k} x\|$ für $n \geq k$ gilt $\|x - P_{H_n} x\| \leq \varepsilon \forall n \geq k$. Es folgt: $P_{H_n} x \rightarrow x$.

c) Es gilt wegen $P_{H_n} x \rightarrow x$ auch $\|P_{H_n} x\|^2 \rightarrow \|x\|^2$; mit (A.40) folgt die Behauptung.

d) $\sum_{i=1}^n \langle x, e_i \rangle \langle e_i, y \rangle = \langle P_{H_n} x, P_{H_n} y \rangle \rightarrow \langle x, y \rangle$.

e) Folgt aus c). □

A.4 Lösungen einer homogenen Differenzengleichung

Bei gegebenen reellen Zahlen $a_0 = 1, a_1, a_2, \dots, a_p$, mit $a_p \neq 0$ suchen wir die Gesamtheit V aller komplexen Lösungsfolgen $\{d_j\} = \{d_j : j = 0, 1, 2, \dots\}$ der folgenden homogenen Differenzengleichungen

$$\sum_{k=0}^p d_{j-k} a_k = 0 \quad \text{für} \quad j = p, p+1, \dots \quad (\text{A.43})$$

Mit $A(z) := \sum_{k=0}^p a_k z^k, z \in \mathbb{C}$, gilt: Ist z_0 eine $(n+1)$ -fache Nullstelle von A , so sind die $(n+1)$ Folgen

$$\{j^r z_0^{-j} : j = 0, 1, 2, \dots\}, \quad r = 0, \dots, n, \quad (\text{A.44})$$

sämtlich Lösungen von (A.43), die sogenannten *Basislösungen*, von denen es genau p gibt, wenn man alle Nullstellen von A mit ihren jeweiligen Vielfachheiten durchläuft.

Zum *Beweis* von (A.44) sei z_0 eine $(n+1)$ -fache Nullstelle von A . Dann verschwinden in z_0 alle r -ten Ableitungen $A^{(r)}$ der Ordnung $r \leq n$, d.h. es gilt

$$A^{(r)}(z_0) = \sum_{k=r}^p a_k k(k-1) \cdots (k-r+1) z_0^{k-r} = 0$$

für $r = 1, \dots, n$. Für diese Nullstelle z_0 gilt also

$$\sum_{k=0}^p a_k k(k-1) \cdots (k-r+1) z_0^k = 0 \quad \text{für} \quad r = 1, \dots, n.$$

Mit vollständiger Induktion über r folgt daraus weiterhin

$$\sum_{k=0}^p a_k k^r z_0^k = 0 \quad \text{für} \quad r = 0, \dots, n. \quad (\text{A.45})$$

Für $\{d_j\} = \{j^r z_0^{-j}\}$ ergibt sich nun aus (A.45)

$$\sum_{k=0}^p a_k (j-k)^r z_0^{-(j-k)} = z_0^{-j} \sum_{k=0}^p a_k (j-k)^r z_0^k = 0, \quad r = 0, \dots, n,$$

also die Eigenschaft, dass $\{d_j\}$ die Differenzengleichung (A.43) löst.

Sind $f_1 = \{f_{1j}\}, \dots, f_p = \{f_{pj}\}$ alle Basislösungen von (A.43), so ist die mit beliebigen Konstanten $\kappa_1, \dots, \kappa_p$ aus \mathbb{C} gebildete Folge

$$d_j = \sum_{l=1}^p \kappa_l f_{lj}, \quad j = 0, 1, 2, \dots \quad (\text{A.46})$$

ebenfalls eine Lösung der Differenzengleichung.

Mit $\mathbf{d} = (d_0, \dots, d_{p-1})^T$, $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_p)^T$ und der $p \times p$ -Matrix $F = (f_{i,j-1})_{i,j=1,\dots,p}$ gilt

$$\mathbf{d} = F^T \boldsymbol{\kappa}. \quad (\text{A.47})$$

Man kann zeigen, dass F nicht ausgeartet ist (siehe Meschkowski (1959), S. 94-98 oder Brockwell/Davis (1991), § 3.6). Damit bilden die p Basislösungen eine Basis von V .

Wird also \mathbf{d} , d.h. die p Startwerte einer Lösungsfolge, vorgegeben, so ist $\boldsymbol{\kappa} = (F^T)^{-1} \mathbf{d}$ und damit die komplette Lösungsfolge eindeutig bestimmt.

Im Hinblick auf unsere Anwendungen interessieren uns häufig nur reelle Lösungsfolgen $\{d_j\}$. Wegen $d_j = \bar{d}_j \forall j$ folgt dann aus (A.46) mit $S_{lj} := (\kappa_l f_{lj} + \bar{\kappa}_l \bar{f}_{lj})/2$ die Darstellung

$$d_j = \sum_{l=1}^p S_{lj}. \quad (\text{A.48})$$

Jeder Summand S_{lj} gehört zu einer Nullstelle z_0 von A , für die entweder $z_0 \in \mathbb{R}$ oder $z_0 \neq \bar{z}_0$ gilt. Wir machen deshalb die folgende Fallunterscheidung:

1. Fall ($A(z_0) = 0, z_0 \in \mathbb{R}$): Dann hat S_{lj} die Gestalt

$$\kappa j^r z_0^{-j}, \quad j = 0, 1, 2, \dots, \quad (\text{A.49})$$

mit $\kappa \in \mathbb{R}$, $r = 0, \dots, n$, wobei $n+1$ die Vielfachheit der Nullstelle z_0 ist.

2. Fall ($A(z_0) = 0, z_0 \neq \bar{z}_0$): Mit z_0 ist auch \bar{z}_0 Nullstelle von A mit gleicher Vielfachheit. Deshalb gibt es einen Index l' mit $f_{l'j} = \bar{f}_{lj}$. Dann gilt mit $\kappa := \kappa_l + \bar{\kappa}_{l'} =: \rho e^{-i\theta}$, $\rho > 0$, sowie $z_0 = \lambda e^{i\varphi}$, $\lambda > 0$, $\theta, \varphi \in (-\pi, \pi]$, $\varphi \notin \{0, \pi\}$, die Darstellung

$$\begin{aligned}
S_{lj} + S_{l'j} &= j^r (\kappa z_0^{-j} + \bar{\kappa} \bar{z}_0^{-j})/2 \\
&= \rho j^r \lambda^{-j} (e^{i(j\varphi+\theta)} + e^{-i(j\varphi+\theta)})/2 \\
&= \rho j^r \lambda^{-j} \cos(j\varphi + \theta), \quad j = 0, 1, 2, \dots
\end{aligned}$$

Die Zusammenfassung der entsprechenden Summanden mit z_0 und \bar{z}_0 ergibt also einen Beitrag

$$\rho j^r \lambda^{-j} \cos(j\varphi + \theta), \quad j = 0, 1, 2, \dots, \quad (\text{A.50})$$

mit $\rho > 0$, $\varphi, \theta \in (-\pi, \pi]$, $\varphi \notin \{0, \pi\}$, $r = 0, \dots, n$, wobei $n+1$ die Vielfachheit der Nullstelle z_0 ist. Die Parameter ρ und θ sind unter den angegebenen Einschränkungen frei wählbar, während λ und φ durch z_0 bestimmt sind.

A.5 Konvergenzbegriffe in der Stochastik

In diesem Teil des Anhangs stellen wir Definitionen und Rechenregeln für die Konvergenz nach Wahrscheinlichkeit (auch stochastische Konvergenz genannt), die Konvergenz im r -ten Mittel und die Verteilungskonvergenz bereit. Dabei werden wir nur Zufallsvariable mit Werten in \mathbb{R}^k , $k \geq 1$, betrachten, da in diesem Zusammenhang nur solche im Text vorkommen.

Konvergenz nach Wahrscheinlichkeit und L_r -Konvergenz

Im Folgenden seien X, X_1, X_2, \dots und Y, Y_1, Y_2, \dots \mathbb{R}^k -wertige Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Definition A.18 (Stochastische Konvergenz)

Man sagt, dass die Folge $\{X_n\}$ nach Wahrscheinlichkeit oder stochastisch gegen $X \in \mathbb{R}^k$ konvergiert, kurz: $X_n \rightarrow X$ n.W. oder $X_n - X = o_P(1)$, falls für $n \rightarrow \infty$

$$P\{|X_n - X| \geq \varepsilon\} \rightarrow 0, \quad \forall \varepsilon > 0, \quad (\text{A.51})$$

gilt, wobei $|\cdot|$ eine beliebige Norm auf \mathbb{R}^k ist.

Die Aussage (A.51) hängt nicht von der Norm $|\cdot|$ ab, da in \mathbb{R}^k alle Normen äquivalent sind, d.h. eine andere Norm $\|\cdot\|$ erfüllt für ein festes $\varrho > 1$ die Einschachtelungsungleichungen $\varrho^{-1}\|x\| \leq |x| \leq \varrho\|x\| \quad \forall x \in \mathbb{R}^k$. Wenn also (A.51) für eine Norm gilt, so folgt die gleiche Aussage sofort für alle Normen auf \mathbb{R}^k . Wählt man z.B. die Maximumsnorm, so impliziert die Gleichung

$$\{|X_n - X| \geq \varepsilon\} = \bigcup_{i=1}^k \{|X_{ni} - X_i| \geq \varepsilon\},$$

dass die Aussage $X_n - X = o_P(1)$ äquivalent zur komponentenweise Aussage $X_{ni} - X_i = o_P(1) \quad \forall i = 1, \dots, k$ ist.

Für den Spezialfall, dass $X = a$ eine Konstante ist, bleibt die Definition $X_n \rightarrow a$ n.W. auch dann noch sinnvoll, wenn die Zufallsvektoren X_n sogar auf mit n variierenden Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{A}_n, P_n)$ erklärt sind, was in der Schreibweise aber regelmäßig unterdrückt wird.

Für manche Beweise ist das folgende Teilfolgenprinzip hilfreich, das die stochastische Konvergenz auf die fast sichere Konvergenz zurückführt. Wegen obiger Bemerkung bzgl. komponentenweiser Konvergenz reicht es, den Fall $k = 1$ zu betrachten. Bekanntlich konvergiert eine reelle Folge $\{X_n\}$ fast sicher gegen X , kurz $X_n \rightarrow X$ f.s., wenn

$$P\{\limsup_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} X_n = X\} = 1$$

gilt.

Satz A.19 (Teilfolgenprinzip für stochastische Konvergenz)

Genau dann konvergiert eine reelle Folge $\{X_n\}$ stochastisch gegen X , wenn zu jeder Teilfolge $\{n'\} \subset \{n\}$ eine weitere Teilfolge $\{n''\} \subset \{n'\}$ existiert mit $X_{n''} \rightarrow X$ f.s. .

Zum Beweis siehe Bauer (1990), Korollar 20.8.

Definition A.20 (Straffheit)

Man nennt eine Folge von Wahrscheinlichkeitsmaßen $\{\mu_n\}$ über $(\mathbb{R}^k, \mathbb{B}^k)$ straff (Englisch: tight), wenn es für alle $\varepsilon > 0$ ein $M(\varepsilon) < \infty$ gibt mit

$$\mu_n \{x \in \mathbb{R}^k : |x| \geq M(\varepsilon)\} < \varepsilon, \quad \forall n \geq 1. \quad (\text{A.52})$$

Die Folge $\{X_n\}$ heißt straff oder nach Wahrscheinlichkeit beschränkt, kurz: $X_n = \mathcal{O}_P(1)$, wenn die Folge ihrer Verteilungen $\{\mu_n := \mathcal{L}(X_n)\}$ straff ist.

Falls $\{c_n\}$ eine Folge in $(0, \infty)$ ist, erklärt man die stochastischen Landau-Symbole $\mathcal{O}_P(c_n)$ und $o_P(c_n)$ durch

$$X_n = \mathcal{O}_P(c_n) :\Leftrightarrow X_n = Y_n \cdot c_n \quad \text{und} \quad Y_n = \mathcal{O}_P(1) \quad (\text{A.53})$$

$$X_n = o_P(c_n) :\Leftrightarrow X_n = Y_n \cdot c_n \quad \text{und} \quad Y_n = o_P(1) \quad (\text{A.54})$$

Sind X_n keine Zufallsvektoren, sondern feste Vektoren, so reduzieren sich die stochastischen Landau-Symbole auf die üblichen Landau-Symbole $\mathcal{O}(c_n)$ und $o(c_n)$.

Bemerkung A.21 (Rechenregeln für stochastische Konvergenz)

Sind X, X_1, X_2, \dots und Y, Y_1, Y_2, \dots \mathbb{R}^k -wertige Zufallsvariable über (Ω, \mathcal{A}, P) und ist $g : \mathbb{R}^k \rightarrow \mathbb{R}$ eine weitere messbare Abbildung, so gelten für die stochastische Konvergenz die folgen Rechenregeln

- (i) Ist g stetig im Punkt $a \in \mathbb{R}^k$ und gilt $X_n \rightarrow a$ n.W., so folgt $g(X_n) \rightarrow g(a)$ n.W.; in Landau-Symbolik:

$$X_n = a + o_P(1) \text{ impliziert } g(X_n) = g(a) + o_P(1).$$

(ii) Gilt $X_n = X + o_P(1)$ und $Y_n = Y + o_P(1)$, so folgt

$$X_n + Y_n = X + Y + o_P(1)$$

(iii) Gilt $X_n = X + o_P(1)$ und $Y_n = Y + o_P(1)$, so folgt

$$X_n^T Y_n = X^T Y + o_P(1).$$

Zum Beweis siehe Behnen/Neuhaus (2003), Abschnitt 8.

In Anhang A.2 hatten wir schon den Begriff der L_r -Konvergenz für \mathbb{C} -oder \mathbb{R} -wertige Zufallsvariable eingeführt ($k = 1$). Wegen der Markov-Ungleichung

$$P\{|X| \geq \varepsilon\} \leq \frac{1}{\varepsilon^r} E|X|^r, \quad \forall \varepsilon > 0, r > 0, \quad (\text{A.55})$$

folgt für X, X_1, X_2, \dots aus $L_r^{\mathbb{R}}$ und $n \rightarrow \infty$

$$X_n \rightarrow X \text{ in } L_r^{\mathbb{R}}(P) \text{ impliziert } X_n \rightarrow X \text{ n.W.} \quad (\text{A.56})$$

Häufig verwendet wird auch das folgende Kriterium zum Nachweis der L_2 -Konvergenz und auch der stochastischen Konvergenz.

Satz A.22 (*Kriterium zur L_2 -Konvergenz*)

Für X, X_1, X_2, \dots aus $L_2^{\mathbb{R}}(P)$ folgt aus $EX_n \rightarrow \mu$ und $\text{Var} X_n \rightarrow 0$ die Aussage

$$X_n \rightarrow \mu \text{ in } L_2^{\mathbb{R}}(P) \text{ für } n \rightarrow \infty \quad (\text{A.57})$$

und damit auch $X_n \rightarrow \mu$ n.W.

Beweis: $E|X_n - \mu|^2 = E|X_n - EX_n|^2 + |EX_n - \mu|^2 \rightarrow 0$ für $n \rightarrow \infty$. \square

Verteilungskonvergenz

Wir wollen nun den Begriff der Verteilungskonvergenz aufgreifen und einige für diesen Text wichtige Aussagen zusammen tragen. Die Beweise, auf die an dieser Stelle vollständig verzichtet wird, findet man in dem Buch von Billingsley (1968).

Seien im Folgenden X, X_1, X_2, \dots \mathbb{R}^k -wertige Zufallsvariable.

Definition A.23 (*Verteilungskonvergenz in \mathbb{R}^k*)

Man sagt, dass die Folge $\{X_n\}$ nach Verteilung gegen X konvergiert, kurz $X_n \xrightarrow{D} X$, falls für $n \rightarrow \infty$

$$P\{X_n \leq x\} \rightarrow P\{X \leq x\}, \quad \forall x \in C(F), \quad (\text{A.58})$$

gilt, wobei $C(F)$ die Menge der Stetigkeitspunkte der Verteilungsfunktion $F(x) = P\{X \leq x\}$ ist.

Offensichtlich kommt es in (A.58) nur auf die induzierten Verteilungen P^{X_n} und P^X an, weshalb die Zufallsvariablen X_n und X auf verschiedenen Wahrscheinlichkeitsräumen definiert sein dürfen, was wir in der Schreibweise jedoch vernachlässigen.

Ist $F_n(x) = P\{X_n \leq x\}$ die Verteilungsfunktion (auf \mathbb{R}^k) von X_n , so schreibt man auch $F_n \xrightarrow{\mathcal{D}} F$ oder $P^{X_n} \xrightarrow{\mathcal{D}} P^X$ statt $X_n \xrightarrow{\mathcal{D}} X$. Auch gemischte Schreibweisen wie z.B. $X_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ im 1-dimensionalen Fall mit der Standardnormalverteilung $\mathcal{N}(0, 1)$ sind üblich.

Äquivalente Formulierungen zur Verteilungskonvergenz sind im folgenden Satz angegeben.

Satz A.24 (*Charakterisierung der Verteilungskonvergenz in \mathbb{R}^k*)

Die folgenden Aussagen sind äquivalent

$$X_n \xrightarrow{\mathcal{D}} X \quad (\text{A.59})$$

$$Ee^{it^T X_n} \rightarrow Ee^{it^T X}, \quad \forall t \in \mathbb{R}^k, \quad (\text{A.60})$$

$$Eg(X_n) \rightarrow Eg(X) \text{ für jede stet., beschr. Funktion } g: \mathbb{R}^k \rightarrow \mathbb{R}. \quad (\text{A.61})$$

Die Abbildung $t \mapsto Ee^{it^T X}$ wird die charakteristische Funktion von X genannt. Die Äquivalenz (A.59) \Leftrightarrow (A.60) heißt Stetigkeitssatz für charakteristische Funktionen. (X_n , X und t werden in (A.60) natürlich als Spaltenvektoren betrachtet.)

Im folgenden Resultat wird die Verteilungskonvergenz von Vektoren auf eindimensionale Zufallsvariable zurückgeführt.

Satz A.25 (*Die Cramér–Wold–Technik*)

$X_n \xrightarrow{\mathcal{D}} X$ (Verteilungskonvergenz in \mathbb{R}^k) ist äquivalent zu

$$c^T X_n \xrightarrow{\mathcal{D}} c^T X, \quad \forall c \in \mathbb{R}^k. \quad (\text{A.62})$$

Beweis: Aus $X_n \xrightarrow{\mathcal{D}} X$ folgt nach (A.60) $\forall c \in \mathbb{R}^k$

$$Ee^{is(c^T X_n)} \rightarrow Ee^{is(c^T X)}, \quad \forall s \in \mathbb{R}, \quad (\text{A.63})$$

also wegen (A.60) \Rightarrow (A.59) die Aussage (A.62). Umgekehrt folgt aus (A.62) und (A.59) die Aussage (A.60) auf $\{c^T X_n\}$ angewandt

$$Ee^{ic^T X_n} = Ee^{i \cdot 1 \cdot (c^T X_n)} \rightarrow Ee^{i \cdot 1 \cdot (c^T X)} = Ee^{ic^T X}, \quad \forall c \in \mathbb{R}^k,$$

also $X_n \xrightarrow{\mathcal{D}} X$. □

Im Umgang mit verteilungskonvergenten Zufallsvariablen sind die nachfolgenden Rechenregeln hilfreich.

Bemerkung A.26 (Rechenregeln für Verteilungskonvergenz)
Es gelten

- (i) $X_n \rightarrow X$ n.W. $\Rightarrow X_n \xrightarrow{\mathcal{D}} X$.
- (ii) Ist $\{Y_n\}$ eine weitere Folge von Zufallsvariablen in \mathbb{R}^k , so gilt:
 - a) Aus $X_n \xrightarrow{\mathcal{D}} X$ und $X_n - Y_n = o_P(1)$ folgt $Y_n \xrightarrow{\mathcal{D}} X$.
 - b) Aus $X_n \xrightarrow{\mathcal{D}} X$ und $Y_n \rightarrow a$ n.W. mit einer Konstanten $a \in \mathbb{R}^k$ folgt $Y_n^T X_n \xrightarrow{\mathcal{D}} a^T X$.
- Beide Aussagen nennt man das *Lemma von Slutsky*.
- (iii) Aus $X_n \xrightarrow{\mathcal{D}} a, a \in \mathbb{R}^k$ (konstant), folgt $X_n \rightarrow a$ n.W.
- (iv) Aus $X_n \xrightarrow{\mathcal{D}} X$ und $Y_n \rightarrow a$ n.W., $a \in \mathbb{R}^k$ (konstant), folgt $(X_n^T, Y_n^T)^T \xrightarrow{\mathcal{D}} (X^T, a^T)^T$.
- (v) Für reelle Zufallsvariable $X_n \xrightarrow{\mathcal{D}} X$ mit $\sup_{n \geq 1} EX_n^2 < \infty$ gilt Konvergenz der ersten Momente, d.h. $EX_n \rightarrow EX$.

Satz A.27 (*Continuous mapping theorem*)

Ist $g: \mathbb{R}^k \rightarrow \mathbb{R}^m$ eine messbare Abbildung und gilt für die stets Borel-messbare Menge $D_g := \{x \in \mathbb{R}^k : g \text{ ist unstetig in } x\}$ die Aussage $P\{X \in D_g\} = 0$, so folgt aus $X_n \xrightarrow{\mathcal{D}} X$ auch $g(X_n) \xrightarrow{\mathcal{D}} g(X)$.

Satz A.28 (*Prohorov*)

Ist $\{X_n\}$ straff, so gibt es eine Teilfolge $\{X_{n(r)} : r \in \mathbb{N}\}$ und eine Zufallsvariable X mit $X_{n(r)} \xrightarrow{\mathcal{D}} X$ für $r \rightarrow \infty$.

In Wahrscheinlichkeitsmaßen ausgedrückt: Ist $\{\mu_n\}$ straff, so gibt es eine Teilfolge $\{\mu_{n(r)}\}$ und ein Wahrscheinlichkeitsmaß μ mit $\mu_{n(r)} \xrightarrow{\mathcal{D}} \mu$ für $r \rightarrow \infty$. Umgekehrt ist jede verteilungskonvergente Folge $\{X_n\}$ straff.

Satz A.29

Seien $X_n, Y_{nj}, Y_j, n \geq 1, j \geq 1, Y$ \mathbb{R}^k -wertige Zufallsvektoren mit

$$Y_{nj} \xrightarrow{\mathcal{D}} Y_j \text{ für } n \rightarrow \infty, \quad \forall j \geq 1, \quad (\text{A.64})$$

$$Y_j \xrightarrow{\mathcal{D}} Y \text{ für } j \rightarrow \infty, \quad (\text{A.65})$$

$$\limsup_{n \rightarrow \infty} P\{|X_n - Y_{nj}| \geq \varepsilon\} \rightarrow 0 \text{ für } j \rightarrow \infty, \forall \varepsilon > 0. \quad (\text{A.66})$$

Dann folgt

$$X_n \xrightarrow{\mathcal{D}} Y \text{ für } n \rightarrow \infty. \quad (\text{A.67})$$

Satz A.30 (Zentraler Grenzwertsatz (CLT) von Lindeberg)

Für $n \in \mathbb{N}$ seien X_{1n}, \dots, X_{nn} stochastisch unabhängige reelle Zufallsvariable mit $EX_{in} = 0$ für $i = 1, \dots, n$ und $\sigma_{in}^2 := \text{Var} X_{in} \in (0, \infty)$. Mit $\tau_n^2 := \sigma_{1n}^2 + \dots + \sigma_{nn}^2$ gelte die sogenannte Lindeberg-Bedingung

$$\tau_n^{-2} \sum_{i=1}^n EX_{in}^2 1\{|X_{in}| \geq \varepsilon \cdot \tau_n\} \rightarrow 0 \text{ für } n \rightarrow \infty \forall \varepsilon > 0. \quad (\text{A.68})$$

Dann gilt

$$\frac{1}{\tau_n} \sum_{i=1}^n X_{in} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \text{ für } n \rightarrow \infty, \quad (\text{A.69})$$

wobei $\mathcal{N}(0, 1)$ die Standard-Normalverteilung bezeichne.

Die folgende Aussage ist eine Version der sogenannten Delta-Methode.

Satz A.31 (Delta-Methode)

Für \mathbb{R}^k -wertige Zufallsvariable Z, X_1, X_2, \dots und eine reelle Folge $\{r_n\}$ mit $r_n \rightarrow \infty$ gelte mit $a \in \mathbb{R}^k$

$$r_n (X_n - a) \xrightarrow{\mathcal{D}} Z. \quad (\text{A.70})$$

Ist dann $g : \mathbb{R}^k \rightarrow \mathbb{R}^m$ messbar und differenzierbar in a mit $m \times k$ -Jacobi-Matrix $g'(a)$, so folgt

$$r_n (g(X_n) - g(a)) \xrightarrow{\mathcal{D}} g'(a) Z. \quad (\text{A.71})$$

Beweis: Der Einfachheit halber beweisen wir nur den Fall $k = m = 1$. Wegen $r_n \rightarrow \infty$ gilt $X_n - a = \frac{1}{r_n} (r_n (X_n - a)) = o(1) \cdot \mathcal{O}_P(1) = o_P(1)$. Sei nun $h : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch $h(0) = 0$ und

$$h(x) = (g(a+x) - g(a) - g'(a) \cdot x) / x, \text{ für } x \neq 0.$$

Wegen der Differenzierbarkeit von g ist h stetig in 0, sodass nach Bemerkung A.21 (i) $h(X_n - a) \rightarrow 0$ n.W. folgt. Wegen (A.70) impliziert dies

$$\begin{aligned} & r_n (g(X_n) - g(a)) - g'(a) r_n (X_n - a) \\ &= r_n (X_n - a) \cdot h(X_n - a) = \mathcal{O}_P(1) o_P(1) = o_P(1). \end{aligned}$$

Da $g'(a) r_n (X_n - a) \xrightarrow{\mathcal{D}} g'(a) Z$ folgt aus dem Lemma von Slutsky die Behauptung (A.71). \square

A.6 Die Moore–Penrose–Inverse

Zu jeder $n \times m$ -Matrix $A \in \mathbb{R}^{n \times m}$ gibt es genau eine $m \times n$ -Matrix $A^+ \in \mathbb{R}^{m \times n}$, die *Moore–Penrose–Inverse*, derart, dass die beiden Matrizen

$$AA^+ \in \mathbb{R}^{n \times n} \quad \text{und} \quad A^+A \in \mathbb{R}^{m \times m} \quad (\text{A.72})$$

symmetrisch sind und

$$AA^+A = A \quad \text{sowie} \quad A^+AA^+ = A^+ \quad (\text{A.73})$$

gelten.

Falls $m = n$ gilt und A invertierbar ist, so folgt $A^+ = A^{-1}$.

Man kann A^+ definieren als Limes

$$A^+ = \lim_{\delta \rightarrow \infty} A^T (AA^T + \delta^2 E_n)^{-1}, \quad (\text{A.74})$$

wobei E_n die $n \times n$ -Einheitsmatrix bezeichne, vgl. Albert (1972), S. 19f.

Für jedes $\delta \neq 0$ ist $A^T (AA^T + \delta^2 E_n)^{-1}$ eine stetige Funktion der Komponenten von A , sodass nach (A.74) A^+ messbar von den Einträgen der Matrix A abhängt.

Für $m = n = 1$ gilt $a^+ = a^{-1}$ falls $a \neq 0$ und $a^+ = 0$ für $a = 0$. Folglich ist $a \mapsto a^+$ nicht stetig. Jedoch ist die Abbildung $A \mapsto A^+$ für $m \times m$ -Matrizen stetig in allen A , für die A^{-1} existiert.

Literaturverzeichnis

- Akaike, H. (1969): Fitting autoregressive models for prediction. *Ann. Inst. Math. Statist.* **21**, 243–247.
- Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In: B. N. Petrov and F. Csaki (eds.) *2nd International Symposium on Information Theory*, 267–281. Akademiai Kiado, Budapest.
- Akaike, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **AC-19**, 716–723.
- Akaike, H. (1978): Time series analysis and control through parametric models. In: D. F. Findley (ed.) *Applied Time Series Analysis*. Academic Press, New York.
- Albert, A. (1972): *Regression and the Moore-Penrose Pseudoinverse*. Academic Press, New York.
- Ash, R.B. and Gardner, M.F.(1975): *Topics in Stochastic Processes*. Academic Press, New York.
- Bauer, H. (1990): *Maß- und Integrationstheorie*. De Gruyter, Berlin.
- Bauer, H. (1991): *Wahrscheinlichkeitstheorie*. 4. Auflage, de Gruyter, Berlin.
- Baxter, G. (1962): An asymptotic result for the finite predictor. *Math. Scand.* **10**, 137–144.
- Behnen, K. und Neuhaus, G. (2003): *Grundkurs Stochastik*. 4.Auflage, PD-Verlag, Heidenau.
- Beran, J. (1994): *Statistics for Long-Memory Processes*. Chapman Hall, Englewood, NJ.
- Berk, K.N. (1974). Consistent Autoregressive Spectral Estimates. *Ann. Statist.* **2**, 489–502.
- Berkes, I., Horváth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. *Bernoulli* **9**, 201–228.
- Billingsley, P. (1968): *Convergence of Probability Measures*. Wiley, New York.
- Black, F. und Scholes, M. (1973): The Pricing of Options and Corporate Liabilities. *J. Polit. Economy* **81**, 637–654.
- Bollerslev, T. (1986): Generalized Autoregressive Conditional Heteroscedasticity. *J. Econometrics* **31**, 307–327.
- Bosq, D. (1996): *Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics* **110**, Springer, Berlin.
- Bougerol, P. (1993): Kalman Filtering with Random Coefficients and Contractions. *SIAM J. Control Optim.* **31**, 942–959.

- Box, G.E.P. and Jenkins, G.M. (1970): *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods*. Second Edition. Springer, New York.
- Brown, B.M. (1971): Martingale central limit theorems. *Ann. Math. Statist.* **42**, 54–66.
- Campbell, M.J. and Walker, A.M. (1977): A Survey of statistical work on the Mackenzie River series of annual Canadian lynx trappings for the years 1821–1934 and a new analysis. *Journal of the Royal Statistical Society series A* **140**, 432–436.
- Coulon-Prieur, C. and Doukhan, P. (2000): A triangular central limit theorem under a new weak dependence condition. *Statist. Prob. Letters* **47**, 61–68.
- Dahlhaus, R. (1985): Asymptotic Normality of Spectral Estimates. *J. Mult. Analysis* **16**, 412–431.
- Doukhan, P. (1994): *Mixing: Properties and Examples. Lecture Notes in Statistics* **85**, Springer, Berlin.
- Doukhan, P. and Louhichi, S. (1991): A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications* **84**, 313–342.
- Dudley, R.M. (1989): *Real Analysis and Probability*. Wadsworth & Brooks/Cole, Belmont.
- Dunford, N. and Schwartz, J.T. (1957): *Linear Operators, Part I*. Interscience Publishers, New York.
- Engle, R. (1982): Autoregressive conditional heteroscedastic models with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.
- Forster, O. (1976): *Analysis I*. Vieweg-Verlag, Wiesbaden.
- Gänssler, P. und Stute, W. (1977): *Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin.
- Giraitis, L. and Robinson, P. M. (2001): Whittle estimation of ARCH models. *Econometric Theory* **17**, 608–631.
- Gradstein, I.S. und Ryshik, I.M. (1981): *Tafeln Band I und Band II*. Verlag Harri Deutsch, Thun.
- Hamilton, J.D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.
- Hannan, E.J. (1970): *Multiple Time Series*. J. Wiley, New York.
- Hannan, E.J. (1980): The estimation of the order of an ARMA process. *Ann. Statist.* **8**, 1071–1081.
- Hannan, E.J. and Deistler, M. (1988): *The Statistical Theorie of Linear Systems*. J. Wiley, New York.
- Heuser, H. (1981): *Lehrbuch der Analysis, Teil 2*. B.G.Teubner, Stuttgart.
- Hurvich, C.M. and Tsai, C.L. (1989): Regression and time series model selection in small samples. *Biometrika* **76**, 297–307.
- Izenman, A. J. (1983): J. R. Wolf and H. A. Wolfer: An Historical Note on the Zurich Sunspot Relative Numbers. *J. R. Statist. Soc. A*, **146**, Part 3, 311–318.
- Janas, D., Dahlhaus, R. (1996): A Frequency Domain Bootstrap for Ratio Statistics in Time Series Analysis. *Ann. Statist.* **24**, 1934–1963.
- Kromer, R. E. (1970): Asymptotic Properties of the Autoregressive Spectral Estimator. Ph.D. dissertation, Stanford University.
- Levinson, N. (1947): The Wiener RMS error criterion in the filter design and prediction. *J.Math.Physics* **25**, 261–278.

- Lütkepohl, H. (1993): *Introduction to Multiple Time Series Analysis*. 2nd edition, Springer, Berlin.
- Merton, R.C. (1973). Theory of rational option pricing. *Bell J. Econom. Management Sci.* **4**, 141–183.
- Meschkowski, H. (1959): *Differenzengleichungen*. Vandenhoeck & Ruprecht, Göttingen.
- Mikosch, T and Straumann, D. (2002): Whittle estimation in a heavy-tailed GARCH(1,1)-model. *Stochastic Process. Appl.* **100**, 187–222.
- Nelson, D. (1990): Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.
- Neumann, M.H. and Paparoditis, E. (2003): Goodness-of-fit tests for markovian time series models. *Preprint*, Univ. Braunschweig.
- Reinsel, G.C. (2003): *Elements of Multivariate Time Series Analysis*. Springer, Berlin.
- Remmert, R. (1984 u. 1991): *Funktionentheorie I*. Springer, Berlin.
- Rudin, W. (1970): *Real and Complex Analysis*. McGraw-Hill, New York.
- Schwabe, H. (1844): Sonnenbeobachtungen im Jahre 1843. *Astronomische Nachrichten* **21**, 233–236.
- Shephard, N. (1996): Statistical aspects of ARCH and stochastic volatility. In: D.R. Cox, D. Hinkley and O.E. Barndorff-Nielsen (eds.) *Likelihood, Time Series with Econometric and other Applications*. Chapman and Hall, London.
- Shibata, R. (1980): Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8**, 147–164.
- Spitzer, F. (1976). *Principles of Random Walk*. 2nd edition. Springer, Berlin.
- Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series. Lecture Notes in Statistics* **181**, Springer, Berlin.
- Tong, H. (1983): *Threshold Models in Non-Linear Time Series Analysis. Lecture Notes in Statistics* **21**, Springer, Berlin.
- Tong, H. (1990): *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.
- Toussoun, O. (1925): Mémoire sur l'Histoire du Nil. *Mémoires de l'Institut d'Egypte* **18**, 366–404.
- Waldmeier, M. (1961): *The Sunspot-Activity in the Years 1610-1960*. Schulthess & CO AG Verlag, Zürich.
- Whittle, P. (1963): On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix *Biometrika* **50**, 129–134.
- Wolf, J.R. (1852): Neue Untersuchungen über die Periode der Sonnenflecken und ihre Bedeutung. *Mittheilungen der Naturforschenden Gesellschaft in Bern* **255**, 249–270.
- Wolfer, A. (1925): Observed sunspot relative numbers, 1749-1924. *Terrestrial Magnetism and Atmospheric Electricity* **30**, 83–86.
- Yule, G.U. (1927): On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Phil.Trans.Roy. Soc. Section A*, 389–420.

Index

A

abgeschlossene Hülle, 369
AIC (Akaike Information Criterion),
278, 287
AICC–Ordnungswahl, 288
Amplitudenspektrum, 341
Anpassungsfehler, 337
Arbeitslosenzahlen, 1
ARCH–, 30
Argumentfunktion, 341
ARMA–Gleichung, 109–112
Autokorrelation(sfunktion), 18, 47
 empirische, 31, 171
 empirische partielle, 41, 72
 partielle, 39, 40
 Stichproben–, 31
Autokovarianz(funktion), 18, 47
 Charakterisierung der, 48
 empirische, 31, 171
 erzeugende Funktion der, 108
 Grundeigenschaften der, 48
 Stichproben–, 31
Autokovarianz(matrix), 325

B

Bandweite, 263, 268
Bartlett, *siehe* Lag Window
Basislösungen, 127, 372
Beobachtungsgleichungen, 356
beschränkt nach Wahrscheinlichkeit,
375
Bessel–Ungleichung, 371
BIC–Kriterium, 290

Blackman–Tukey, *siehe* Lag Window

C

Canadian Lynx Data, 5
Cauchy–Schwarz–Ungleichung, 366
charakteristische Funktion, 377
Chi–Quadrat–Verteilung, 44
CLT, *siehe* Zentraler Grenzwertsatz
continuous mapping theorem, 378
Cramér–Wold–Technik, 377

D

Daniell, *siehe* Lag Window
DAX, 5
d–Differenz, 98
Delta–Methode, 379
Differenzgleichung, 126–129, 372–374
 erster Ordnung, 25
 zweiter Ordnung, 44
Differenzenoperator, 95

E

equivalent degrees of freedom (*EDF*),
262
Ergodensatz, 154
Eulersche Konstante, 306
exponentielles Glätten, 11, 72, 100

F

Fejér–Kern, 272
Filter, 91
 high pass–, 91, 96
 Kalman–, 358
 lineare, 93

lineare L_1 –, 119
 low pass–, 91, 98
 multivariate, 335
 Filterfunktion, 91
 Final Prediction Error (FPE), 278, 283, 286
 Fourier–Frequenz, 232
 Fouriertransformation (finite), 234
 FPE , 286
 Frequenz, 52
 Fourier–, 232
 Nyquist–, 52

G

geometrisches Abklingen, 116
 Gewichtsfunktion, 244
 gleitendes Mittel, 10, 98
 einseitiges, 11
 symmetrisches, 11
 Glättung, 244, 255
 Gram–Schmidt–Orthogonalisierung, 66
 multivariate, 355

H

Hamming, *siehe* Lag Window
 Hanning, *siehe* Lag Window
 Herglotz–Lemma, 49
 multivariates, 328
 Hermite, 48, 328
 Hilbertraum, 367

I

Identifizierbarkeit, 133
 i.i.d., 2
 Innenproduktraum, 366
 Innovation, 67
 integrierter Periodogrammschätzer, 255, 262
 Inversionsformel, 53
 multivariate, 331
 Invertibilität, 117
 Isometriesatz, 79
 Iterationseigenschaft, 66, 68, 369

K

Kalman–Filter, 358
 Kausalität, 116
 Kern, 263
 Kleinste–Quadrate(LS)–Schätzer, 9, 195

Konfidenzellipse, 191, 212, 352
 Konvergenz
 im quadratischen Mittel, 365
 in L_r , 365
 nach Verteilung, 376
 stochastische oder n.W., 374
 Korrelation(smatrix), 327
 Korrelogramm, 27, 55
 Kospektrum, 335, 339
 Kovarianzfunktion, 63
 Kreisfrequenz, 52
 Kreuzspektraldichten, 331
 Kronecker–Lemma, 44, 365
 Kullback–Leibler–Information, 287, 291

L

Lag, 18
 Lag Number, 244
 Lag Window, 244
 Bartlett, 256
 Blackman–Tukey, 246
 Daniell, 258
 Hamming, 246
 Hanning, 246
 Modified Bartlett, 246
 Parzen(1), 246
 Parzen(2), 246
 Truncated Daniell, 246, 259
 Lag–Window–Schätzer, 244
 as. Bias, 250
 Lag–Operator, 111
 Landau–Symbole, 375
 Laurent–Entwicklung, 115
 Lemma von Slutsky, 378
 Levinson–Rekursion, 70, 355
 Lindeberg–Bedingung, 379
 lineare Filter, 93
 linearer Prozess, 142, 156, 159
 Lipschitz–Konstante, 160
 Log–Likelihood–Funktion, 198
 long memory, 323
 L_2 –stetig, 64

M

m –Abhängigkeit, 159, 169
 Maximum–Likelihood (ML)–Methode, 193
 Maß mit orthogonalen Werten (MOW), 78

Mean Square Error (*MSE*), 253
 Methode der kleinsten Quadrate, 9
 Moore–Penrose–Inverse, 185, 380
 Moving Average, 23

N

Normalgleichungen, 65, 370
 Nyquist–Frequenz, 52

O

Ordnungswahl in *ARMA*–Zeitreihen,
 279
 orthogonal, 368–371

P

Parallelogrammregel, 368
 Parseval–Gleichung, 371
 Parzen, *siehe* Lag Window
 Periodenlänge, 52
 Periodogramm, 229, 232
 asymptotische Verteilung, 239, 241
 Eigenschaften, 235
 geglättetes, 255
 integriertes, 262, 263
 asymptotische Normalität, 264
 Nicht–Konsistenz, 236
 zentriertes, 229
 Phasenspektrum, 341
 positiv (semi)definit, 48–49
 prewhitening, 253, 353
 principle of parsimony, 110
 Projektion, 65, 369

Q

QML–Methode, 313
 quadratischer Kohärenzkoeffizient, 337
 Quadratspektrum, 335, 339
 Quotientenstatistik, 269

R

Random Walk, 45
 Rechteckverteilung, 44
 reeller Fall, 114
 Regressoren, 110
 Reihe, *siehe* Zeitreihe
 Rendite, 6
 Riemann–Approximation, 84

S

Saisonbereinigung, 12

Satz des Pythagoras, 368
 Satz von
 Lebesgue, 365
 Bochner, 52
 Lindeberg (*CLT*), 379
 Prohorov, 378
 schwache Abhängigkeit, 153
 Schätzer
 bedingter *ML*–, 195
 kleinste–Quadrate–(*LS*), 195–196
 Lag–Window–, 244
 Yule–Walker–, 187–188
 Skalarprodukt, 366
 Sonnenflecken, 35
 span, 369
 Spektraldarstellung, 77
 Spektraldichte, 50
 bei *ARMA*–Reihen, 115
 Lag–Window–Schätzer, 243
 Schätzer
 Bias, 249
 parametrischer, 230
 Spektraldichte–Matrix, 331
 Spektralmatrix, 328
 Spektralmaß, 49
 Symmetrie, 53
 Spektralsatz, 75, 78
 multivariater, 332
 Spektralverteilungsfunktion, 263
 Spencer’s 15–Punkte Formel, 108
 stationär, 18
 Stetigkeitssatz
 für charakter. Funktionen, 377
 stochastisches Integral, 81–87, 332–336
 straff, 375
 Substitutionsregel, *siehe* stochastisches
 Integral

T

TAR–Modelle, 292
 Teilfolgenprinzip, 375
 Threshold, 292
 Transferfunktion, 91
 Transformationsformel, 89
 Trend, 9
 Truncation Point, 244

V

vec–Operator, 347

Verkehrsunfallzahlen, 7
 Verteilungskonvergenz, 376–378
 Volatilität, 299, 300
 Vorhersage, 65–68, 354–359

W

Wasserstände des Nils, 4
 Wechselkurse, 7
 weißes Rauschen, 22
 Gaußsches, 39
 multivariates, 327
 Wellenlänge, 52
 Wold-Zerlegung, 140
 Wolfer's Sunspot Series, 36

X

$X_{t-1} - X_t$ -Plot, 5, 41

Y

Yule-Walker
 –Gleichungen, 29, 129, 212
 –multivariate, 346
 –Schätzer, 31, 129, 185–188

Z

Zeitreihe, 1, 17
 L_2 -, 18

$ARCH(p)$ -, 299
 $ARMA(p, q)$ -, 29
 $GARCH(p, q)$ -, 298
 deterministische, 139
 gefilterte, 91
 $AR(\infty)$ -, 202
 $ARCH$ -, 30, 297
 $ARCH(1)$ -, 30
 $AR(1)$ -, 25
 $ARIMA$ -, 278
 $ARMA(p, q)$ -, 109
 $AR(p)$ -, 28, 110
 autoregressive (AR)-, 23–28, 110, 201
 Gaußsche, 38
 Moving Average (MA)-, 23, 106, 110
 multivariate (r-variate), 325
 rein nicht-deterministische, 139
 zyklische, 30, 52
 Zentraler Grenzwertsatz (CLT), 162,
 170, 171, 176, 187, 188, 196, 200,
 207, 239, 241, 251, 264, 270, 314,
 315, 351, 352, 379
 z -Transformation, 110
 matrixwertige, 344
 Zustandsgleichungen, 356
 Zustandsraummodell, 355

