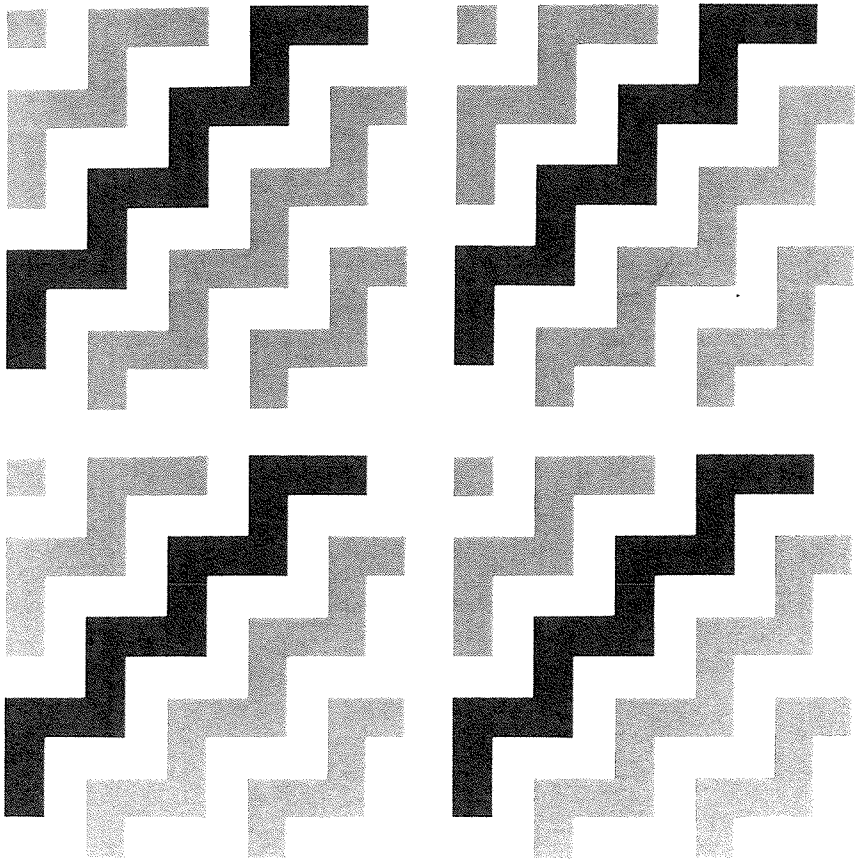
TIME SERIES MODELS by A.C. Harvey

# 時系列モデル入門

A·C·ハーベイ 著 国友直人／山本 拓 訳



東京大学出版会

時系列モデル入門

A·C·ハーベイ 著 国友直人 山本 拓 訳

東京大学出版会

定価 2884円
（本体 2800円）

ISBN 4-13-042022-4 C3033 ¥2800E

Translated from
TIME SERIES MODELS
©A.C. Harvey, 1981
Published 1981 by Philip Allan Publishers Limited, Oxford



# はしがき

本書では時系列データの分析と時系列モデルを扱う．学部の最終学年生および大学院初年生を対象とした統計学と計量経済学のコースで使われることを主な目的として本書は書かれている．その主眼は経済時系列の分析におかれているが，ここでふれられていることは，工学，地学をはじめ時系列データ解析を必要とする他の部門においてもまた教科書として適当であろう．

本書は私の前著 *Econometric Analysis of Time Series* の姉妹編とみなすことができる．いくつかの相互参照はあるものの，二つの本はそれぞれ本質的には自己完結的である．本書を通じて前著への言及は Harvey (1981 a) によって示される．

Harvey (1981 a) と同様，この本でも実用的価値のあるモデルおよびテクニックに主な関心が向けられる．すなわち，最大限の数学的厳密さをもって種々の定理を証明することよりも，むしろ種々のモデルによってどのようなことができるのか，およびそのモデルを応用する方法について理解することに力を注いでいる．読者は微分積分学および行列代数の初歩に多少の馴染みがあることを前提としている．しかしながら本書の多くの部分は，全く行列の知識をもたずに読むこともできる．統計的推論についての基礎知識もまた前提としている．それについては，例えば Silvey (1970) または Harvey (1981 a) の第 3，4，5 章などが推定および推論についての節への予備知識として十分であろう．本書で使用する統計的推論についての主要な結果の概略は序章にまとめられている．

時系列モデルの作成および予測の短期コースの教科書として用いる場合には，＊印のついた節をとばして 1，2，5，6 章を用いればよい．しかし，より全般的な時系列分析コースでは，3 章におけるスペクトラム分析についての内容を

含むべきであろう．もっとも個々の教員は，4章，7章における話題についてはやや選択的に扱うことを望むかもしれない．第4章を書き加えることにした理由は，状態空間モデル，およびそのより伝統的な時系列分析法との関係について最近急速に関心が高まってきたためである．残念ながら，現状では状態空間モデルについての多くの叙述は工学の著作のみにかぎられている．私は4章で示されている内容によって統計学や計量経済学になじんだ者に，状態空間モデルがより親しみやすいものになることを望んでいる．

本書では，時系列モデルとは確率変数間の変動の関係について，いかなる先験的理論にもよらず説明するよう構成されたモデルである，と定義している．一変量の分析においては，変数の変動は，それ自身の過去およびその時間に関する位置によってのみ説明される．多変量分析においても同じことが行なわれるが，そこでは変数の組に対して分析が行なわれる．実際にはいくつかの変数を同時にモデル化しようという決定は，ある種の先験的知識を含んでいる．しかしながら重要な点は，そこにおいてもモデルの形式について何らの先験的制約が課されないということである．

回帰モデルに関しての7章を含めたことは，ある意味では前段と矛盾するかもしれないが，そこでは，静学的モデルにおいてモデルに含まれる説明変数の形については所与と考えている．そして一般的な線型モデルのテキストではふれられていない時系列に固有な回帰分析におけるいくつかの側面に関心が払われている．

方程式は各節ごとに番号が打たれている．章の番号は，他章の式に言及する場合以外は除かれている．例題もそれぞれの節において番号をつけられており，方程式と同じように言及される．図と表はそれぞれの章において通し番号がつけられており，それが出てくる節とは独立である．それぞれの章の末尾にある付論では本文にあげられていない文献のリストが与えられているが，それはより進んだ研究のためである．

Harvey (1981 a) と同様に，ある節は＊印がつけてある．それらの節はより難しいか，あるいはよりつっこんだ奥深い内容を含んでいる．したがってそれらの節は省略しても連続性を失うことないだろう．しかし，大学院での時系列分析のコースにおいては，それらの節のほとんどは含まれるべきであろう．

本書では太字の字体は用いられていない．一般的な規則として，小文字の記号はスカラーまたはベクトルを表わし，一方，二，三の例外を除いては，大文字は行列を表わす．例外的に3章では大文字の記号が多くの場合にスカラーを表わすのに使われている．そこでは行列は3章 **3.7** 節の最初の部分のみに使われるだけである．対数 (log) は，他の説明がないかぎり自然対数を表わすものとして使われている．



本書の一部は，ロンドン経済大学 (LSE) における講義で用いられた案を基礎としている．そして特定のトピックの説明について，より明快に説明するように私を仕向けさせた学生たちにこの場を借りて感謝する．さらに，親切にも本書の種々の草稿段階においてコメントをくれた同僚や友人に感謝する．中でも特に Dick Baillie, Tom Cooley, James Davidson, Rob Engle, Katarina Juselius, Colin McKenzie および Bianca De Stavola に感謝しなければならない．もちろん，残されたいかなる誤りもひとえに筆者の責任である．最後に難しい草稿をきわめて効率的にタイプしてくれた Jill Duggan, Hazel Rice, Sue Kirkbride および Maggie Robertson に感謝する．

ロンドンにて
1981年2月
Andrew Harvey

# 訳者はしがき

現在では工学をはじめとして経済学や経営学などの分野でも，時系列データの統計的分析手法として時系列分析がしばしば用いられている．ところが使われている用語が耳慣れなかったり，これまでに経験したことのない考え方に出会ったりしてとまどう学生や研究者が，特に社会科学系の分野には少なくない．訳者は現在たまたま国立大学の経済学部に所属し，統計学や計量経済学を教えているため，このような人達から時系列分析の概念や用語に関してたずねられることがしばしばあり，適当なテキストを挙げるように求められることがある．こうした折には訳者はこれまで英文ではあるが特に本書を指定することにしていた．

もちろん，時系列分析の統計理論に関する英文の書物は必ずしも少なくない．大雑把にいってその大部分は二つのタイプに分類することができよう．第一は，ある程度の数学的素養を前提として時系列モデルの統計的推測に関する議論を展開するタイプである．その例としては Hannan (1970)，Anderson (1971)，Fuller (1976) などをあげることができるが，これらの書物は統計学を専門とする研究者の要求をある程度満足させるであろう．第二のタイプとしては，時系列データが与えられたときに統計モデルをいかにして作り，将来値を予測するかといった問題を扱う Box=Jenkins (1970，第 1 版) とその解説書の一群がある．そのアプローチは「ボックス=ジェンキンズ法」と呼ばれ，時系列分析と同一視されることもしばしばある．このように，ある意味で現代の統計学の方向を象徴している二つのアプローチが時系列分析という名の下に存在しているために，この分野をこれから勉強しようとする学生や研究者のテキストとしてはこれらの書物はいずれも問題をはらんでいるように思われる．それに比較してあえていうならば，本書は二つのタイプの中間に位置づけることができ，時系列分析への二つのアプローチについてそれぞれほどよくバランスのとれた解説を行なっているように見うけられる．



さて本書の著者であるアンドリュウ・ハーベイ氏は伝統的に統計学が盛んなロンドン経済大学 (LSE) で教授として教鞭を取っている．同教授はこれまでに時系列の統計分析に関する数多くの論文を発表しており，時系列モデルについての同教授の鋭い洞察は本書のいたる所に見ることができる．そして本書の特長の一つとして，特に時系列モデルについての話題は新鮮かつ豊富で，しかもその数学的取り扱いは比較的平易であることを挙げることができる．このことからも，これから時系列データの分析あるいは時系列理論を勉強しようとする統計学に関連した諸分野，例えば工学・経済学・経営学などの学部学生や修士コースの大学院生にとって，現在入手可能な最良のテキストとして本書を奨めることができよう．実際，訳者自身も時系列モデルの話題の扱い方などについて，本書から少なからず知的刺激を受けたことをつけ加えねばならない．

私事になるが，本書の存在を知ったのは，訳者の一人がノースウェスタン大学経済学部に在職の折に同僚の Marc Nerlove 氏 (現在ペンシルバニア大学) とデンマーク人の Lars Muus 氏 (現在コーネル大学) の示唆によるものであった．両氏にはこの場をかりて感謝したい．訳出にあたっては著者の了解の下で原文に散見された誤り・誤植等を訂正した．最後に東京大学出版会の大瀬令子さんには翻訳の機会を与えていただき感謝する次第である．

なお，授業等で本書を教科書として使用し，各章末の練習問題を学生に宿題として課すインストラクターのために訳者はその解答集を作成した．希望者は直接訳者に請求されたい．

1985 年 5 月

国 友 直 人
山 本　拓

# 目　　次

はしがき

訳者はしがき

**1章　序　　論**

1.1　時系列データの分析と時系列モデル …… 1

時系列モデルの本質(2)　トレンド(4)　多変量時系列(8)　本書の概略(9)

1.2　推　　定 …… 10

最尤法(10)　線型回帰(11)　予測誤差の分解(13)　多変量モデルの尤度関数の分解(17)　尤度関数の最大化(17)　二段階推定量(18)

1.3　検　　定 …… 18

古典的検定の方法(19)　非入れ子型仮説とモデルの診断(20)

練習問題(21)

**2章　定常確率過程と時間領域におけるその性質**

2.1　基本概念 …… 23

定常性(24)　自己共分散関数と自己相関関数(26)　コレログラム(27)　ラグ作用素(28)　自己回帰移動平均過程(29)

2.2　自己回帰過程 …… 30



1次自己回帰モデルにおける定常性(30)　2次自己回帰モデルにおける定常性(31)　$p$次自己回帰モデルにおける定常性(33)　自己共分散関数と自己相関関数(34)

2.3　移動平均過程 …… 36

自己共分散関数と自己相関関数(37)　反転可能性(37)

2.4　混合過程 …… 38

定常性と反転可能性(38)　自己共分散関数と自己相関関数(40)　共通因子(42)

2.5　集計問題と欠足観測値問題 …… 42

和の定理(43)　時間集計と欠足観測値(44)

2.6　多変量時系列 …… 45

定常性(46)　相互相関関数(47)

2.7　ベクトル値自己回帰移動平均過程 …… 48

自己共分散関数と自己相関関数(48)　多変量自己回帰移動平均過程における定常性と反転可能性(50)　識別性(51)　多変量モデルと一変量モデル(51)

練習問題(52)

**3章　周波数領域**

3.1　はじめに …… 55

スペクトラム(56)

3.2　周期的トレンド …… 58

周期的関数(59)　フーリエ解析(60)　循環的トレンド・モデル(64)

3.3　確率過程のスペクトラム表現 …… 66

複素スペクトラム表現(69)　ウォルドの分解定理(70)

3.4　周波数領域における自己回帰移動平均過程の性質 …… 71

伝達関数の導出(72)　自己回帰過程(73)　混合過程と随伴多項式(75)

3.5 線型フィルターの性質 …… 76
ゲインと位相(77) 周波数応答関数(79) 循環的変動の平滑化および除去(79) 季節調整(82) 見せかけの周期変動(83)
3.6 スペクトラムの推定 …… 85
標本スペクトラム密度の性質(86) 標本スペクトラムの平均化(87) 時間領域における加重和(89) 高速フーリエ変換(91) 自己回帰スペクトラムの推定(92)
3.7 多変量スペクトラム解析 …… 93
ゲイン，位相およびコヒーレンス(93) 推定(96)
付論 A：三角関数の定義式(97) B：直交関係(99) C：フーリエ変換(100)
練習問題(101)

**4章 状態空間モデルとカルマン・フィルター**

4.1 状態空間 …… 103
カルマン・フィルター(104) ARMA モデルの状態空間表現(105)
4.2 最小平均二乗推定 …… 106
事前情報と標本情報の結合(108)
4.3 更新方程式と予測方程式 …… 109
予測(109) 更新(110) カルマン・フィルターの一般形(112)
4.4 最尤推定 …… 113
定常モデル(114) 非定常モデル(115)
4.5 予測，平滑化および定常状態 …… 116
多期間予測(116) 平滑化(117) 定常解(119)
付論：行列の逆転公式(120)
練習問題(121)

**5章 自己回帰移動平均モデルの推定**

5.1 はじめに …… 123
5.2 自己回帰モデル …… 124



厳密最尤推定(125) 回帰法(125) ユール=ウォーカー方程式(126)
5.3 移動平均過程および混合過程 …… 127
MA(1)モデルの条件付二乗和(127) 一般の場合の条件付二乗和(129) 厳密最尤推定(130) 初期推定値(131) 定常性と反転可能性(133)
5.4 漸近的特性 …… 134
自己回帰過程(134) MA および混合過程(135)
5.5 仮説検定と信頼区間 …… 136
ワルド検定および尤度比検定(137) ラグランジュ乗数検定(138) 信頼区間(139)
5.6 小標本特性 …… 139
自己回帰過程(139) 移動平均過程と混合過程(140)
5.7 多変量モデル …… 144
ベクトル AR(1)過程の推定(145) 最尤推定量の漸近的分布と関連する検定統計量(146) ベクトル MA(1)過程(147) 見かけ上無関係なベクトル値 ARMA 過程(148)
練習問題(149)

**6章 時系列モデルの作成手順と予測**

6.1 はじめに …… 151
6.2 独立性検定 …… 152
フォン・ノイマン比(153) 風呂敷検定統計量(154) 累積ピリオドグラム(156)
6.3 モデル選択 …… 157
識別(159) 診断(161) 適合度(163)
6.4 予測 …… 163
最適予測とその平均二乗誤差(165) 有限標本予測と反転可能性(166) 母数推定値を用いる予測(167)
6.5 定常性 …… 169

自己回帰和分移動平均過程(171)　予測(172)　確率的トレンド項と誤差項の加法モデル(175)

6.6 季　節　性 …… 176

季節自己回帰移動平均過程(177)　非確率的季節要素モデル(179)　可変的季節パターン(179)　季節調整(181)

6.7 時系列モデルの一般的クラス …… 181

トレンド要素，季節要素および誤差要素(183)　過剰階差問題(185)　モデル選択(187)

6.8 多変量モデルの作成 …… 189

定常性(190)　識別，推定および診断(190)　モデルの性質(193)　予測(193)

練習問題(194)

**7章　時系列回帰分析**

7.1 逐次最小二乗法 …… 197

逐次残差(198)　割引逐次最小二乗法(199)

7.2 系列相関のある攪乱項 …… 200

厳密最尤推定法(201)　累積ピリオドグラム(203)

7.3 時変係数回帰モデル …… 205

確率的係数モデル(206)　定常性への再帰モデル(207)　酔歩係数回帰モデル(209)　過去の観測値の割引き(212)

7.4 周波数領域における回帰分析 …… 212

周波数領域への変換(213)　系列相関(216)

7.5 系列相関のノン・パラメトリックな扱い …… 218

練習問題(220)

文　　献 …… 223

索　　引 …… 228



# 1章 序　論

## 1.1 時系列データの分析と時系列モデル

典型的な時系列データは，等間隔にとらえた時間に対応して観測された変数 $y$ の観測値の集まりから成りたっている．たとえば，経済変数は一般にストックまたはフローに分類される．貨幣供給はストックの例であり，投資やGNPはフローの例である．ストックとフローの区別は集計問題や欠足観測値の問題を扱うときには重要であるけれども，多くの目的にとって問題とはならない．本書では多くの場合それらがストックであるかフローであるかということを無視して，$T$ 個の観測値の系列は $y_1, y_2, \cdots, y_T$ として表わす．

時系列の研究には，時系列データの分析と時系列モデルの分析という二つの側面がある．時系列データの分析の目的は系列の特徴をまとめ，その顕著な特徴を性格づけることにある．これは時間領域または周波数領域のいずれにおいても行なわれる．時間領域においては，異時点間における観測値の関係に主な関心が注がれる．一方，周波数領域において研究されるのは周期的変動である．時系列分析のこの二つの側面は，競合的であるというよりも補完的であるといえよう．同一の情報が異なった方法によって処理されるのであり，それによって時系列の性質についての異なった洞察が与えられることが期待されよう．

時系列データをモデル分析する主な理由は，それによって将来の値についての予測を可能にすることである．時系列モデルの，たとえば計量経済モデルと対峙させる顕著な特徴は，それが $y_t$ を他の変数と関係づける何らの試みもなされないことである．すなわち $y_t$ の動きは，それ自身の過去に関してのみ，

またはそれの時間的関係に関してのみ「説明」される．そして予測は外挿によって与えられるといったところであろう．

### 時系列モデルの本質

図 1.1 は一定の水準 $\mu$ の周辺に変動する一連の観測値の系列を示している．もし観測値が互いに独立であるならば，時系列の次の観測値 $y_{T+1}$ の予測は単に $\mu$ であろう．また，もしそれが未知であるならば，標本平均のような $\mu$ の適当な推定値を用いればよい．しかしながらこの場合，観測値は明らかに独立ではなく，それぞれ遠く離れたものどうしよりも，隣接したものどうしの方がより近い値を持つ傾向を示している．この構造は系列相関(serial correlation)として知られているが，時系列観測値の特徴の典型的なものであり，系列相関の形を考慮することによって将来の観測値についてのよりよい予測値を得ることができる．すなわち図 1.1 におけるような連続する観測値間の関係が与えられたときには，予測値を単純に $\mu$ とするよりも，$y_T$ と $\mu$ の間のどこかにあるとする方がもっともらしく思われよう．

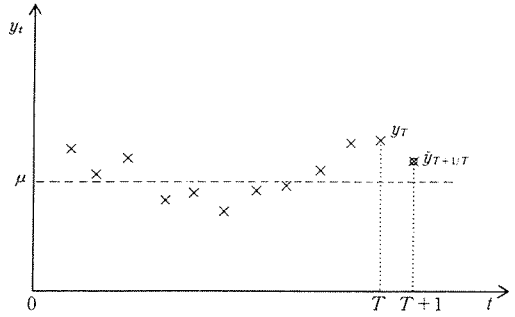



**図 1.1** 系列相関を持つ時系列

予測への統計的アプローチは時系列モデルの作成に基礎をおいている．ここで時系列モデルとは，問題となっている観測値を生み出すことが可能であろうとみなされるメカニズムを定義するものである．そのようなモデルは多くの場合に確率的(stochastic)なものである．もしモデルから同じ期間についての観測値のいくつかの組が生成されるとすれば，異なった観測値の組は同一の確率的法則に従うであろう．

一階の自己回帰モデル



$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t \tag{1.1}$$

は，確率過程(stochastic process)の簡単な例である．不確実性は変数 $\epsilon_t$ から生じる．これは平均がゼロ，分散が $\sigma^2$ である互いに独立な撹乱項であるので異時点間における相関がゼロである．モデルのその他の性質は，母数 $\mu$ および $\phi$ によって決定される．もし $|\phi| < 1$ であれば，観測値はその確率過程の平均である $\mu$ のまわりを変動する．

前述の(1.1)における母数(parameter)は，最小二乗法(ordinary least squares：OLS)回帰によって推定することができる．推定値 $\tilde{\mu}$ および $\tilde{\phi}$ が与えられれば，この系列の次の観測値は次式により予測することができる．

$$\tilde{y}_{T+1|T} = \tilde{\mu} + \tilde{\phi}(y_T - \tilde{\mu}) \tag{1.2}$$

推定値 $\tilde{\phi}$ が 1 に近ければ近いほど，より多くの荷重が $y_T$ に与えられる．これは図 1.1 の観測値について与えられた直観的議論と整合的である．

上記(1.1)において $\phi$ が $-1$ から 1 の間に含まれるということは，その確率過程が定常(stationary)であることを意味する．観測値の時系列が定常過程から生成されている場合には，それらはある一定の水準のまわりに変動し，時間とともに散らばり方が増大したり減少したりする傾向はない．以上の性質は定常時系列の唯一の特徴ではないが，最も顕著なものであり，図 1.1 における系列を簡単にながめただけで，それがこのような特徴を示しているということがわかる．

より離れた過去の値 $y_{T-2}$，$y_{T-3}$ などを(1.1)に追加することもできる．そのようにすれば，より複雑な時間的依存関係の形をモデル化することができる．たしかに十分に高い次数の遅れを導入することによって，ほとんどどのような時系列の特徴も復元することができる．ただし，このような方法で時系列をモデル化することの不利な点としては，多数の遅れの変数値が必要な場合には，多数の母数を推定しなければならないことである．この問題に対する一つの解は，モデルのクラスを拡張して $\epsilon_t$ についての遅れ(ラグ)の値を含むようにすることである．観測される変数のラグ変数と撹乱項のラグ変数の双方を含むモデルは，自己回帰移動平均(autoregressive-moving average：ARMA)として知られている．この確率過程は動学的モデル作りにおいて中心的な役割を演ずる．なぜならば，このモデルは定常時系列の倹約的な表現(parsimonious representation)をもたらすのである．いい換えると，比較的少数の母数から成るモデルを作ることができるのである．

モデル(1.1)を修正して$\epsilon_t$についての一期のラグ変数を導入すると，

(1.3) $$y_t-\mu=\phi(y_{t-1}-\mu)+\epsilon_t+\theta\epsilon_{t-1}$$

となる．ここで$\theta$は移動平均母数である．このモデルは次数(1,1)のARMA過程であり，これを簡単に$y_t\sim$ARMA(1,1)と表わす．

実際の時系列が定常であることは非常に少ないけれども，定常性の概念は時系列分析にとって基本的なものである．定常ARMA過程はその上に，より一般的なモデルを作る基礎となる形である．さらにこのモデルの利用は，純粋な時系列モデルに限られるものではない．例えば時系列データを用いる回帰モデルにおける撹乱項を定常な確率過程とみなすことも可能であろう．この考え方はとくに工学や計量経済学において，人間の時間的行動をモデル化する場合に広く用いられている．

### トレンド

多くの時系列はある一定の水準のまわりを変動するわけではなく，傾向的に上昇または下降の動きを示す．図1.2(a)に与えられている観測値の場合を考えよう．観測値は上昇勾配を描く直線のまわりに確率的に散らばっているとみなすことができる．ゆえに，適切なモデルは以下のように書けるであろう．

(1.4) $$y_t=\alpha+\beta t+\epsilon_t,\qquad t=1,\cdots,T$$

ここで$\alpha$と$\beta$は母数であり，$\epsilon_t$は(1.1)で定義されたような「ホワイト・ノイズ」(white noise)である．

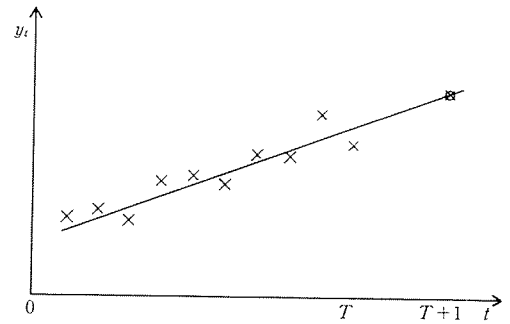
上記の(1.4)は古典的な線型回帰モデルであるから，その母数は最小二乗法によって推定することができる．また$y_t$の将来値の予測は当てはめられた式の非確率的部分を外挿することによって行なわれる．すなわち，

(1.5) $$\hat{y}_{T+l/T}=a+b(T+l)=a+bT+bl,\qquad l=1,2,\cdots.$$

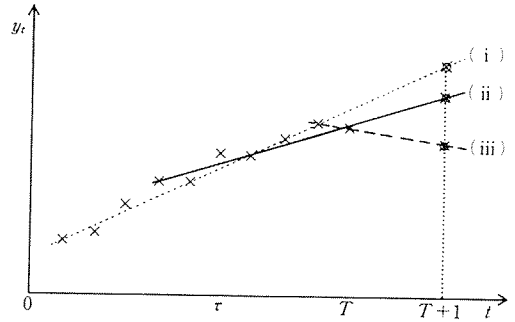
ここで$a$と$b$は$\alpha$と$\beta$の最小二乗推定量を表わす．さらに撹乱項がARMA$(p,q)$過程に従うとすることによって，より一般的なモデルを作ることができる．もしそのようなモデルが作られるとすると，上記(1.5)による予測は系列相関の形を考慮に入れるように修正されなければならないであろう．

モデル(1.4)を予測目的のために使うことの不利な点は，そのトレンドが大域的(global)なものと仮定されているということである．すなわちすべての時点において，トレンド項の母数はずっと一定のまま成立すると想定されている．したがって図1.2(b)に表わされるような状況，すなわち勾配が$t=\tau$以降で






(a) 大域的トレンドモデルによる予測




(b) (i)大域的トレンドによる予測，(ii)局所的トレンドによる予測，(iii)極端な局所的トレンドによる予測

**図1.2** 大域的線型トレンドと局所的線型トレンド

下落しているような場合は，(1.4)では適切に扱うことができないのである．この点については，撹乱項がホワイト・ノイズであるか，またはより一般的な定常ARMA過程が想定されているかという問題は関係がない．

図1.2(b)の時系列に対するより好ましい予測方法は，データにおける変化に対して母数が適応していくのを許すことであろう．もしより多くの加重(ウエイト)が最も最近の観測値に与えられるならば，予測は局所的(local)トレンドの推定値に基づいて行なうことができる．図1.2(b)における線は二つの接近法を対比させてある．この考え方を極端にすると，最後の二つの観測値のみから直線を射影することを意味するであろう．すなわち，

(1.6) $$\hat{y}_{T+l/T}=y_T+(y_T-y_{T-1})\,l,\qquad l=1,2,\cdots.$$

さてホルト=ウィンターズ(Holt=Winters)の方法とは，局所的な直線的トレンドからの予測を作るための一般的な手続きである．その予測関数は以下のように与えられる．

(1.7) $$\hat{y}_{T+l/T} = \hat{\alpha}_T + \hat{\beta}_T l, \qquad l = 1, 2, \cdots$$

ここで $\hat{\alpha}_T$ と $\hat{\beta}_T$ は次のような漸化式から計算される．

(1.8 a) $$\hat{\alpha}_t = \lambda_0 y_t + (1-\lambda_0)(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1})$$

(1.8 b) $$\hat{\beta}_t = \lambda_1(\hat{\alpha}_t - \hat{\alpha}_{t-1}) + (1-\lambda_1)\hat{\beta}_{t-1}, \qquad t = 3, \cdots, T$$

ここで初期値は $\hat{\alpha}_2 = y_2$，$\hat{\beta}_2 = y_2 - y_1$ であり，平滑化定数(smoothing constants) $\lambda_0$，$\lambda_1$ は $0 < \lambda_0$，$\lambda_1 \leq 1$ を満たす．ここで $\lambda_0 = \lambda_1 = 1$ とおくと，この方法によって(1.6)に基づく予測がもたらされる．一方，それらの母数がゼロに近づいていくと，予測は大域的トレンドモデルによって行なわれたものに近づいていく．ここで $\hat{\alpha}_T$ は(1.5)における $a + bT$ に対応していることに留意されたい．

上記の $\lambda_0$ と $\lambda_1$ の値は，利用者の先験的(a priori)情報のもとに決定することができる．これらの値が1に近ければ近いほどシステムはより早く変化方向に適応していき，短期の予測の正確性に主な重点が置かれるのである．

このようなホルト=ウィンターズ(Holt=Winters)のシステムは本質的には恣意的(*ad hoc*)なものであり，果してこのような形式の予測を生み出すようなモデルは実際に存在するのかどうかというごく自然な疑問が生じる．$\lambda_0 = \lambda_1 = 1$ のような極端な場合には，この答は非常にはっきりしている．まず観測値の一階の階差が次のように定義されるものとしよう．

(1.9) $$\Delta y_t = y_t - y_{t-1}.$$

もう一回，階差をとることにより次式を得る．

(1.10) $$\Delta^2 y_t = y_t - y_{t-1} - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2}$$

さらに想定されたモデルが以下のようなものであるとしよう．

(1.11) $$\Delta^2 y_t = \epsilon_t, \qquad t = 3, \cdots, T$$

ここで $\epsilon_t$ はホワイト・ノイズである．すると $y_{T+1}$ の適当な予測は以下のように与えられる．

(1.12) $$\hat{y}_{T+1/T} = 2y_T - y_{T-1}.$$

この式は(1.6)において $l = 1$ としたときの予測と同一である．$l > 1$ の予測は，次のような差分方程式から得ることができる．

(1.13) $$\hat{y}_{T+l/T} = 2\hat{y}_{T+l-1/T} - \hat{y}_{T+l-2/T}, \qquad l = 1, 2, \cdots$$



ここで $\hat{y}_{T/T} = y_T$ および $\hat{y}_{T-1/T} = y_{T-1}$ である．モデル(1.11)はどのような $l$ の値についても(1.6)を解としてもたらす．

一般的には，観測値の二階階差をとったモデルがMA(2)過程に従うと想定される場合には，それはホルト=ウィンターズ法によって作られた予測に一致する．もし $\theta_1$ と $\theta_2$ が次式における移動平均項の母数であるならば，

(1.14) $$\Delta^2 y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

となる．すなわち(1.8)における $\lambda_0 = 1 - \theta_2$ および $\lambda_1 = \dfrac{1 + \theta_1 + \theta_2}{1 - \theta_2}$ に対応している．このことから，(1.14)をデータに対して当てはめて $\lambda_0$ と $\lambda_1$ を推定するという可能性が示唆される．

上記(1.14)における確率過程は，自己回帰和分移動平均(autoregressive-integrated-moving average：ARIMA)過程と呼ばれるモデルの種類の一つの特殊例となっている．ARIMA$(p, d, q)$過程においては，観測値の $d$ 階の階差がARMA$(p, q)$過程としてモデル化されている．ゆえに(1.14)はARIMA$(0, 2, 2)$である．そのようなモデルは局所的なトレンドをとらえることができるので，予測に際して非常に魅力的となる．

ARIMAモデルに基づいた時系列予測への接近法はボックス=ジェンキンズ(Box=Jenkins, 1976)によって発展させられた．彼等の手法の第一段階は近似的に定常な系列を得ることである．これは通常は観測値に対して対数をとるなどの変換を適用すること及び一階または二階の階差をとることによって行なわれる．しかる後に，結果としてもたらされた系列が分析され，その特徴に基づいて適当なARMAモデルが選ばれる．このモデルが次にデータに当てはめられ，残差を検証して，もしモデルが満足すべきものであると診断されれば，それが予測に用いられる．ここでもしこの残差の「診断」(diagonostic checking)がモデルの定式化の重大な不適切さを指摘すれば他の定式化を開発し，そして以上の過程がくり返される．

ボックス=ジェンキンズによる以上の方法は，いくつかの批判にさらされている．その中の一つは，この方法で適当なモデルを選ぶにはかなりの熟練を要するということである．それにひきかえホルト=ウィンターズのような手続きは，準自動的なやり方で適用することができる．しかしながらこの批判が受け入れられたとしても，ARIMA過程を用いて時系列を研究することの有利さは否定しがたい．第一に，多くの準自動的な手続きは，もとの時系列についてある特定の確率過程を仮定していることを示すことができる．そしてこのことを

認識すると，前述の方法の長所と短所がより明らかに描き出されることになろう．第二に，ARIMA モデルからの予測の平均平方誤差は明示的に評価することができる．このことは予測の信頼性を評価するに当っての一つの科学的な基準を提供することになる．

### 多変量時系列

これまでの議論は一変量(univariate)時系列の分析とその時系列モデルに限られていた．いくつかの系列を同時に考慮するときには，$y_t$ はもはや一つの観測値ではなくベクトルとみなされる．そして $y_1, \cdots, y_t$ の系列は多変量(multivariate)時系列と呼ばれる．

いくつかの系列を同時にモデル分析するのは，系列がお互いに関係づけられていることを示唆する先験的(a priori)情報が存在することを意味している．しかしながら，一変量時系列モデルと同様に具体的な先験的な知識というものは必要とされるわけではない．もし $y_t$ が確率過程とみなされるのであれば，観測値ベクトルのそれぞれの要素は，それ自身の過去の値とシステム内における他のすべての変数の過去の値に依存すると想定される．もし一次のラグのみが適当であるならば，(1.1)の一般化を得ることができる．これは次のように書くことができよう．

$$y_t = \Phi y_{t-1} + \epsilon_t. \tag{1.15}$$

もし $y_t$ が $N\times 1$ ベクトルであるならば，ランダムな攪乱項 $\epsilon_t$ もまた $N\times 1$ ベクトルであり，大文字の $\Phi$ は $N\times N$ の母数行列である．攪乱項は時間に関しては無相関であるが，同時点間では相関があるかもしれない．ゆえに，$E(\epsilon_t \epsilon_t') = \Omega$ は $N\times N$ 行列として定義される．一変量の場合と同様に，(1.15)の型のモデルは通常は定常な変量のみに当てはめられる．そこで(1.15)を書くにあたって，$y_t$ は一階または二階の階差をとり，かつ対数をとったあとの変数の組を表わすものと理解される．またゼロでない平均も $y_t$ の中に組み込むこともできよう．

多変量時系列モデルと計量経済学で用いられる動学的な同時方程式モデル体系との関係は，Harvey(1981 a)の最終章で吟味されている．なお(1.15)の形のようなモデルは，計量経済学における予測においてしばしば用いられるけれども，そこでは経済理論が $\Phi$ の要素について先験的(a priori)な制約をおく役割を演じている．



### 本書の概略

以下2章では，時間領域における確率過程の分析にとって必要な基本概念を述べる．ARMAモデルのクラスの背後にある考え方を詳しく説明し，そしてそのクラスの中の種々のメンバーの時間領域での性質を調べる．最終節において一変量時系列の概念を多変量過程に拡張する．

3章は，周波数領域における分析にかかわっている．これはスペクトラム分析(spectral analysis)として知られているものである．いくつかの ARMA モデルの周波数領域の性質を調べ，そして対応したモデルの時間領域における性質と対比する．しかる後に，時系列にほどこされる種々の操作の影響について，スペクトラム分析がいかに重要な洞察を与えるかということを示す．季節調整は一つの例である．

4章は，状態空間モデルとカルマン・フィルターについての序論である．状態空間モデルは主に制御工学との関係において発展し，比較的近年になってその重要性がより広く知られるようになってきた．状態空間モデルは観測不能な要素(unobservable components)の取扱い，および時間的に変動する母数を含むモデルの取扱いの方法を提供するのみでなく，ARMA モデルの厳密な最尤推定についての自然な接近法の基礎を提供する．

推定についての話題は5章で扱われる．5章の主な論点は，条件付二乗和関数と呼ばれるものの最小化を含むような最尤推定の近似についてである．検定の手続きもまた5章で展開される．

そして6章では，いかにして検定の手続きがモデル選択の一般的戦略に用いることができるか，が示される．6章はまた予測を作る手続き及びそれに関連した平均二乗誤差を計算するという問題を扱っている．

最終章は時系列回帰分析のいくつかの側面を取り扱う．これは，時系列モデルの本質的な特徴はモデルが具体的行動を記述する先験的情報を含まないことであるという前言から考えると，場違いのように見えるかもしれない．しかしながら回帰モデルの非確率的部分は，ここの取扱いにおいては単に二次的な興味でしかない．その定式化は与えられるものと考え，攪乱項および時間と共に変化し得る母数のモデル分析の方法にもっぱら関心が注がれることになる．

## 1.2 推　　定

時系列モデルにおいて母数の推定に用いられる多くの方法は，最尤法の原理に基づいている．最尤法を採用することの便宜は，その結果として大標本において有効推定量が得られることにある．一方，それを算出するには，通常何らかの逐次的方法が必要となる．いくつかの数値的最適化の方法についての序論的解説は，Harvey (1981a) の 4 章に見い出されるであろう．ただしガウス=ニュートン (Gauss=Newton) 法と呼ばれる一つの方法については，以下でもやや詳しく述べよう．ここでは幾分は教育的見地からガウス=ニュートン法を強調する．というのはこの方法によって非線型の枠組における推定法および検定についての多少の理論的洞察が得られるからである．

以下で示される内容は，Harvey (1981 a) の 2 章から 4 章までの簡単なまとめである．まず最尤推定量の性質を示し，しかる後に線型回帰分析を説明する．ここでは，読者は回帰分析にはすでに親しみがあると想定している．回帰分析をここで取り上げる主な理由は，記号を紹介するとともに，最小二乗法と最尤法の関係を確立するためである．次に予測誤差の分解を述べるが，これは尤度関数を扱いやすい形式に解きほぐす方法を提供するという意味において，時系列分析における基本的な結果である．そして一変量モデルに関しては，尤度関数を最大化する問題は，適当な近似によって二乗和関数を最小にする問題に置き換えることができるということが示される．

### 最　尤　法

観測値 $y_1, \cdots, y_T$ の同時密度関数が，ベクトル $\psi = (\psi_1, \cdots, \psi_n)'$ なる 1 組の $n$ 個の未知の母数に依存すると想定しよう．そして $L(y_1, \cdots, y_T; \psi)$ と表わそう．いったん標本が抽出されたなら $L(y_1, \cdots, y_T; \psi)$ は尤度関数と再解釈され，それは与えられた標本における異なった $\psi$ のもっともらしさを示す．したがって尤度は $\psi$ の関数である．そこで最尤推定量 $\tilde{\psi}$ は，$L(\psi)$ を最大化する $\psi$ の値と定義する．

最尤推定量を見つけ出す通常の方法は，対数尤度関数 $\log L$ を $n$ 個の母数 $\psi_1, \psi_2, \cdots, \psi_n$ のそれぞれについて微分することである．これらの微係数をゼロとおくことにより，以下のような尤度方程式 (likelihood equations) を得る[1]．



$$\frac{\partial \log L}{\partial \psi} = 0. \tag{2.1}$$

一般的には尤度方程式は非線型である．そこで最尤推定値は何らかの逐次方法によって見つけ出さなければならない．

古典的な最尤推定理論は，独立で同一分布を持つ観測値について発展してきた．時系列モデルは，その定義によって独立でない観測値にかかわるものである．しかしながら，標準的な結果が時系列モデルにも適用でき，独立な観測値に対して持っていたのと全く同様の一連の結果が得られる．ここで主要な点は，いくつかの正則条件の下に $\tilde{\psi}$ が一致性を持つこと，および $\sqrt{T}(\tilde{\psi} - \psi)$ が平均ゼロで共分散行列が漸近的に情報行列の逆行列となるような多変量正規分布に収束することである．ここで漸近的情報行列は以下のように定義されるものである．

$$IA(\psi) = \operatorname{plim} T^{-1}\left(-\frac{\partial^2 \log L}{\partial \psi \partial \psi'}\right). \tag{2.2}$$

この結果を表す別の方法は $\tilde{\psi}$ が漸近的に正規分布に従い，平均ゼロで共分散行列として以下のようなものを持つということである．

$$\operatorname{Avar}(\tilde{\psi}) = T^{-1} IA^{-1}(\psi). \tag{2.3}$$

このことは簡単に次のように表現される．

$$\tilde{\psi} \sim AN(\psi, T^{-1} IA^{-1}(\psi)). \tag{2.4}$$

実際には $\operatorname{Avar}(\tilde{\psi})$ は標本から推定しなければならない．漸近的に有効な検定統計量を作り上げるためには，以下のような条件を満たす $\operatorname{avar}(\tilde{\psi})$ の推定値を作らなければならない．

$$\operatorname{plim} T \cdot \operatorname{avar}(\tilde{\psi}) = IA^{-1}(\psi). \tag{2.5}$$

もしこの条件が満たされれば，$\operatorname{avar}(\tilde{\psi})$ の対角要素の平方根は漸近的標準誤差 (asymptotic standard error) となる．

### 線型回帰

古典的な線型回帰モデルは，以下のような形式で表わされる．

---

1) もし $f(x)$ が $n$ 変数ベクトル $x = (x_1, x_2, \cdots, x_n)'$ の関数であるなら，$\partial f/\partial x$ は第 $i$ 要素が $\partial f/\partial x_i$ であるような $n \times 1$ ベクトルを表す．対応する行ベクトル $(\partial f/\partial x)'$ は，$\partial f/\partial x'$ と書かれよう．二次微係数からなる $\partial^2 f(x)/\partial x \partial x'$ は $\partial^2 f(x)/\partial x_i \partial x_j$ をその第 $ij$ 要素として持つ $n \times n$ 行列である．

(2.6) $$y_t = x_t'\beta + \epsilon_t, \qquad t = 1, \cdots, T,$$

ここで$x_t$は，くり返される標本において固定したものとして扱うことができる変数の$k\times 1$の観測値ベクトルであり，$\beta$は$k\times 1$の母数ベクトルであり，そして$\epsilon_t$は確率的な撹乱項である．

もし観測値が正規分布しているならば，すなわち，もし$\epsilon_t \sim NID\ (0, \sigma^2)$であるならば，対数尤度関数は次のように与えられる．

(2.7) $$\log L\ (\beta, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t - x_t'\beta)^2.$$

尤度方程式は次のようになる．

(2.8) $$\frac{\partial \log L}{\partial \beta} = \sigma^{-2}\sum_{t=1}^{T} x_t\,(y_t - x_t'\beta) = 0,$$

(2.9) $$\frac{\partial \log L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(y_t - x_t'\beta)^2 = 0.$$

ここで(2.8)と(2.9)を解くことにより，$\beta$と$\sigma^2$の最尤推定量が次のように求められる．

(2.10) $$\tilde{\beta} = \left(\sum_{t=1}^{T} x_t x_t'\right)^{-1}\sum_{t=1}^{T} x_t y_t,$$

(2.11) $$\tilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(y_t - x_t'\tilde{\beta})^2.$$

このように尤度方程式の直接的な解を求めることが可能となる．もっともこのようなことが可能なのは**1.1**節で述べたように，一般的であるというよりはむしろ例外である．なお(2.10)における最尤推定量は$\beta$の最小二乗(ordinary least squares：OLS)推定量と同一である．そしてそれは$\beta$の最小分散不偏推定量(minimum variance unbiased estimator：MVUE)であることを示すことができる．もし正規性の仮定が落されても，最小二乗推定量は依然として線型推定量のクラスの中で最小の分散を持つことになる．それは最良線型不偏推定量(best linear unbiased estimator：BLUE)と呼ばれている．

モデル(2.6)は以下のように表わすこともできる．

(2.12) $$y = X\beta + \epsilon,$$

ここで$y$は$T\times 1$ベクトルで$y = (y_1, \cdots, y_T)$，$X$は$T\times k$行列で$X = (x_1, \cdots, x_T)'$，そして$\epsilon$は$T\times 1$ベクトルで$\epsilon = (\epsilon_1, \cdots, \epsilon_T)'$である．最小二乗推定量(2.10)は，次のように表わすことができる．



(2.13) $$b = \tilde{\beta} = (X'X)^{-1}X'y.$$

もし撹乱項の平均がゼロであっても，それが系列相関を持っていたり，または異なった分散を持っているときには，モデルは一般化線型回帰モデルと呼ばれるものとなる．そのときは(2.12)におけるベクトル$\epsilon$は，$E(u) = 0$でかつ$E(uu') = \sigma^2 V$であるようなベクトル$u$に置き換えられることになる．ここで$\sigma^2$はスカラーであり，$V$は正値定符号行列である．もし$V = I$であるならば古典的モデルが得られ，そして(2.13)は$\beta$の最良線型不偏推定量となる．より一般的には，$\beta$の最良線型不偏推定量は次のような一般化最小二乗(generalised least squares：GLS)推定量である．

(2.14) $$\tilde{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y.$$

ふたたび，もし撹乱項が正規分布しているならば$\tilde{\beta}$は$\beta$の最尤推定量となる．

### 予測誤差の分解

予測誤差の分解は時系列分析における基本的な結果であり，それによって観測値の同時密度関数を扱いやすい形式に書くことが可能になる．このことは比較的複雑な時系列モデルの最尤推定の適切な近似を示唆すると同時に，厳密な最尤推定への道を開く．

多変量正規分布，すなわち$y \sim N(\mu, \sigma^2 V)$から抽出された平均$\mu$，そして共分散行列$\sigma^2 V$を持つ$T$個の時系列観測値の集合を考えよう．ここでスカラー量の$\sigma^2$を導入した動機は分析上の便利さのためである．それを含めることは主な結果にとっては末梢的な問題である．以下の結果は$\sigma^2$を1と等しいと置いたとしても本質的に変化しない．多くの時系列モデルにおいては$\mu$の要素は同一であるか，またある場合にはゼロである．しかしながら$\mu$を制約しないでおくことは，一見するよりもはるかに大きな一般性を許すことになる．とくにもし$\mu$が$\mu = X\beta$，ここで$X$と$\beta$は(2.12)で定義されたもの，とすれば一般化線型回帰モデルが得られる．

観測値の同時密度は次のように与えられる．

(2.15) $$\log L\ (y) = -(T/2)\log 2\pi - (T/2)\log \sigma^2 - (1/2)\log |V| - (1/2)\sigma^{-2}(y-\mu)'V^{-1}(y-\mu).$$

以下のように書くことによって，これは二つの部分に分けることができる．

(2.16) $$\log L\ (y) = \log L\ (y_1, \cdots, y_{T-1}) + \log l\ (y_T | y_{T-1}, \cdots, y_1).$$

上記(2.16)の右辺の第二項は，それまでのすべての観測値に条件づけられた

(conditional)最後の観測値の分布である．尤度をこのように分けるには条件付確率の定義が用いられている．条件付確率の最も初歩的な関係は以下のように表される．

$$Pr(A)=Pr(A/B)\cdot Pr(B).$$

ここで$A$と$B$は事象である．$\log L(y)$は標本$y_1,\cdots,y_T$をもたらす同時確率変数$y$の密度関数の対数であるから，上式において$A$を$y$とし$B$を$y_T$を除いた$y$の部分集合と置くことによって(2.16)が得られたのである．

さて，ここで$y_{T-1},\cdots,y_1$が既知であるとしたときに，$y_T$を推定する問題を考えよう．もし$\hat{y}_{T/T-1}$がこのような基礎の上に作られた$y_T$の推定量とするならば，推定または予測の誤差は次のように二つに分けることができる．

$$(2.17)\qquad y_T-\hat{y}_{T/T-1}=[y_T-E(y_T/y_{T-1},\cdots,y_1)]+[E(y_T/y_{T-1},\cdots,y_1)-\hat{y}_{T/T-1}].$$

ここで$E(y_T/y_{T-1},\cdots,y_1)$は$y_{T-1},\cdots,y_1$を条件付きとした$y_T$の分布の平均である．そこで(2.17)よりわれわれは以下のような結果を得る．

$$(2.18)\qquad \mathrm{MSE}(\hat{y}_{T/T-1})=\mathrm{Var}(y_T/y_{T-1},\cdots,y_1)+E\{[\hat{y}_{T/T-1}-E(y_T/y_{T-1},\cdots,y_1)]^2\}.$$

この時(2.18)の右辺の第一項は$\hat{y}_{T/T-1}$から独立である．ゆえに$y_{T-1},\cdots,y_1$を条件とした$y_T$の最小平均二乗推定量(minimum mean square estimator：MMSE)は次式で与えられる．

$$\hat{y}_{T/T-1}=E(y_T/y_{T-1},\cdots,y_1).$$

この$\hat{y}_{T/T-1}$に伴われる予測誤差の分散は$E[(y_T-\hat{y}_{T/T-1})^2]$である．これはMSE$(\hat{y}_{T/T-1})$と同一であり，$\hat{y}_{T/T-1}$の構造よりVar$(y_T/y_{T-1},\cdots,y_1)$に等しい．この量は以下で$\sigma^2f_T$と表わす．第二項は

$$\log l(y_T/y_{T-1},\cdots,y_1)=-(1/2)\log 2\pi-(1/2)\log\sigma^2-(1/2)\log f_T-(1/2)\sigma^{-2}(y_T-\hat{y}_{T/T-1})^2/f_T,$$

より予測誤差$y_T-\hat{y}_{T/T-1}$の分布として解釈される．

上記(2.16)における分割は，最初の$T-1$個の観測値の尤度に関しても繰り返すことができる．それをさらに繰り返すことによって，われわれは最終的に次のような結果を得る．

$$(2.19)\qquad \log L(y)=\sum_{t=2}^{T}\log l(y_t/y_{t-1},\cdots,y_1)+\log l(y_1).$$

なお$t=2,\cdots,T$に関して，$y_{t-1},\cdots,y_1$を条件とした$y_t$の平均は，$\hat{y}_{t/t-1}$，すなわちそれ以前の観測値が与えられた下での$y_t$の最小平均二乗推定量に等しくなければならない．したがって，それぞれの条件付分布は最適予測にまつわる誤差の分布ということになり，一方$l(y_1)$は$y_1$の無条件分布となる．ただし，それ以前の観測値が無いという状況の下で，$\mu_1$が$y_1$の最小二乗推定量であるとみなされるのであるならば，$y_1-\mu_1$が$y_1$にまつわる予測誤差ということになる．したがって$y_1$の分散を$\sigma^2f_1$と表わすことは適切であろう．



以上の議論は表現(2.19)を用いて尤度関数を以下のような$T$個の独立な予測誤差の結合分布として分解することを可能とする．

$$(2.20)\qquad v_t=y_t-\hat{y}_{t/t-1},\qquad t=1,\cdots,T.$$

ここで$\hat{y}_{1/0}=\mu_1$であり，それぞれの予測誤差$v_t$は平均ゼロ，分散$\sigma^2f_t$を持つ．ゆえに(2.15)は以下のように書ける．

$$(2.21)\qquad \log L(y)=-\frac{T}{2}\log 2\pi-\frac{T}{2}\log\sigma^2-\frac{1}{2}\sum_{t=1}^{T}\log f_t-\frac{1}{2}\sigma^{-2}\sum_{t=1}^{T}v_t^2/f_t.$$

予測誤差の分解は，$V^{-1}$のコレスキー(Cholesky)分解として理解することができるであろう．もし$\bar{L}$を対角要素が1である下三角行列とすると，$V^{-1}$は$V^{-1}=\bar{L}'D\bar{L}$として分解できる．ここで$D$は対角成分が$f_1^{-1},f_2^{-1},\cdots,f_T^{-1}$なる対角行列である．この分解は一意であり，(2.20)における予測誤差は$v=\bar{L}y$なる変換によって与えられる．この変換のヤコビアンは$|\bar{L}|=1$であるので，ベクトル$v=(v_1,\cdots,v_T)'$における要素の結合分布は(2.21)で与えられたものとなる．ここで$|V^{-1}|=|\bar{L}'|\cdot|D|\cdot|\bar{L}|=|D|$であり，ゆえに$\log|V|=\sum\log f_t$となることに注意されたい．

コレスキー分解は数値的に$V^{-1}$を分解すること，および$\bar{L}y$なる変換によって直接的に予測誤差を計算するための基礎として用いることができる．他方，もとになっているモデルは状態空間形式にも表わされることもでき，その場合には予測誤差はカルマン・フィルターによって計算される．これが5章でとられているアプローチの概略である．

実際の問題において尤度関数への近似が適当であると考えられる場合には，予測誤差分解によって非常に大きな簡単化を成し遂げることができる．一般的慣例として，初期観測値についてのある種の仮定を置くことによってそのような近似は理論的に正当化され，しばしば一定の分散$\sigma^2$を持つ予測誤差という

結果をもたらす．その場合にはすべての $t$ について $f_t=1$ である．さらにその場合の対数尤度関数は次のようになる．

$$\log L=-\frac{T}{2}\log 2\pi-\frac{T}{2}\log\sigma^2-\frac{1}{2}\sigma^{-2}\sum\epsilon_t^2. \tag{2.22}$$

ここで記号の $v_t$ から $\epsilon_t$ への変化は，それぞれの誤差が今やモデルにおける基礎となっているホワイト・ノイズである攪乱項 $\epsilon_t$ と一致するということを示している．さてもし尤度関数が(2.22)のように簡略化されるのであれば，それは推定にとって重要な意味合いをもつ．すなわち尤度関数を最大化することは，以下のような二乗和関数を最小化することに等しくなる．

$$S(\psi)=\sum\epsilon_t^2. \tag{2.23}$$

ここで $\psi$ はモデルに含まれる $\sigma^2$ 以外のすべての母数を表わすとする．

**例1** 第1節で導入された以下のような一次の自己回帰モデルを考えよう．

$$y_t=\phi y_{t-1}+\epsilon_t,\qquad |\phi|<1. \tag{2.24}$$

ここで $\epsilon_t\sim NID\ (0,\sigma^2)$ である．以下では尤度関数を作成するに当って，この確率過程は遠い過去のある時点から始まったが，$t=1,\cdots,T$ の時刻にのみ観測されたと想定しよう．

任意の $t$ について過去の観測値がすべて与えられたときには，$y_t$ の分布は平均 $\phi y_{t-1}$ および分散 $\sigma^2$ を持つ．したがって $t=1$ を除く $T-1$ 個の予測誤差は $t=1$ を除く $T-1$ 個の攪乱項と同一である．というのは以下のような関係が成立するからである．

$$v_t=y_t-\hat{y}_{t|t-1}=y_t-\phi y_{t-1}=\epsilon_t,\qquad t=2,\cdots,T. \tag{2.25}$$

先の(2.19)に対応して，対数尤度関数は $\epsilon_2,\cdots,\epsilon_T$ までの密度関数と $y_1$ の無条件密度に分割される．このとき後者は平均ゼロ，分散 $\sigma^2/(1-\phi^2)$ なる正規分布に従う(これについては2章(2.7)を参照)．ゆえに，最初の予測誤差は攪乱項 $\epsilon_1$ ではなく $y_1$ それ自身に等しい．$v_1=y_1$ と置くことによりすべての $T$ 観測値についての対数尤度は，以下のごとく(2.21)のような形に表わすことができる．

$$\log L(y)=-\frac{T}{2}\log 2\pi-\frac{T}{2}\log\sigma^2+\frac{1}{2}\log(1-\phi^2)$$
$$-\frac{1}{2}\sigma^{-2}(1-\phi^2)y_1^2-\frac{1}{2}\sigma^{-2}\sum_{t=2}^{T}(y_t-\phi y_{t-1})^2 \tag{2.26}$$

ここで，$t=2,\cdots,T$ については $f_t=1$ であり $f_1=(1-\phi^2)^{-1}$ であることに留意しておこう．ここで上記(2.26)の第三および第四項は漸近的に無視し得る．そしてこれらの項を除くと尤度関数は(2.22)の形に簡単化される．このようにすることは $y_1$ を固定されたものとみなすことによって形式的に正当化される．このとき $\phi$ の最尤推定量は線型となり，$y_t$ を $y_{t-1}$ の上に回帰することによって得られる．



### 多変量モデルの尤度関数の分解

各時点において $N\times 1$ ベクトルとして観測される多変量時系列に対しても，予測誤差の分解が拡張される．すなわち，$v_t$ を平均ゼロ，共分散行列 $F_t$ を持つ $t$ 時点における予測誤差からなる $N\times 1$ ベクトルとすると，議論は以前と全く同様に行なわれる(多変量の場合には，スカラー量 $\sigma^2$ を取り出す意味はそれほどない)．この場合，対数尤度は以下のように分解される．

$$\log L(y_1,\cdots,y_T)=-\frac{TN}{2}\log 2\pi-\frac{1}{2}\sum_{t=1}^{T}\log|F_t|$$
$$-\frac{1}{2}\sum_{t=1}^{T}v_t'F_t^{-1}v_t. \tag{2.27}$$

ここで，$y_1\sim N(\mu_1,F_1)$ および $v_1=y_1-\mu_1$ である．もとのモデルを状態空間型式に表せば，やはりカルマン・フィルターを(2.27)の導出に用いることができる．しかしながら，一変量の場合と同様に $v_t$ をその過程を形成しているホワイト・ノイズとみなすことにより，$\log L$ を近似することがしばしば可能となる．もし $\epsilon_t\sim N(0,\Omega)$ であるなら，その近似は次のような形で与えられる．

$$\log L=-\frac{TN}{2}\log 2\pi-\frac{T}{2}\log|\Omega|-\frac{1}{2}\sum\epsilon_t'\Omega^{-1}\epsilon_t. \tag{2.28}$$

### 尤度関数の最大化

数値的な最適化については数多くの異なったアプローチがある．しかしながら，尤度関数の最大化の問題に関しては，以下のような漸化式で記述される重要な計算手順がある．

$$\psi^*=\hat{\psi}+[I^*(\hat{\psi})]^{-1}\{\partial\log L/\partial\psi\}. \tag{2.29}$$

ここで $I^*(\psi)$ は情報量行列またはその近似を示し，(2.29)においてそれは各反復時の推定値 $\hat{\psi}$ で評価されるものとする．微係数ベクトル $\partial\log L/\partial\psi$ もまた $\hat{\psi}$ で評価し，(2.29)の適用は改善(更新)された推定値 $\psi^*$ をもたらす．この繰り返しは収束するまで続けられる．実際にはこの方法において，たとえば可変なステップ幅の導入などのいくつかの変更がなされる．それについては Har-

vey (1981 a, 4.3 節) を参照されたい.

もし $I^*(\psi)$ が $-\log L$ のヘッセ (Hessian) 行列と定義されるときは, (2.29) における反復法はニュートン=ラフソン (Newton=Raphson) 法として知られているものとなる. このヘッセ行列の期待値は情報行列それ自身であり, それを用いる方法はスコアリング法 (method of scoring) として知られている. もし尤度関数の最大化が, (2.23) における二乗和関数を最大化にするのと等価であるような特殊な場合には,

$$\frac{\partial \log L}{\partial \psi} = \frac{1}{\sigma^2} \sum z_t \epsilon_t$$

となる. ここで

$$z_t = -\partial \epsilon_t / \partial \psi$$

である. 行列 $I^*(\psi)$ を以下のように定義すると

$$I^*(\psi) = \sigma^{-2} \sum z_t z_t',$$

次のような反復法が導かれる.

(2.30) $$\psi^* = \hat{\psi} + (\sum z_t z_t')^{-1} \sum z_t \epsilon_t.$$

これはガウス=ニュートン (Gauss=Newton) 法として知られている. 反復の各段階における $\epsilon_t$ およびその微係数ベクトルは前段階の推定値 $\hat{\psi}$ で評価したものである. そして推定値 $\hat{\psi}$ は $\epsilon_t$ を $z_t$ に回帰したものによって修正 (更新) される. この反復法は母数 $\sigma^2$ から完全に独立である. 反復が収束したときに, $\sigma^2$ は残差 $\epsilon_t(\tilde{\psi})$ より以下の算式により推定される.

(2.31) $$\tilde{\sigma}^2 = T^{-1} \sum \epsilon_t^2(\tilde{\psi}).$$

### 二段階推定量

上記の (2.29) は漸近的に有効な二段階 (two-step) 推定量を導出する等式として利用することもできる. もし $\hat{\psi}$ が $\psi$ の一致 (consistent) 推定量であり, いくつかの正則 (regularity) 条件が満たされれば, $\psi^*$ は最尤推定量と同一の漸近分布 (2.4) を持つ. しかしその小標本特性は非常に異なるかもしれない.

## 1.3 検　　定

仮説検定の方法は, 最尤法によるアプローチから系統的に導くことができる. 基本的な検定方法は尤度比検定である. しかし他の二つの検定, ワルド検定



(Wald test) とラグランジュ乗数検定も同等の大標本特性を有し, しばしばより魅力的なものとなる. これらすべての検定は以下の最初の項で説明される. それにつづいて, このような古典的枠組にうまくあてはまらない検定についてのある種の側面を取り扱う. より詳しい議論は Harvey (1981 a) の 5 章を参照されたい.

### 古典的検定の方法

尤度比 (likelihood ratio : LR) 検定は, 母数ベクトル $\psi$ についての一連の制約群の妥当性を検証することに主に用いられる. もし制約条件が線型であるときには, それらは次式で表現される.

(3.1) $$R\psi = r,$$

ここで, $R$ は $m \times n$ の定数からなる行列, $r$ は $m \times 1$ の定数からなるベクトル, そして $m$ は制約の数であり, それは $n$ より小さくなければならない.

帰無仮説 $H_0$ の下に, $\psi$ は (3.1) の制約を満たす. 制約が課せられたときの $\psi$ の最尤推定量 $\tilde{\psi}_0$ が, 制約を受けないときの推定量 $\tilde{\psi}$ と対比してみよう. もし $H_0$ の下での最大化された尤度関数 $L(\tilde{\psi}_0)$ が, 制約されないで最大化された尤度関数 $L(\tilde{\psi})$ よりもはるかに小さいときには, 帰無仮説に反する証拠があることになる. この考え方は尤度比

(3.2) $$\lambda = L(\tilde{\psi}_0)/L(\tilde{\psi}),$$

に基づく検定が, 望ましい統計的性質を持つことを示しているネイマン=ピアソン (Neyman=Pearson) の補題として知られている.

しばしば尤度比は, $H_0$ の下での厳密な分布が知られている統計量に変換することが可能である. もしこれができないときには, 大標本における検定が行なわれる. それは統計量

(3.3) $$\mathrm{LR} = -2 \log \lambda$$

が $H_0$ の下で漸近的に $\chi_m^2$ (自由度 $m$ のカイ二乗) 分布するという性質を利用する.

尤度比検定の弱点は, モデルを帰無仮説と対立仮説の双方の下で推定しなければならない点にある. ワルド検定 (Wald test) と呼ばれるもう一つの方法は, 制約のないモデルからの $\psi$ の推定値のみを必要とし, 検定統計量は通常次式で与えられる.

(3.4) $$W = (R\tilde{\psi} - r)' [R I^{-1}(\tilde{\psi}) R']^{-1} (R\tilde{\psi} - r).$$

ここで，$I(\tilde{\psi})$は制約のない推定値$\tilde{\psi}$で評価された情報行列である．帰無仮説$H_0$の下で，$W$は漸近的に$\chi^2_m$分布し，その大標本特性は尤度比検定の場合と同様であることが知られている．

もし帰無仮説の下でのモデル推定の方が容易であるならば，ラグランジュ乗数(Lagrange multiplier)検定が適当かもしれない．この検定統計量はやはり$H_0$の下で漸近的に$\chi^2_m$分布し，次式で与えられる．

$$\text{(3.5)} \qquad \mathrm{LM} = \left(\frac{\partial \log L}{\partial \psi}\right)' I^{-1}(\tilde{\psi}_0)\left(\frac{\partial \log L}{\partial \psi}\right).$$

ここで，$\partial \log L/\partial \psi$は制約された推定値$\tilde{\psi}_0$で評価する．ワルド検定と同様に，ラグランジュ乗数検定の大標本特性は尤度比検定の場合と同一である．ただし，制約のないモデルの推定は必要ない．

制約のないモデルの尤度関数の最大化が二乗和関数の最小化と等価な場合には，(3.5)をわずかに修正することによって特に便利な検定統計量がもたらされる．残差$\epsilon_t$とその微係数ベクトル$\partial \epsilon_t/\partial \psi$を$\psi = \tilde{\psi}_0$で評価したものとし，$\epsilon_t$を$\partial \epsilon_t/\partial \psi$の上に回帰すると，重相関係数$R^2$を得る．すると統計量

$$\text{(3.6)} \qquad \mathrm{LM}^* = TR^2$$

はラグランジュ乗数統計量(3.5)と漸近的に等価であることが知られている．そしてそれは(3.5)と同様に$\chi^2_m$変数として検定できる．さらにしばしば$F$分布に基づく「修正されたラグランジュ乗数検定」が便利に使用される．このアプローチのもととなる一般的原理は本節で与えた例から明らかであろう．

### 非入れ子型仮説とモデルの診断

古典的検定の手続きは入れ子型仮説(nested hypothesis)にかかわっている．そこでは帰無仮説の下でのモデルはより一般的モデルの特殊ケースとなっている．しかしながら，しばしばいずれのモデルも他のモデルの特殊ケースではないような二つのモデルを比較する必要が生じる．そのようなモデルは非入れ子型(non-nested)と呼ばれ，それらを扱うには基本的に二つの方法がある．第一の方法はそれらを区別するある基準を定義することである．第二の方法はそれぞれの定式化を他に対して検定する検定手続きを用いることである．第二の方法では四つの結果が可能である．なぜなら一つのモデルが採択され他方が棄却される場合に加えて，双方が採択されたり，双方が棄却される可能性もあるからである．



最大化された尤度関数は，モデルを区別する一つの基準を提供する．しかしながら，その場合に推定される母数の数を考慮に入れる必要があろう．この考え方は，例えば赤池の情報量基準(Akaike Information Criterion：AIC)として定式化される．これは以下のものを最小にするモデルを選ぶ決定法である．

$$\text{(3.7)} \qquad \mathrm{AIC} = -2\log L(\tilde{\psi}) + 2n.$$

ここで，$n$は母数の数である．一つのモデルを他のモデルに対して検定する方法はより複雑であり，本書では扱わない．

モデルの診断(diagnostic checking)は古典的枠組にはそのままでは当てはまらないような検定の一つの側面である．もしモデルが適切なものであるならば，残差は近似的にランダムとなるべきであり，多くの診断的チェックの手続きは独立性からの乖離を検出しようとするものである．これらの手続きは図的解釈によるものもあろうし，また正式な検定統計量の形をとるものかもしれない．しかしモデルの診断の本質的な特徴は，研究者がモデルの不適切さの原因について確固たる見解がないことである．したがって診断のための検定統計量は，ラグランジュ乗数による接近のように特定の対立仮説を想定して導くこともできるが，帰無仮説の棄却は必ずしもその特定の対立仮説を意味するとは限らないことになる．標準的な例としては線型回帰分析におけるダービン=ワトソン(Durbin=Watson)検定である．この検定は一階の自己回帰に従う撹乱項の対立仮説に対して高い検出力を持つように導かれている．しかしながら，検定統計量が有意な値を取る場合には，それが誤った関数形や主要変数の脱落なども含めた広範囲な定式化の誤りを示唆していると考えることもできよう．

### 練習問題

**1.** 以下の数値は1967年～71年の米国の個人消費支出(10億ドル)である．29.2, 31.2, 33.8, 36.4, 39.4．大域的線型トレンド法(すなわち線型回帰)を用いて，傾きを推定せよ．そして1972年の値を予測せよ．

実際には1972年～74年の値は，43.3, 47.3, 52.9であった．局所的線型トレンド法(すなわちホルト=ウィンターズ法)を用いて，平滑化定数を0.1とし，上記の間で当てはめられた式を出発点とし，それぞれの年の平均と傾きを更新せよ．そして1975年の値を予測せよ．このデータを扱うにあたっての方法の適切さを論じよ［ヨーク大学，1980年］．

**2.** 以下の割引二乗和

$$S(\mu)=\sum_{t=1}^{T}\delta^{T-t}(y_t-\mu)^2, \qquad 0<\delta\leqq 1,$$

を $\mu$ に関して最小化して求めた $\mu$ の推定量は，幾何級数的に減少するウェイトによる観測値の加重和であることを示せ．もしこの推定量が $y_{T+1}$ の予測 $\tilde{y}_{T+1/T}$ として用いられるなら，この予測は次の漸化式を満たすことを示せ．

$$\tilde{y}_{T+1/T}=(1-\delta^*)y_T+\delta^*\tilde{y}_{T/T-1}, \qquad \delta^*=\delta\left(\frac{1-\delta^{T-1}}{1-\delta^T}\right).$$

ここで，$\tilde{y}_{T/T-1}$ は最初の $T-1$ の観測値に基づいた $y_T$ の予測値である．$\delta$ の値と局所的および全域的トレンドとの関係について論じよ．



# 2章
# 定常確率過程と時間領域におけるその性質

## 2.1 基本概念

本書では主として時系列を確率過程としてとらえてモデル分析を行なう．ある確率過程の各時点における観測量は確率変数であって，何らかの確率法則に従って時間とともに変化するものと考える．したがって，一つの確率過程は時間について順序のついた確率変数の列として定義されよう．

一つの統計モデルは観測値を発生させるメカニズムを定める．例えば1次の移動平均過程

(1.1) $$y_t=\epsilon_t+\theta\epsilon_{t-1}, \qquad t=1,\ \cdots,\ T$$

を考えてみよう．ここで $\{\epsilon_t\}$ は平均ゼロで分散一定である互いに無相関な確率変数列であって，$\theta$ は未知母数である．$\epsilon_0,\ \epsilon_1,\ \cdots,\ \epsilon_T$ の一組の実現値に対応して一組の観測値 $y_1,\ y_2,\ \cdots,\ y_T$ が得られる．$\epsilon_0,\ \epsilon_1,\ \cdots,\ \epsilon_T$ が別の実現値をとればそれに対応した観測値の組が得られることになる．すなわちモデル(1.1)は期間 $t=1,\ 2,\ \cdots,\ T$ におけるこのような実現値の無限の組の集合を生成するものととらえることができよう．いいかえると，このモデルは確率変数 $y_1,\ y_2,\ \cdots,\ y_T$ の同時分布を規定している．

一つの確率過程のモーメントは確率変数 $y_1,\ y_2,\ \cdots,\ y_T$ の分布に関して定義される．時刻 $t$ における確率過程の平均(mean)とは

(1.2) $$\mu_t=E(y_t), \qquad t=1,\ \cdots,\ T$$

であるが，これは確率変数 $y_t$ の可能なすべての実現値についての平均と解釈されよう．2次モーメントについても同様に解釈される．時刻 $t$ における $y_t$

の分散は

(1.3) $$\mathrm{Var}\,(y_t)=E[(y_t-\mu_t)^2], \qquad t=1,\cdots,T$$

によって，また$y_t$と$y_{t-\tau}$の共分散は

(1.4) $$\mathrm{Cov}\,(y_t,y_{t-\tau})=E[(y_t-\mu_t)(y_{t-\tau}-\mu_{t-\tau})], \qquad t=\tau+1,\cdots,T$$

によってそれぞれ定義される．

さて，もし$y_t$の実現値が複数系列観測されるとすれば，以上で定義されたモーメントはその「複数の系列についての」平均(ensemble)によって推定することができる．例えば(1.2)は，

(1.5) $$\hat{\mu}_t=m^{-1}\sum_{j=1}^{m}y_t^{(j)}, \qquad t=1,\cdots,T$$

によって推定されよう．ただしここで$y_t^{(j)}$とは時刻$t$における$y_t$の$j$番目の系列の観測値であり，$m$は実現系列の数である．しかしながら大部分の時系列では，観測されるのは単一の実現系列のみにすぎない．この場合には観測値を生成している確率過程について，何らかの制約条件を課すことなしには(1.2)や(1.4)などのモーメントについての意味のある統計的推測を行なうことはできない．実は以上のことから定常性(stationarity)の概念が要請されるのである．

## 定　常　性

観測量の一つの実現系列(realisation)のみが利用可能であるときには，各時刻の観測量から，何らかの意味で時間について平均化された観測量を考えねばならなくなる．それには(1.2)，(1.3)，(1.4)で与えられる量が時刻$t$に依存しないことが必要となる．すなわち，たとえば$t=1,\ 2,\ \cdots,\ T$に対して$\mu_t=\mu$となっていれば，観測値の時間平均をとることによって$\mu$を推定することが可能である．

ある確率過程が定常(stationary)であるためには，すべての$t$に対し条件

(1.6) $$E(y_t)=\mu,$$

(1.7) $$E[(y_t-\mu)^2]=\gamma(0),$$

(1.8) $$E[(y_t-\mu)(y_{t-\tau}-\mu)]=\gamma(\tau), \qquad \tau=1,2,\cdots$$

が満たされねばならない．この条件の中で(1.6)と(1.7)はデータを生成している確率過程の平均と分散を定義し，(1.8)は時差(lag)$\tau$の自己共分散を定義している．(1.6)と(1.7)の意味については図1.2(b)に関する説明ですでにふれている．さて(1.6)～(1.8)は単一の観測値系列から次の量によって推定するこ



とができる．

(1.9) $$\hat{\mu}=\bar{y}=T^{-1}\sum_{t=1}^{T}y_t,$$

(1.10) $$\hat{\gamma}(0)=c(0)=T^{-1}\sum_{t=1}^{T}(y_t-\bar{y})^2,$$

(1.11) $$\hat{\gamma}(\tau)=c(\tau)=T^{-1}\sum_{t=\tau+1}^{T}(y_t-\bar{y})(y_{t-\tau}-\bar{y}), \qquad \tau=1,2,3,\cdots.$$

ここでもし確率過程がエルゴード性(ergodic)を持っていれば，これらの統計量はそれぞれ平均，分散，自己共分散の一致推定量となる．エルゴード性については本書では正確に定義を与えないが，基本的には十分遠い観測量がほぼ無相関となることを要請している．本章で考察されるすべてのモデルでは定常性がエルゴード性を意味するので，この問題にさらに深入りする必要はないであろう．

さて条件(1.6)～(1.8)は弱(weak)定常性あるいは共分散(covariance)定常性の定義である．しばしば強(strict)定常性の条件を課すこともあるが，それはより強い条件であり，時刻$t_1,\ t_2,\ \cdots,\ t_r$における$r$個の観測量の同時分布が時刻$t_1+\tau,\ t_2+\tau,\ \cdots,\ t_r+\tau$における$r$個の観測量の同時分布と同一であることを意味している．本章では定常性とは常に弱定常性を意味する．もちろん確率過程が弱定常で正規分布に従っていれば，同時に厳密な意味で定常であることには注意する必要があろう．

定常確率過程の最も簡単な例は平均と分散が一定で互いに無相関である確率変数列である．この種の確率過程は「ホワイト・ノイズ」(white noise)として知られているが，この用語自体は工学の文献から借りてきたものである．本書全体を通して記号$\epsilon_t$は常にホワイト・ノイズ変数を表わし，もし他に明確に言及されない限り平均はゼロ，分散は$\sigma^2$とする．ホワイト・ノイズ系列の確率変数は互いに無相関であるから，ゼロでない時差の自己共分散はすべてゼロとなる．すなわち

(1.12) $$E(\epsilon_t\epsilon_{t-\tau})=\begin{cases}\sigma^2, & \tau=0\\ 0, & \tau\neq 0\end{cases}$$

が成立する．

### 自己共分散関数と自己相関関数

ある確率過程が定常であるならば $\gamma(\tau)$ を $\tau$ に対しプロットすることによって時間領域での性質を要約することができる．これは自己共分散関数(autocovariance function)として知られている．$\gamma(\tau)=\gamma(-\tau)$ となることより，負の $\tau$ に対してプロットする必要はない．

自己共分散は確率過程の分散で割ることによって基準化され，自己相関

(1.13) $$\rho(\tau)=\gamma(\tau)/\gamma(0), \qquad \tau=0,\pm1,\pm2,\cdots$$

が得られる．$\rho(\tau)$ の非負の $\tau$ に対するプロットによって自己相関関数(autocorrelation function)が得られる．ここで定義から $\rho(0)=1$ となる．

ところで移動平均モデル(1.1)の時間領域での性質は比較的容易に求められる．$\epsilon_t$ は平均ゼロ，分散 $\sigma^2$ のホワイト・ノイズ確率変数であるから，

(1.14 a) $$\mu=E(\epsilon_t)+\theta E(\epsilon_{t-1})=0$$

となるとともに，

(1.14 b) $$\begin{aligned}\gamma(0)&=E[(\epsilon_t+\theta\epsilon_{t-1})(\epsilon_t+\theta\epsilon_{t-1})]\\&=E(\epsilon_t^2)+\theta^2E(\epsilon_{t-1}^2)+2\theta E(\epsilon_t\epsilon_{t-1})\\&=(1+\theta^2)\sigma^2\end{aligned}$$

となる．同様にして

(1.14 c) $$\begin{aligned}\gamma(1)&=E[(\epsilon_t+\theta\epsilon_{t-1})(\epsilon_{t-1})+\theta\epsilon_{t-2})]\\&=E(\epsilon_t\epsilon_{t-1})+\theta E(\epsilon_{t-1}^2)+\theta E(\epsilon_t\epsilon_{t-2})+\theta^2E(\epsilon_{t-1}\epsilon_{t-2})\\&=\theta E(\epsilon_{t-1}^2)\\&=\theta\sigma^2,\end{aligned}$$

かつ

(1.14 d) $$\gamma(\tau)=0, \qquad \tau=2,3,4,\cdots$$

となる．したがって，平均・分散・自己共分散はいずれも $t$ と独立となり定常確率過程であることがわかる．

自己共分散関数と自己相関関数は全く同一の形状をしていて，確率過程の性質については同一の情報を提供している．ただし自己相関関数は次元を持たないので，より一般的にプロットして用いられる．たとえば(1.14 c)を基準化すると，

(1.15) $$\rho(1)=\theta/(1+\theta^2)$$

となる．図2.1は $\theta=0.5$ に対する自己相関関数を示している．

$\theta$ が正ならば，前後する $y_t$ は正の相関を持ち，確率過程は独立確率変数列



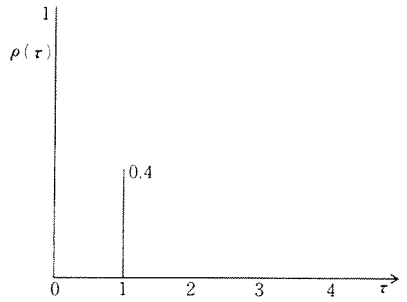



**図 2.1**　MA(1)過程の自己相関関数　$\theta=0.5$

$\epsilon_t$ よりもさらに滑らかに変動する．他方もし $\theta$ が負ならば，時系列は正の $y_t$ の値に負の値が続き，あるいは負の値に正の値が続く傾向が生じ，独立確率変数列 $\epsilon_t$ よりもさらに不規則に変動する．このことが自己相関関数に反映して $\theta<0$ ならば $\rho(1)$ は負となる．

### コレログラム

さて(1.9)，(1.10)，(1.11)で定義される統計量はそれぞれ標本平均(sample mean)，標本分散(sample variance)，標本自己共分散(sample autocovariance)と呼ばれる．標本自己共分散は理論的な自己共分散と同じように基準化される．すると

(1.16) $$r(\tau)=c(\tau)/c(0), \qquad \tau=1,2,\cdots$$

となり，$r(\tau)$ の非負の $\tau$ に対するプロットは標本自己相関関数あるいはコレログラム(correlogram)と呼ばれる．

標本自己相関関数はデータを発生させている確率変数の自己相関関数の理論値の推定値である．したがって標本によるバラツキの影響を受ける．すなわちコレログラムは自己相関の理論値の性質を反映するが，全く同一な性質を再生するわけではない．たとえばホワイト・ノイズ過程からの標本自己相関は理想的には常にゼロであるべきであるが，一般にはゼロに近い値をとる．

このコレログラムは観測値系列の性質を時間領域で分析する主要な用具である．ただここでコレログラムの解釈に際していくつか知っておくべきことがある．すなわち第一には，推定された自己相関の標本変動であり，第二には，種々の確率過程に対する自己相関関数の構造の特色である．以下では数節を費

して自己回帰移動平均過程のクラスに限定して種々の特殊な場合に関する自己相関関数の性格を検討する．そのことが6章で解説されるモデル作成法の基礎となるであろう．標本変動とコレログラムに関連した検定方法については6章で同じく考察される．

### ラグ作用素

ラグ作用素 $L$ は変換

(1.17) $$Ly_t = y_{t-1}$$

によって定義され，時系列分析における種々の代数演算上で有用な役割をはたす．$L$ を $y_{t-1}$ に作用することによって $Ly_{t-1} = y_{t-2}$ となるので，(1.17)に代入することによって，$L(Ly_t) = L^2 y_t = y_{t-2}$ 等々となり，一般には

(1.18) $$L^\tau y_t = y_{t-\tau}, \qquad \tau = 1, 2, 3, \cdots$$

となる．論理的首尾一貫性を保つために $L^0 y_t = y_t$ によって $L^0$ を定義すれば(1.18)は総ての非負整数について成立することになる．

このラグ作用素は，通常の代数的数値と同様に操作することができる．$\epsilon_{t-j}$ $(j = 0, 1, \cdots,)$ の係数が $\phi^j$ であって $|\phi| < 1$ となる無限次移動平均過程を考えよう．モデルは

(1.19) $$y_t = \sum_{j=0}^{\infty} (\phi L)^j \epsilon_t$$

と書けるが，ここでもし $L$ が性質 $|L| \leqslant 1$ を満たしているとみなせば $|\phi L| < 1$ であることから，系列 1，$\phi L$，$(\phi L)^2, \cdots$ は無限次幾何級数の和とみなすことができよう．そこで(1.19)は，

(1.20) $$y_t = \epsilon_t / (1 - \phi L)$$

となり，さらに整理すると1次自己回帰過程

(1.21) $$y_t = \phi y_{t-1} + \epsilon_t$$

として表現されるであろう．

次に1次差分作用素(first difference operator) $\Delta$ は，$\Delta = 1 - L$ となることより，ラグ作用素と同様に扱うことができる．これら二つの作用素の関係はしばしば有用であり，たとえば

(1.22) $$\Delta^2 y_t = (1-L)^2 y_t = (1 - 2L + L^2) y_t = y_t - 2y_{t-1} + y_{t-2}$$

などと用いられる．



### 自己回帰移動平均過程

無限個のラグを移動平均表現に用いることによって確率過程の一つの一般的クラスを作ることができる．その表現は，

(1.23) $$y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

であるが，ここで $\psi_0$，$\psi_1$，$\psi_2$，…は母数であって，条件

(1.24) $$\sum_{j=1}^{\infty} \psi_j^2 < \infty$$

を課せばこの確率過程には分散が存在することになる．この形に表現されるモデルのことを非決定論的(indeterministic)過程，あるいは線型(linear)過程と呼ぶ．

任意の線型過程は定常であり，その性質は自己共分散関数によって表現されるが，この確率過程はまた自己回帰移動平均過程のクラスのモデルによって十分に近似される．次数 $(p, q)$ の自己回帰移動平均過程は通常 ARMA $(p, q)$ と略され

(1.25) $$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

と表現される．次節で示されるが，このモデルが定常過程であるためには自己回帰母数 $\phi_1$，$\phi_2$，…，$\phi_p$ に制約条件が課されねばならない．

任意の ARMA 過程についてはラグ作用素を使った随伴多項式(associated polynomials)を定義することによって縮約して表現することができる．ここで

(1.26 a) $$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$$

(1.26 b) $$\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$$

とおけば，モデル(1.25)は

(1.27) $$\phi(L) y_t = \theta(L) \epsilon_t$$

となる．この表現にはいくつかの技術的及び記号上の便利さがある．たとえば定常過程にゼロでない平均 $\mu$ を導入すると，

(1.28) $$y_t = \mu + \phi^{-1}(L)\, \theta(L)\, \epsilon_t$$

となる．$\phi(L)$ を両辺に乗ずれば，(1.27)の右辺に一つの母数

$$\theta_0 = \phi(L)\, \mu = \phi(1)\, \mu = (1 - \phi_1 - \cdots - \phi_p)\, \mu$$

を加えたモデルが得られる．

以下では3節にわたって純粋な自己回帰過程，純粋な移動平均過程，混合(すなわち ARMA)過程の性質を調べる．便宜上，確率過程の平均はゼロ，す

なわち $\mu=0$ と仮定されるが，それによって一般性を失うことはない．

## 2.2 自己回帰過程

次数 $p$ の自己回帰過程は

(2.1) $$y_t=\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\epsilon_t, \qquad t=1,\cdots,T$$

と表現されるが，これを $y_t\sim \mathrm{AR}(p)$ と表わすことにしよう．自己回帰過程はこれまでにしばしば用いられているが，その理由としてはモデルが自然に解釈されることと，MA 過程や混合過程に比較して推定が容易であることをあげることができよう．

AR モデルについて吟味すべき第一の問題は定常性の条件である．その条件は結局モデルが条件(1.24)を満足する(1.23)に表現されるか否かに帰着される．もしこの表現を導くことができれば自己相関関数を得ることができる．

### 1 次自己回帰モデルにおける定常性

AR(1)過程は

(2.2) $$y_t=\phi y_{t-1}+\epsilon_t, \qquad t=1,\cdots,T$$

と表現される．時系列は時刻 $t=1$ より観測されるが，確率過程自体は非常に遠い過去の時点から出発しているとみなされる．$y_t$ の過去の値の代入を繰り返すと，

(2.3) $$y_t=\sum_{j=0}^{J-1}\phi^j\epsilon_{t-j}+\phi^J y_{t-J}$$

を得る．(2.3)は二つの部分から構成され，その第一項は確率過程を生成しているホワイト・ノイズ変数の過去の値の移動平均である．第二項は時刻 $t-J$ における $y_t$ の値に依存している．$y_{t-J}$ を固定した数値とみなして期待値をとれば，

(2.4) $$E(y_t)=E\Big(\sum_{j=0}^{J-1}\phi^j\epsilon_{t-j}\Big)+E(\phi^J y_{t-J})=\phi^J y_{t-J}$$

となる．ここでもし $|\phi|\geq 1$ ならば確率過程の平均は初期値に依存する．すなわちこのとき表現(2.3)は非確率的要素を含むので，時系列の将来値を予測する場合にはいかに遠い将来であっても $y_{t-J}$ の情報を用いる必要がある．他方もし $\phi$ の絶対値が 1 より小さく，かつ $J$ が十分大きければ(2.4)の右辺の非確



率項は無視できよう．$J\to+\infty$ となるにつれて第二項は消滅してゆき，もし確率過程が遠い過去のある時点から出発したとすると(2.2)を

(2.5) $$y_t=\sum_{j=0}^{\infty}\phi^j\epsilon_{t-j}, \qquad t=1,\cdots,T$$

と表現することがゆるされよう．

ここで(2.5)と(1.23)を比較すると，$|\phi|<1$ となる AR(1)過程は純粋に確率的になり係数の二乗和は幾何級数を用いて，

(2.6) $$\sum_{j=0}^{\infty}\phi^{2j}=1/(1-\phi^2)$$

となる．総ての $t$ に対して $y_t$ の期待値はゼロであり，

(2.7) $$\gamma(0)=E(y_t^2)=E\Big(\sum_{j=0}^{\infty}\phi^j\epsilon_{t-j}\Big)^2=\sum_{j=0}^{\infty}\phi^{2j}E(\epsilon_{t-j}^2)$$

$$=\sigma^2\sum_{j=0}^{\infty}\phi^{2j}=\sigma^2/(1-\phi^2)$$

となる．

### 2 次自己回帰モデルにおける定常性*

2 次自己回帰過程は

(2.8) $$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\epsilon_t, \qquad t=1,\cdots,T$$

によって定義される．1 次自己回帰モデルと同様に(2.8)を確率的要素と非確率的要素の二つの部分に分解することができる．非確率的部分は出発値に依存するが，もし確率過程が定常であれば出発点を遠い過去のある時点とみなし，その影響を無視することができよう．

非確率的要素の性質を調べるために最も容易な方法は(2.8)において攪乱項を除いてみることであろう．すると同次差分方程式

(2.9) $$\bar{y}_t-\phi_1\bar{y}_{t-1}-\phi_2\bar{y}_{t-2}=0$$

が得られる．ここで記号 $\bar{y}_t$ は確率過程の平均を意味している．(2.9)の根は特性方程式

(2.10) $$x^2-\phi_1 x-\phi_2=0$$

の根に依存している．(たとえば Goldberg(1958)を参照)．さて(2.10)は 2 次

---

注　*のついた節は多少難しいので省略しても連続性は損われない．

方程式であるから二根 $m_1$, $m_2$ は

(2.11) $$(x-m_1)(x-m_2)=0$$

を満足しているので，公式より

(2.12) $$m_1, m_2=(\phi_1 \pm \sqrt{\phi_1^2+4\phi_2})/2$$

を得る．

さて(2.12)の解は判別式が正・ゼロ・負のいずれかによって三つの場合が考えられる．第一に二根とも実根の場合，(2.9)の解は，

(2.13) $$\bar{y}_t = k_1 m_1^J + k_2 m_2^J$$

で与えられる．ここで $k_1$, $k_2$ は初期値 $\bar{y}_{t-J}$, $\bar{y}_{t-J+1}$ に依存する定数である．もし $m_1$ と $m_2$ の絶対値がともに1以下であれば，$J$ が十分大きいときには $\bar{y}_t$ はゼロに近づく．

次に $\phi_1^2+4\phi_2<0$ ならば，二根は共役複素数となる．$\bar{y}_t$ の解は(2.13)となるが，さらに

(2.14) $$\bar{y}_t = k_3 r^J \cos(\lambda J + k_4)$$

と表現することができる．ここで $k_3$ と $k_4$ は時系列の初期値に依存する定数で，$r$ は根の半径，$\lambda$ は

(2.15) $$\lambda = \tan^{-1}[Im(m_1)/Re(m_1)] = \tan^{-1}\sqrt{(-\phi_1^2-4\phi_2)/\phi_1} = \cos^{-1}[\phi_1/(2\sqrt{-\phi_2})]$$

で定義されラジアン(角速度)で測る．$\bar{y}_t$ の時間的径路は循環的になり，もし根の半径が1より小さければ減衰し，$J$ が十分大きければ $\bar{y}_t$ は無視できるほど小さくなる．

もし根が実根で等根の場合には(2.9)の解は(2.13)とは違った型になるが，この場合にもやはり根の絶対値が1より小さいことが，$\bar{y}_t$ が無視できるほど小さくなるために必要となる．

ここでさらに $J$ が無限大になることがゆるされ，(2.10)の根の絶対値が1より小さければ(2.8)の非確率部分は消滅する．このとき線型過程(1.23)となるが，その係数は多項式

(2.16) $$\phi(L)=1-\phi_1 L-\phi_2 L^2$$

(2.17) $$\psi(L)=\psi_0+\psi_1 L+\psi_2 L^2+\cdots+\psi_\tau L^\tau+\cdots$$

を定義し，(2.8)と(1.23)がそれぞれ

(2.18) $$y_t=\phi^{-1}(L)\,\epsilon_t$$

(2.19) $$y_t=\psi(L)\,\epsilon_t$$



となることに注意することによって求めることができる．ここで(2.18)と(2.19)を比較すると，

(2.20) $$\phi(L)\,\psi(L)=1$$

となり，これを展開して

$$(1-\phi_1 L-\phi_2 L^2)(\psi_0+\psi_1 L+\psi_2 L^2+\cdots)=1$$

より整理すると，

(2.21) $$\psi_0+(\psi_1-\phi_1)L+(\psi_2-\phi_1\psi_1-\phi_2\psi_0)L^2+(\psi_3-\phi_1\psi_2-\phi_2\psi_1)L^3+\cdots=1$$

となる．さらに(2.21)の右辺における $L$, $L^2$, $L^3$, … の係数がすべてゼロとなることより，

(2.22) $$\psi_1-\phi_1=0$$

$$\psi_j-\phi_1\psi_{j-1}-\phi_2\psi_{j-2}=0, \qquad j\geqq 2$$

となる．ここで $j\geqq 2$ に対する $\psi_j$ は初期値 $\psi_0=1$, $\psi_2=\phi_1$ とした2次の差分方程式(2.22)によって定められる．この差分方程式は(2.9)と全く同一であるから解は(2.12)で与えられる．もし二根の絶対値がともに1より小さければ $\psi_j$ は $j\to\infty$ となるにつれてゼロに近づく．そしてゼロに近づく速さは(1.24)を十分満足し，(2.10)の根の半径が1より小さいという条件が定常性を満たす十分条件であることを容易に示すことができる．

定常性の条件を未知母数 $\phi_1$, $\phi_2$ によって表現すると，

(2.23 a) $$\phi_1+\phi_2<1$$

(2.23 b) $$-\phi_1+\phi_2<1$$

(2.23 c) $$\phi_2>-1$$

となる(たとえばGoldberg(1958, p. 171)参照)．同次差分方程式の根が複素数となる必要条件は $\phi_2<0$ であり，十分条件は

(2.24) $$\phi_1^2+4\phi_2<0$$

である．以上の条件は図2.2に要約され，境界点を除く三角形の内部の点が定常過程に対応している．斜線部分は複素根に対応している．

### $p$ 次自己回帰モデルにおける定常性

AR($p$)モデル(2.1)はもし方程式

(2.25) $$x^p-\phi_1 x^{p-1}-\cdots-\phi_p=0$$

の根の絶対値が総て1より小さいとき，あるいは同じことであるが根が単位円

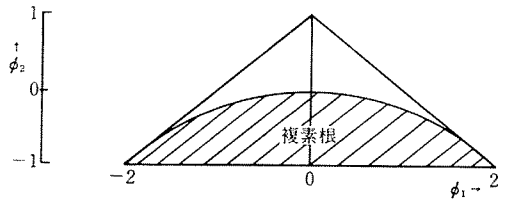



**図 2.2** 定常 RA(2)過程における $\phi_1$ と $\phi_2$ 許容領域

内にあれば定常である．この条件は随伴方程式(1.26 a)に関する条件に直すこともできる．$L$ の多項式

(2.26) $$1-\phi_1 L-\cdots-\phi_p L^p=0$$

において $x$ を $1/L$ に置き換え $L^p$ を乗ずると(2.25)が得られる．したがって定常性の条件は(2.26)の根が全て単位円周上外に存在することとなる．

### 自己共分散関数と自己相関関数

AR(1)モデル(2.2)においては $|\phi|<1$ ならば平均はゼロで分散は(2.7)で与えられた．時差(ラグ)$\tau$の自己共分散は $y_t$ を $y_{t-\tau}$ と $\epsilon_t$, $\epsilon_{t-1}$, …, $\epsilon_{t-\tau+1}$ の線型結合として表現することによって求めることができる．すなわち(2.3)において $J=\tau$ とおくと

$$\gamma(\tau)=E(y_t y_{t-\tau})=E\left[\left(\phi^\tau y_{t-\tau}+\sum_{j=0}^{\tau-1}\phi^j\epsilon_{t-j}\right)y_{t-\tau}\right]$$

となる．ここで $\epsilon_t$, …, $\epsilon_{t-\tau+1}$ はすべて $y_{t-\tau}$ と無相関であるから，

(2.27) $$\gamma(\tau)=\phi^\tau E(y_{t-\tau}^2)=\phi^\tau\gamma(0),\qquad \tau=1,2,\cdots$$

となる．自己共分散関数は $\tau$ のみに依存しているので，確率過程が定常であることに対応している．

もし定常性が始めから仮定できればより直接的に自己共分散を求めることができる．すなわち(2.2)の両辺に $y_{t-\tau}$ を乗じて期待値をとると

(2.28) $$E(y_t y_{t-\tau})=\phi E(y_{t-1}y_{t-\tau})+E(\epsilon_t y_{t-\tau}),\qquad \tau=0,1,2,\cdots$$

となる．定常過程では，$E(y_{t-1}y_{t-\tau})=E(y_t y_{t-\tau+1})=\gamma(t-1)$ となるので，$\tau>0$ に対して(2.28)の最終項は $\epsilon_t$ が過去の $y_t$ と無相関であることからゼロとなる．ゆえに

(2.29) $$\gamma(\tau)=\phi\gamma(\tau-1),\qquad \tau=1,2,\cdots$$

を得る．これは1次差分方程式となり，解は(2.27)で与えられる．さらに，



$E(\epsilon_t y_t)=\sigma^2$ に注意すると確率過程の分散も同様の計算で得ることができる．

自己相関関数は

(2.30) $$\rho(\tau)=\phi^\tau,\qquad \tau=0,1,2,\cdots$$

と表現され，$\phi$ が正ならば図2.3(a)に示されるように指数的に滑らかに減衰する．$\phi$ が負のときには自己相関関数は指数的に減衰するが，図2.3(b)に示されるように正値と負値を交互に取り振動する．$\phi$ が正のとき時系列自体も「順々に変化し」前後の変化は比較的小さい．これに対して $\phi$ が負のときには，観測値の前後の値は負の相関のためにきわめて不規則な時系列のパターンを生む．

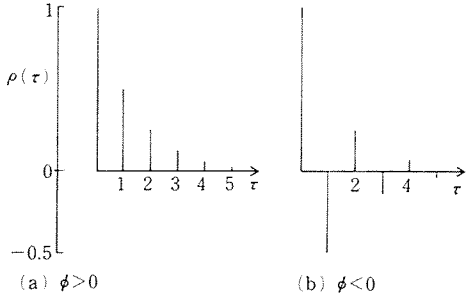



**図 2.3** AR(1)過程における自己相関関数

AR(2)過程の分散と自己共分散は(2.28)を一般化することによって得ることができる．$y_{t-\tau}$ を(2.8)に乗じて期待値をとると，

(2.31) $$E(y_t y_{t-\tau})=\phi_1 E(y_{t-1}y_{t-\tau})+\phi_2 E(y_{t-2}y_{t-\tau})+E(\epsilon_t y_{t-\tau})$$

となる．$\tau>0$ に対しては最後の項はゼロであるから，

(2.32) $$\gamma(\tau)=\phi_1\gamma(\tau-1)+\phi_2\gamma(\tau-2),\qquad \tau=1,2,\cdots$$

である．これを $\gamma(0)$ で割ると自己相関関数に関する2次の差分方程式

(2.33) $$\rho(\tau)=\phi_1\rho(\tau-1)+\phi_2\rho(\tau-2),\qquad \tau=1,2,\cdots$$

が得られる．ここで $\rho(-1)=\rho(1)$ であるから，$\tau=1$ とおくことによって，

$$\rho(1)=\phi_1+\phi_2\rho(1)$$

となるので，$\rho(0)=1$ より $\rho(1)=\phi_1/(1-\phi_2)$ によって自己相関関数の初期値が与えられる．方程式(2.33)は同次方程式(2.9)と同一である．したがって，その解はAR(2)過程の非確率的要素と同じような特性を示すこととなる．とくに複素根に対しては図2.4で例示されるように減衰的周期変動を示す．

AR(2)過程の分散は(2.31)で $\tau=0$ と置くことに得られるが，この場合最後

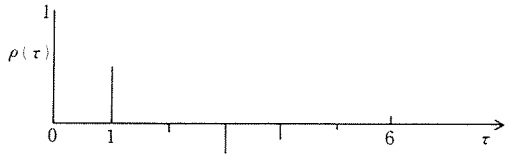



**図 2.4** 複素根を持つ AR(2) 過程の自己相関関数

の項はゼロとはならず,

$$E(\epsilon_t y_t) = E[\epsilon_t(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t)]$$
$$= \phi_1 E(\epsilon_t y_{t-1}) + \phi_2 E(\epsilon_t y_{t-2}) + E(\epsilon_t^2)$$
$$= 0 + 0 + \sigma^2 = \sigma^2$$

となるので,

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \sigma^2 \tag{2.34}$$

が得られる. これを $\gamma(0)$ で割って整理すると,

$$\gamma(0) = \frac{\sigma^2}{1 - \rho(1)\phi_1 - \rho(2)\phi_2} \tag{2.35}$$

となるが, $\rho(1) = \phi_1/(1-\phi_2)$ および $\rho(2) = \phi_1\rho(1) + \phi_2$ であることより $\gamma(0)$ を $\phi_1$ と $\phi_2$ で表わすと

$$\gamma(0) = \left(\frac{1-\phi_2}{1+\phi_2}\right) \frac{\sigma^2}{[(1-\phi_2)^2 - \phi_1^2]} \tag{2.36}$$

となる.

以上の方法は任意の AR 過程の時間領域における性質を導く際に用いることができる. $y_{t-\tau}$ を (2.1) に乗じ期待値をとって $\gamma(0)$ で割れば $p$ 次の差分方程式

$$\rho(\tau) = \phi_1\rho(\tau-1) + \cdots + \phi_p\rho(\tau-p), \qquad \tau = 1, 2, \cdots \tag{2.37}$$

を得る. $\tau = 0$ として整理すると分散は

$$\gamma(0) = \sigma^2/[1 - \rho(1)\phi_1 - \cdots - \rho(p)\phi_p] \tag{2.38}$$

となる. あらゆる定常 AR 過程に対して自己相関関数は $\tau \to \infty$ のとき $\rho(\tau)$ が減衰してゼロに収束する. そして実際の時系列が, たとえば循環変動を示すかどうかは特性方程式 (2.25) の根に依存している.

## 2.3 移動平均過程

$q$ 次移動平均過程は,

$$y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \qquad t = 1, \cdots, T \tag{3.1}$$



と表現され, これを $y_t \sim \mathrm{MA}(q)$ と書く. 任意の有限次の移動平均過程は常に定常である. 条件 (1.24) は自動的に満たされ, 自己共分散の導出にも母数 $\theta_1, \cdots, \theta_q$ に何の制約も課す必要はない. しかしながら MA 過程の自己回帰表現がしばしば必要となるが, それには MA 母数は AR 母数の定常性条件と類似の条件を満足せねばならなくなる. その条件を満足しているとき MA 過程は反転可能であるという.

### 自己共分散関数と自己相関関数

さて (3.1) の期待値をとると直ちに平均がゼロであり,

$$\gamma(0) = E(y_t^2) = (1 + \theta_1^2 + \cdots + \theta_q^2)\sigma^2 \tag{3.2}$$

となることがわかる. 自己共分散は

$$\gamma(\tau) = \begin{cases} (\theta_\tau + \theta_1\theta_{\tau+1} + \cdots + \theta_{q-\tau}\theta_q)\sigma^2, & \tau = 1, \cdots, q \\ 0 & \tau > q \end{cases} \tag{3.3}$$

で与えられる. ((1.14 c) と (1.14 d) を参照). ここで $q$ 次以上の時差の自己共分散はゼロであるので, 自己共分散関数と自己相関関数は $\tau = q$ において明瞭な切断点をもっている. このことは AR 過程の自己共分散関数がゼロへ向って徐々に収束してゆくことと対照的である. もし $q$ が小さければ $|\rho(1)|$ の最大値は 1 よりも相当小さい. 実際

$$|\rho(1)| \leq \cos[\pi/(q+2)] \tag{3.4}$$

となることが知られているので, $q = 1$ の場合には $|\rho(1)|$ の最大値は 0.5 であることを意味している. このことはもちろん (1.15) より直接示すこともできる.

### 反転可能性

MA(1) 過程において $\epsilon_t$ の過去の値を繰り返し代入することによって, $y_t$ を $y_t$ の過去の値による表現

$$y_t = \theta y_{t-1} - \theta^2 y_{t-2} + \cdots - (-\theta)^J y_{t-J} + \epsilon_t - (-\theta)^{J+1}\epsilon_{t-J-1} \tag{3.5}$$

を得る ((2.3) を参照). ここで $y_t$ が遠い過去の時刻でシステムに生じた撹乱に依存していないとすると, 母数 $\theta$ の絶対値は 1 より小さくなければならない. さらに, もし $J$ を無限大に近づければ (3.5) の最後の項は消滅するので, $y_t$ は係数が減少してゆく無限次自己回帰モデル

$$y_t = -\sum_{j=1}^{\infty} (-\theta)^j y_{t-j} + \epsilon_t \tag{3.6}$$

を得る．

MA(1)モデルにおいて $|\theta|>1$ のときには反転可能ではないが定常である．ただしこのとき，自己相関関数は母数 $1/\theta$ の反転可能な MA 過程によって再生することができる．それは(1.15)において

$$\rho(1)=\frac{1/\theta}{1+(1/\theta)^2}=\frac{\theta}{1+\theta^2} \tag{3.7}$$

となることから容易にわかる．したがって $|\theta|=1$ の場合を除いて一つの自己相関関数は二つの確率過程に対応していて，その中の一つのみが反転可能となっている．反転可能な過程に限定すれば母数の識別問題は解決され得るが，さらにそのことによって予測が非効率的になることを避けることができる．(**6.4** 節参照)．

MA(1)過程で $|\theta|=1$ となるのはきわめて特殊な状況であり，母数 $\theta$ は自己相関関数より一意的に識別される．この過程は厳密には反転可能ではないが，モデルに基づく予測は完全に可能であるので，$|\theta|>1$ となる過程と同様の理由ですてることはできない．

反転可能性の概念は高次の MA 過程に拡張することができる．反転可能性の必要条件は随伴 MA 多項式(1.26 b)によって表わされ，

$$\theta(L)=0 \tag{3.8}$$

の根が単位円周外にあることとなる．

## 2.4 混合過程

一般の自己回帰移動平均過程はすでに(1.25)で定義されているが，これを $y_t\sim \mathrm{ARMA}(p,q)$ と表わすことにする．$\mathrm{AR}(p)$ モデルや $\mathrm{MA}(q)$ モデルは特別な場合であるので，それぞれ $\mathrm{ARMA}(p,0)$ と $\mathrm{ARMA}(0,q)$ と表わすことができる．ここで $p$ と $q$ がともにゼロでない場合のことを混合過程と呼ぶことにしよう．

### 定常性と反転可能性

ある混合過程が定常過程であるか否かはその自己回帰部分のみに依存する．その条件は自己回帰部分に対応した随伴多項式(1.27)において $\phi(L)=0$ の根が単位円外に存在することとなる．同様にして反転条件は $\mathrm{MA}(q)$ 過程と同一



であって(3.8)の根が単位円外に存在することとなる．

定常性の条件がモデルの AR 部分のみに依存していることは，混合過程の無限次移動平均表現を求めれば明らかとなる．例えば $\mathrm{ARMA}(1,1)$ 過程

$$y_t=\phi y_{t-1}+\epsilon_t+\theta\epsilon_{t-1},\qquad t=1,\cdots,T \tag{4.1}$$

は最も単純な混合過程であるが，$y_t$ の過去の値を繰り返し代入することによって(2.3)の表現を得ることができよう．別の方法を用いてこのことを示すと，まず(4.1)を

$$(1-\phi L)y_t=(1+\theta L)\epsilon_t \tag{4.2}$$

と表わし，両辺を $(1-\phi L)$ で割ると

$$y_t=\frac{\epsilon_t}{1-\phi L}+\frac{\theta\epsilon_{t-1}}{1-\phi L} \tag{4.3}$$

となる．ここで $L$ が条件 $|L|\leqq 1$ を満たすと仮定し，$1/(1-\phi L)$ を条件 $|\phi|<1$ より 1，$\phi L$，$(\phi L)^2,\cdots$ の無限幾何級数和とみなすと，(4.3)は

$$\begin{aligned} y_t&=\sum_{j=0}^{\infty}(\phi L)^j\epsilon_t+\theta\sum_{j=0}^{\infty}(\phi L)^j\epsilon_{t-1}\\ &=\sum_{j=0}^{\infty}\phi^j\epsilon_{t-j}+\theta\sum_{j=0}^{\infty}\phi^j\epsilon_{t-j-1}\\ &=\epsilon_t+\sum_{j=1}^{\infty}(\theta\phi^{j-1}+\phi^j)\epsilon_{t-j} \end{aligned} \tag{4.4}$$

と表わされる．$\theta=0$ のときには(2.5)に一致している．さらに条件 $|\phi|<1$ の下で(4.4)のウェイトは急速に減衰するので，確率過程の分散と自己共分散は共に存在する．

$p>1$ のときには $\phi(L)$ を因数分解し $\phi(L)^{-1}\theta(L)$ の各因子を展開して無限次 MA 表現を得ることができる．より簡単な方法としては，

$$\theta(L)=\phi(L)\psi(L) \tag{4.5}$$

において，$L$ のベキ係数を比較する方法が考えられる．ただし，ここで $\theta(L)$ と $\phi(L)$ は(1.26)で定義される多項式である．無限次 MA 多項式 $\psi(L)$ はすでに(2.17)で与えられており，(4.5)は AR(2)の場合に MA 係数を求めた(2.20)の一般化にすぎない．係数について整理すると(4.5)より

$$\begin{aligned} &\psi_0=1\\ &\psi_j=\theta_j+\sum_{i=1}^{\min(j,p)}\phi_i\psi_{j-i},\qquad j=1,\cdots,q \end{aligned} \tag{4.6}$$

$$\psi_j = \sum_{i=1}^{\min(j,p)} \phi_i \psi_{j-i}, \qquad j > q$$

を得る．ここで $j \geq \max(p, q+1)$ に対して $\psi_j$ は差分方程式(4.6)と $p$ 個の初期値によって決定される．

**例1** AR(2)モデルにおいてMA係数は，初期値を $\psi_0 = 1$，$\psi_1 = \phi_1$ とした $j \geq 2$ に関する差分方程式(2.22)より導かれることがすでに示されている．

**例2** ARMA(1,1)モデル(4.1)において $\psi_j$ は差分方程式

(4.7) $$\psi_j = \phi\psi_{j-1}, \qquad j \geq 2$$

および初期値

(4.8) $$\psi_1 = \theta + \phi\psi_0 = \theta + \phi$$

によって計算される．

同様の方法によって任意の反転可能なARMA過程の無限次自己回帰表現を得ることもできる．

### 自己共分散関数と自己相関関数

混合モデルの時間領域における性質は，ARモデルとMAモデルの両方の性質を組合わせたものとなる．そこでこれをARMA(1,1)モデルによって例示してみよう．$y_{t-\tau}$ を(4.1)に乗じて期待値をとると，

(4.9) $$\gamma(\tau) = \phi\gamma(\tau-1) + E(\epsilon_t y_{t-\tau}) + \theta E(\epsilon_{t-1} y_{t-\tau}), \quad \tau = 0, 1, 2, \cdots$$

となる．$\tau > 1$ に対しては右端の二項はゼロである．$\tau = 1$ に対しては右辺第二項はゼロとなるが右端の項は

$$E(\epsilon_{t-1} y_{t-1}) = E[\epsilon_{t-1}(\phi y_{t-2} + \epsilon_{t-1} + \theta\epsilon_{t-2})] = \sigma^2$$

となる．$\tau = 0$ に対してはこれら二項ともにゼロとはならずにそれぞれ

$$E(\epsilon_t y_t) = \sigma^2,$$

$$\begin{aligned} E(\epsilon_{t-1} y_t) &= E[\epsilon_{t-1}(\phi y_{t-1} + \epsilon_t + \theta\epsilon_{t-1})] \\ &= \phi E(\epsilon_{t-1} y_{t-1}) + \theta\sigma^2 \\ &= \phi\sigma^2 + \theta\sigma^2 \end{aligned}$$

によって与えられる．したがって自己共分散関数は，

(4.10 a) $$\gamma(0) = \phi\gamma(1) + \sigma^2 + \theta\phi\sigma^2 + \theta^2\sigma^2$$

(4.10 b) $$\gamma(1) = \phi\gamma(0) + \theta\sigma^2$$

(4.10 c) $$\gamma(\tau) = \phi\gamma(\tau-1), \qquad \tau = 2, 3, \cdots$$



となる．(4.10 b)の $\gamma(1)$ へ(4.10 a)を代入すると

(4.11) $$\gamma(0) = \frac{1 + \theta^2 + 2\phi\theta}{1 - \phi^2}\sigma^2$$

となるので，

(4.12) $$\gamma(1) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 - \phi^2}\sigma^2$$

が得られることとなる．さらに(4.12)と(4.10 c)を(4.11)で割れば，

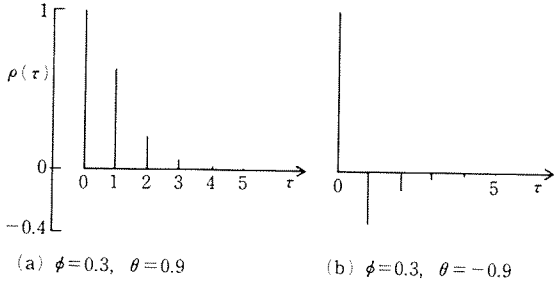
(4.13 a) $$\rho(1) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + \theta^2 + 2\phi\theta}$$

(4.13 b) $$\rho(\tau) = \phi\rho(\tau-1), \qquad \tau = 2, 3, \cdots$$

が得られる．

この自己相関関数を調べるとその動きは，$\tau > 1$ においては1次差分方程式(4.13 b)に支配されることがわかる．すなわち自己相関は指数的に減衰し，$\phi$ が負ならば振動する．このことはAR(1)の場合と全く同一である．しかしながらAR(1)モデルの差分方程式の初期値が $\rho(0) = 1$ のみで与えられるのに対し，ARMA(1,1)モデルでは初期値は上の $\rho(1)$ で与えられる．さらに(4.13 a)が示しているように，$\rho(1)$ は母数 $\phi$ と $\theta$ に依存していて $\rho(1)$ の符号は $\phi + \theta$ の符号によって定まる．

自己相関関数が示すパターンの例として $\phi = 0.3$，$\theta = 0.9$ とおいてみよう．このとき $\rho(1) = 0.649$ となり自己相関関数は図2.5(a)に示されている．移動平均母数を $\theta = -0.9$ に変えると $\rho(1) = -0.345$ となる．このとき自己相関は再び指数的に減衰するが初期値は正値ではなく負値となることが図2.5(b)によって示されている．ARMA(1,1)モデルの場合には自己相関は $\phi$ と $\theta$ に依存して全部で六つの異なったパターンを描きうることが知られている．




(a) $\phi = 0.3$, $\theta = 0.9$　　(b) $\phi = 0.3$, $\theta = -0.9$

**図2.5**　ARMA(1,1)過程に対する自己相関関数

さらに高次の ARMA モデルの性質についても同様にして調べることができる．一般的な自己相関関数のパターンは最初の $q$ 個の自己相関が移動平均母数と自己回帰母数の両方に依存する．そして高次の自己相関は $\rho(q)$，$\rho(q-1)$，…，$\rho(q-p+1)$ を初期値とした $p$ 次差分方程式(2.37)の解として与えられる．

### 共通因子

もし(1.27)において AR 多項式と MA 多項式に共通根があるならば，共通因子(common factor)が存在するという．この場合には $p$ と $q$ を削減することによって同一の性質を持つモデルを得ることができるので，モデルは母数過剰になっている．共通因子を見つけ出すことは重要である．というのは，もしそれが存在するとモデルは識別可能ではなくなり計算上の問題も生じることになる．(たとえば Harvey (1981 a) 3.6 節参照)．

**例 3** ARMA(2, 1)モデルを

$$y_t = 0.2y_{t-1} + 0.15y_{t-2} + \epsilon_t + 0.3\epsilon_{t-1} \tag{4.14}$$

とすると，AR 多項式は因数分解できて

$$(1-0.2L-0.15L^2) = (1-0.5L)(1+0.3L)$$

となる．そこでモデルを書きなおすと

$$y_t = \phi^{-1}(L)\,\theta(L)\,\epsilon_t = \frac{(1+0.3L)}{(1-0.5L)(1+0.3L)}\,\epsilon_t$$

となり，(4.14)は AR(1)モデル

$$y_t = 0.5y_{t-1} + \epsilon_t \tag{4.15}$$

と同一の MA 表現を持つことは明らかである．すなわち(4.15)と(4.14)は同一の自己共分散を持ち，(4.14)は過剰母数となっていることがわかる．

## 2.5 集計問題と欠足観測値問題*

時系列の集計には二種類の形がある．第一は観測系列が二つあるいはそれ以上の時系列の和の場合である．このときには個々の系列は観測不能な要素(unobserved components)といわれ，もしそれぞれの系列が定常 ARMA 過程に従うときに集計された系列が，ARMA 過程に従うか否かが自然な疑問として生じる．

集計化の第二の形は，時間に関する集計である．利用可能な時系列データは



モデルで用いられる時間単位よりも粗い単位を持っているとしよう．たとえば，モデルの定式化における「自然な」時間間隔は四半期であるが年データのみが利用可能な場合である．ここで問題の変数が粗国民生産のようなフローの場合には時間集計(temporal aggregation)が始めから行なわれているとみなせよう．もし集計前の系列が ARMA 過程によって生成されていたとすると，集計系列が ARMA 過程に従うかどうか問題となるが，もしそうならばその確率過程の性質が次の問題となろう．さてもし問題の変数が貨幣供給量のようなストックである場合には問題の性格が多少異なるであろう．ストック変数は本質的に連続的であるが，それを特定の時間単位でモデル化することも無意味ではない．そのとき，もし変数の計測がモデルの要求よりも頻度が少なければ欠足観測値(missing observation)問題が生じる．この状況はフロー変数の場合と類似しているが，集計された情報は利用できない．それにもかかわらず集計問題は欠足観測値問題として扱われることが多い．

以下ではまず集計が行なわれたり，観測値が欠落した場合の観測量の性質に関する結果を示す．個々の要素系列を抽出したり，集計前の観測値を推定する方法のいくつかは 4 章で簡単に言及されるであろう．

### 和の定理

定常で互いに独立で平均ゼロの二つの時系列 $y_t$ と $x_t$ を考えよう．もし $y_t \sim$ ARMA$(p_1, q_1)$ かつ $x_t \sim$ ARMA$(p_2, q_2)$ であるならば，和

$$z_t = y_t + x_t \tag{5.1}$$

は ARMA$(p_3, q_3)$ に従う．ただし $p_3 \leqq p_1 + p_2$ かつ $q_3 \leqq \max(p_1+q_2, p_2+q_1)$ とする．証明は Granger=Morris (1976) を参照されたい．ここで特別な場合には，より簡単化されるために $p_3$ と $q_3$ について不等号条件が必要となる．

**例 1** $x_t$ と $y_t$ は互いに独立の AR(1)過程であって，しかも共通の AR 母数をもつとする．すなわち，

$$y_t = \phi y_{t-1} + \epsilon_t \tag{5.2 a}$$

$$x_t = \phi x_{t-1} + \eta_t \tag{5.2 b}$$

であって，$\epsilon_t$ と $\eta_t$ は互いに独立なホワイト・ノイズとしよう．このとき

$$y_t + x_t = \phi(y_{t-1} + x_{t-1}) + \epsilon_t + \eta_t \tag{5.3}$$

となるので

(5.4) $$z_t = \phi z_{t-1} + \xi_t$$

と表わされる．ここで $\xi_t = \epsilon_t + \eta_t$ はホワイト・ノイズであるから $z_t$ は ARMA(2, 1) モデルではなく AR(1) モデルとなる．

一般的結果から特殊な場合に関する結果をいくつか導いてみよう．もし母数が例 1 のように偶然的に縮約されない場合には次の結果が知られている．

(5.5 a) $$\mathrm{AR}(p) + \text{white noise} = \mathrm{ARMA}(p, p)$$

(5.5 b) $$\mathrm{AR}(p_1) + \mathrm{AR}(p_2) = \mathrm{ARMA}[p_1 + p_2, \max(p_1, p_2)]$$

(5.5 c) $$\mathrm{MA}(q_1) + \mathrm{MA}(q_2) = \mathrm{ARMA}[0, \max(q_1, q_2)]$$

(5.5 d) $$\mathrm{ARMA}(p, q) + \text{white noise} = \mathrm{ARMA}[p, \max(p, q)]$$

(5.5 e) $$\mathrm{AR}(p) + \mathrm{MA}(q) = \mathrm{ARMA}(p, p+q)$$

ここで注意すべきことは，AR 過程にホワイト・ノイズを加えると混合過程になるが，MA 過程にホワイト・ノイズを加えると同じ次数の MA 過程となることである．この結果は比較的容易に導くことができる．

**例 2** MA(1) 過程 (1.1) にそれとは独立なホワイト・ノイズ系列 $\eta_t$ を加えると，

(5.6) $$z_t = \epsilon_t + \theta\epsilon_{t-1} + \eta_t$$

となる．ここで $\mathrm{Var}(\epsilon_t) = \sigma_1^2$，$\mathrm{Var}(\eta_t) = \sigma_2^2$ とすれば，$z_t$ の自己相関関数は

(5.7) $$\rho(\tau) = \begin{cases} \sigma_1^2\theta/[\sigma_1^2(1+\theta^2) + \sigma_2^2], & \tau = 1 \\ 0 & \tau \geq 2 \end{cases}$$

となるが，これは MA(1) 過程の自己相関関数に一致する．ゆえに $z_t$ は MA(1) 過程であって，その母数は (1.15) を解くことによって見つけることができる．

### 時間集計と欠足観測値

さて $y_t$ は ARMA$(p, q)$ 過程に従うフロー変数であるが，各 $m$ 期間ごとに集計されて観測されているものとしよう．すなわち観測量は，

(5.8) $$y_t^* = \sum_{j=0}^{m-1} y_{t-j}, \qquad t = m, 2m, 3m, \cdots$$

によって定義されるが，これは ARMA$(p, q^*)$ 過程に従っていることがわかる．ここで $q^*$ は

$$[(p+1)(m-1) + q]/m$$

に最も近い整数である．この結果については Brewer (1973) 及び Amemiya=Wu (1972) を参照されたい．



**例 3** $y_t$ は AR(1) 過程に従っているが 2 時点ごとに観測されているとしよう．すなわち，

(5.9) $$\begin{aligned} y_t &= \phi(\phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \end{aligned}$$

(5.10) $$y_{t-1} = \phi^2 y_{t-3} + \phi\epsilon_{t-2} + \epsilon_{t-1}$$

であるが (5.9) と (5.10) を加えることによって集計時系列の生成過程についての表現

(5.11) $y_t + y_{t-1} = \phi^2(y_{t-2} + y_{t-3}) + \epsilon_t + (1+\phi)\epsilon_{t-1} + \phi\epsilon_{t-2}$, $t = 2, 4, 6, \cdots$

を得る．これを書き直すと

(5.12) $$y_\tau^* = \phi^2 y_{\tau-1}^* + u_\tau, \qquad \tau = 2t, \quad t = 1, 2, 3, \cdots$$

となるが，$u_\tau$ は MA(1) 過程と同一の自己相関関数を持っている．ゆえに $y_\tau^*$ は ARMA(1, 1) 過程に従っている．

一般に $m$ が増大するにつれて集計時系列は $q \geq p+1$ ならば ARMA$(p, p+1)$ に近づき，$q < p+1$ ならば ARMA$(p, p)$ に近づく．他方，連続する観測量の自己相関は弱くなってゆき集計された時系列はよりホワイト・ノイズに近くなってゆく．

ストック変数に関しては同様の結果を Brewer (1973) が得ている．もし確率過程がある時刻で ARMA$(p, q)$ であって各 $m$ 番目ごとの時刻で観測が行なわれるとすると，観測量は ARMA$(p, q^*)$ に従う．ただし $q^*$ は

$$[p(m-1) + q]/m$$

に最も近い整数である．標本間隔 $m$ が大きくなると観測値を生成する確率過程は，$q \geq p$ ならば ARMA$(p, p)$，$q < p$ ならば ARMA$(p, p-1)$ に近づくことが知られている．

**例 4** $y_t \sim$ AR(1) 過程であるが $y_t$ は 2 時点ごとに観測されるとすると，観測量はやはり AR(1) 過程に従う．このことは $y_{t-1}$ の代入によって

(5.13) $$y_t = \phi^2 y_{t-2} + \{\epsilon_t + \phi\epsilon_{t-1}\}$$

となることから直ちにわかる．というのは右辺のカッコ内の項は時刻 $t = 2, 4, 6, \cdots$ においてはホワイト・ノイズになっているからである．

## 2.6 多変量時系列

一変量時系列の分析に用いた諸概念は多変量時系列に拡張される．多変量時系列ではまず時刻 $t = 1, 2, \cdots, T$ において観測される $N \times 1$ ベクトル $y_t =$

$(y_{1t}, y_{2t}, \cdots, y_{Nt})'$ の要素間の同時的変動に焦点があてられる．そして次に多変量(あるいはベクトル値)ARMA 過程によるモデル分析が試みられることになる．

## 定 常 性

一般性を失うことなく $E(y_t)=0$ と仮定しよう．すると $y_t$ と $y_{t-\tau}$ の自己共分散は $E(y_t y'_{t-\tau})$ によって定義される．この行列の $(i,j)$ 要素は $E(y_{it}y_{jt-\tau})$ であり，これらの要素がすべて $t$ から独立であるとき，ベクトル値過程 $y_t$ は共分散定常となる．この条件は個々の構成系列が共分散定常であることよりもはるかに強い条件である．後者は単に自己共分散行列の対角要素が $t$ と独立となる制約であって非対角要素については何の制約も課さない．共分散行列の非対角要素は特定の時差での異なる時系列同士の相互共分散(cross covariances)を表わし，これらが時刻 $t$ から独立なときにのみ二つの定常過程は同時定常(jointly stationary)となる．

弱定常ベクトル値過程に対し $y_t$ と $y_{t-\tau}$ の共分散は，

$$E(y_t y'_{t-\tau}) = \Gamma(\tau), \qquad \tau = 0, 1, 2, \cdots \tag{6.1}$$

と表わされ，$\Gamma(\tau)$ は一般に時差 $\tau$ の自己共分散行列(autocovariance matrix)と呼ばれる．共分散(covariance)行列は $\tau=0$ とおくことによって得られるが，これは対称行列となる．

一変量時系列の場合には一般に行列 $\Gamma(\tau)$ と $\Gamma(-\tau)$ とは等しくない．ただし $\Gamma(\tau)$ の $(i,j)$ 要素をとって $t=t^*+\tau$ とすると

$$\begin{aligned}\gamma_{ij}(\tau) &= E(y_{it}y_{j,t-\tau}) = E(y_{i,t^*+\tau}y_{j,t^*}) \\ &= E(y_{jt^*}y_{i,t^*+\tau}) = \gamma_{ji}(-\tau)\end{aligned} \tag{6.2}$$

となるので，$\Gamma(\tau)$ の $(i,j)$ 要素は $\Gamma(-\tau)$ の $(j,i)$ 要素に等しい．すなわち定常ベクトル値確率過程の自己共分散行列については，

$$\Gamma(\tau) = \Gamma'(-\tau), \qquad \tau = 0, 1, 2, \cdots \tag{6.3}$$

が成立する．

自己共分散行列を基準化することによって対応する自己相関行列(autocorrelation matrix)が得られる．$\Gamma(\tau)$ の $(i,j)$ 要素を $y_i$ と $y_j$ の分散の平方根で割れば自己相関あるいは相互相関関数が得られる．行列の用語を用いると時差 $\tau$ の自己相関行列 $P(\tau)$ は，

$$P(\tau) = D_0^{-1}\Gamma(\tau)D_0^{-1}, \qquad \tau = 0, \pm1, \pm2, \cdots \tag{6.4}$$



によって定義される．ここで $D_0^2 = \mathrm{diag}\{\gamma_{11}(0), \cdots, \gamma_{NN}(0)\}$ であり，$P(\tau)$ の $(i,j)$ 要素は時差 $\tau$ の $y_i$ と $y_j$ の相互相関関数

$$\rho_{ij}(\tau) = \gamma_{ij}(\tau)/\sqrt{\gamma_{ii}(0)\,\gamma_{jj}(0)} \tag{6.5}$$

を与える．

## 相互相関関数

ベクトル $y_t$ の異なる要素間の情報は自己相関行列の非対角要素に含まれている．とくに2変量時系列に対してはこの行列の構造より一つの非対角要素のみを考えに入れればよく，それを $\tau$ の正値と負値に対してプロットしたものが相互相関関数を与える．

**例1**　$\epsilon_t$ と $x_t$ を平均ゼロで分散がそれぞれ $\sigma^2$，$\sigma_x^2$ の互いに独立なホワイト・ノイズ過程とする．さらに第三の変数 $y_t$ を方程式，

$$y_t = \beta x_{t-1} + \epsilon_t \tag{6.6}$$

によって定義する．このときベクトル値確率過程 $(y_t, x_t)'$ は共分散定常であり相互相関関数は，

$$\rho_{yx}(\tau) = \begin{cases} \beta\sigma_x/(\beta^2\sigma_x^2+\sigma^2)^{1/2}, & \tau = 1 \\ 0 & \tau \neq 1 \end{cases} \tag{6.7}$$

で与えられる．

**例2**　例1において $x_t$ が定常 AR(1)過程

$$x_t = \phi x_{t-1} + \eta_t, \qquad |\phi| < 1 \tag{6.8}$$

に従っているとする．ただしここで $\eta_t$ は平均ゼロ，分散 $\sigma_\eta^2$ のホワイト・ノイズであるとしよう．このときには相互相関関数は多少複雑になる．$\epsilon_t$ と $\eta_t$ は互いに独立であることから $x_{t-\tau}$ を(6.6)に乗じて期待値をとると，

$$\gamma_{yx}(\tau) = \beta\gamma_x(\tau-1), \qquad \tau = 0, \pm1, \pm2, \cdots \tag{6.9}$$

を得る．したがって，

$$\rho_{yx}(\tau) = \beta\phi^{|\tau-1|}\sigma_x/\sigma_y, \qquad \tau = 0, \pm1, \cdots \tag{6.10}$$

となる．ただし $\sigma_x^2 = \sigma_\eta^2/(1-\phi^2)$ かつ $\sigma_y^2 = \beta^2\sigma_x^2 + \sigma^2$ である．二変数 $x_t$ と $y_t$ の関係は(6.7)では明瞭であるのに対し $x_t$ に系列相関が存在すると不明瞭になる．というのは(6.10)が示しているように相互相関関数は $x_t$ と $y_t$ の関係のみならず $x_t$ を生成する確率過程にも依存してくるからである．

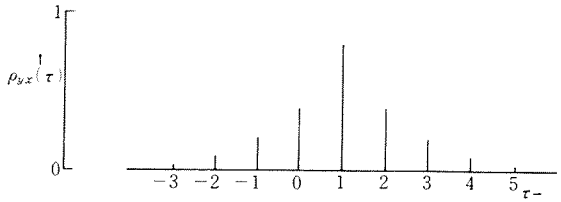



図 2.6　(6.10)で与えられる相互相関関数

## 2.7　ベクトル値自己回帰移動平均過程

非確率要素を含まないベクトル値確率過程 $y_t$ は常に無限次移動平均過程

$$y_t = \sum_{j=0}^{\infty} \Psi_j \epsilon_{t-j} \tag{7.1}$$

と表わされる．ここで $\Psi_j$ は $N \times N$ の母数行列で $\Psi_0 = I$ とする．$\epsilon_t$ は平均ゼロのホワイト・ノイズ変量から成る $N \times 1$ ベクトルである．$\epsilon_t$ はしばしば多変量ホワイト・ノイズ過程(multivariate white noise process)と呼ばれ，その性質は

$$E(\epsilon_t) = 0$$

かつ

$$E(\epsilon_t \epsilon_s') = \begin{cases} \Omega, & t = s \\ 0, & t \neq s \end{cases} \tag{7.2}$$

である．ここで $\Omega$ は $N \times N$ の共分散行列である．すなわち $\epsilon_t$ の要素は同時点において相関を持っている．

ところで移動平均表現(7.1)は(1.23)の直接的な一般化である．一変量の場合と同様にして「ケチの原理」に基づけば $y_t$ を混合過程

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t + \cdots + \Theta_q \epsilon_{t-q} \tag{7.3}$$

によってモデル分析を行なうことができよう．ここで $\Phi_i$ は $N \times N$ の AR 母数行列で $\Theta_j$ は $N \times N$ の MA 母数行列である．このモデルは多変量(あるいはベクトル)ARMA$(p, q)$過程として知られている．

### 自己共分散関数と自己相関関数

ベクトル値 ARMA 過程の時間領域での性質は一変量の場合に用いられた方法によって同様に導かれる．



**例 1**　ベクトル値 MA(1)過程

$$y_t = \epsilon_t + \Theta \epsilon_{t-1}, \tag{7.4}$$

は母数 $\Theta$ の値にかかわらず定常となる．$E(y_t) = 0$ であるので，$y_t$ の共分散行列は

$$\Gamma(0) = E(y_t y_t') = \Omega + \Theta \Omega \Theta' \tag{7.5}$$

であり，さらに自己共分散行列は

$$\Gamma(\tau) = E(y_t y_{t-\tau}') = \begin{cases} \Theta\Omega, & \tau = 1 \\ \Omega\Theta', & \tau = -1 \\ 0, & |\tau| \geq 2. \end{cases} \tag{7.6}$$

によって与えられる．ここで $\Gamma'(-1) = \Gamma(1)$ であり，$|\tau| \geq 2$ に対して自己共分散は切断されることに注意しよう．これは一変量 MA(1)過程と平行した性質である．

**例 2**　ベクトル値 1 次自己回帰過程

$$y_t = \Phi y_{t-1} + \epsilon_t \tag{7.7}$$

が定常となる条件を導いておこう．もし定常であるとすると，自己共分散行列は $y_{t-\tau}'$ を $y_t$ の後から乗じて期待値をとると，

$$E(y_t y_{t-\tau}') = \Phi E(y_{t-1} y_{t-\tau}') + E(\epsilon_t y_{t-\tau}') \tag{7.8}$$

となることに注意する．ここで $\tau \geq 1$ に対して最後の項はゼロ・ベクトルであるから，

$$\Gamma(\tau) = \Phi \Gamma(\tau - 1), \qquad \tau \geq 1 \tag{7.9}$$

となる．これはベクトル値差分方程式であって，その解は

$$\Gamma(\tau) = \Phi^{\tau} \Gamma(0), \qquad \tau \geq 0 \tag{7.10}$$

で与えられる．(2.29)参照．$\tau = 0$ のとき(7.8)の最後の項は，

$$E[\epsilon_t (\Phi y_{t-1} + \epsilon_t)'] = E(\epsilon_t \epsilon_t') = \Omega$$

となる．さらに(6.3)と(7.10)を用いると(7.8)の右辺第一項は，

$$\Phi \Gamma(-1) = \Phi \Gamma'(1) = \Phi \Gamma(0) \Phi'$$

となるので，

$$\Gamma(0) = \Phi \Gamma(0) \Phi' + \Omega \tag{7.11}$$

を得る．この方程式から $\Gamma(0)$ を評価する方法はいくつか考えられよう．一つには繰返し法によって $\Gamma(0)$ を適当な初期値から逐次的に近似値を求める方法があろう．あるいは恒等式

$$\mathrm{vec}\{\Gamma(0)\} = [I - \Phi \otimes \Phi]^{-1} \mathrm{vec}\{\Omega\} \tag{7.12}$$

を直接解くことも考えられよう．ここで記号 $\otimes$ はクロネッカー積を表わし，vec(•)記号は $n \times m$ 行列の列を順番に上から下へ並べて作られた $nm$ ベクトルを意味している．

最後に(6.6)と(6.8)によって定義されたモデルでは，$(y_t, x_t)'$ はベクトル値 AR(1) 過程において，

$$\Phi = \begin{bmatrix} 0 & \beta \\ 0 & \phi \end{bmatrix}$$

となっていることに注意しよう．この場合自己共分散と相互共分散を直接求めることもできる．むろん以上で述べた一般公式によっても求めることができるが，それは読者の練習問題としよう．

### 多変量自己回帰移動平均過程における定常性と反転可能性*

ベクトル値 AR(1) 過程(7.7)において $y_t$ のラグ値の代入を繰り返すと，

$$y_t = \sum_{j=0}^{J-1} \Phi^j \epsilon_{t-j} + \Phi^J y_{t-J} \tag{7.13}$$

が得られる．一変量の場合と同様にして $J$ を十分大きくすると $\Phi^J$ の要素がゼロへ収束し $y_t$ は $y_{t-J}$ に依存しなくなるであろう．その条件を導くために，まず正則な $N \times N$ 行列 $Q$ による行列 $\Phi$ の対角化

$$\Phi = Q \Lambda Q^{-1} \tag{7.14}$$

を考えよう．ここで $\Lambda = \mathrm{diag}\{\lambda_1, \cdots, \lambda_N\}$ であって，各要素 $\lambda_1, \cdots, \lambda_N$ は行列式方程式 (determinantal equation)

$$|\Phi - \lambda I| = 0 \tag{7.15}$$

の根であり，互いに相異なると仮定している．したがって，例えば $N = 2$ で $\Phi$ の $(i, j)$ 要素が $\phi_{ij}$ のときには，行列式方程式は

$$\begin{vmatrix} \phi_{11} - \lambda & \phi_{12} \\ \phi_{21} & \phi_{22} - \lambda \end{vmatrix} = 0 \tag{7.16}$$

となって二根は 2 次関数

$$\lambda^2 - (\phi_{11} + \phi_{22})\lambda + (\phi_{11}\phi_{22} - \phi_{12}\phi_{21}) = 0 \tag{7.17}$$

を解くことによって求められ，根は実根あるいは共役複素根になる．

ところで(7.14)の対角化は $N$ 個の根の総てが相異なるときにのみ可能である．この仮定の下で，

$$\Phi^J = Q\Lambda Q^{-1} \cdot Q\Lambda Q^{-1} \cdots Q\Lambda Q^{-1} = Q\Lambda^J Q^{-1}$$

となる．ここで $Q$ は $J$ に依存せず，しかも $\Delta^J = \mathrm{diag}\{\lambda_1^J, \cdots, \lambda_N^J\}$ であるから，行列式方程式(7.15)の総ての根の絶対値が 1 より小さいという条件の下で，

$$\lim_{J \to \infty} \Phi^J = \lim_{J \to \infty} Q\Lambda^J Q^{-1} = 0 \tag{7.18}$$



となる．すなわち(7.7)が定常であるためには条件

$$|\lambda_i| < 1, \qquad i = 1, \cdots, N \tag{7.19}$$

が必要である．

さて以上の結果は $p$ 次自己回帰要素を含むベクトル値 ARMA 過程に一般化される．もし，

$$|\lambda^p I - \lambda^{p-1}\Phi_1 - \cdots - \Phi_p| = 0 \tag{7.20 a}$$

のすべての根の絶対値が 1 より小さければモデルは定常となる．同様にして，もし

$$|\lambda^q I + \lambda^{q-1}\Theta_1 + \cdots + \Theta_q| = 0 \tag{7.20 b}$$

のすべての根の絶対値が 1 より小さければ反転可能となる．ここで(7.3)のラグ作用素による表現を，

$$\Phi_p(L)\, y_t = \Theta_q(L)\, \epsilon_t \tag{7.21}$$

としよう．ただし，

$$\Phi_p(L) = I - \Phi_1 L - \cdots - \Phi_p L^p \tag{7.22 a}$$

でかつ

$$\Theta_q(L) = I + \Theta_1 L + \cdots + \Theta_q L^q \tag{7.22 b}$$

とおくことにする．このとき定常性の条件と反転可能性の条件は行列式多項式 $|\Phi_p(L)|$ と $|\Theta_q(L)|$ に関する条件として表わされ，一変量の場合と同様に方程式の根が単位円外に存在することになる．

### 識　別　性

一変量 ARMA モデルの識別性 (identifiability) に関する共通因子の意味についてはすでに本章 **2.4** 節で述べた．同様の問題は多変量モデルにおいても考えられるが識別性の条件はずっと複雑になる．この問題に関しては Hannan (1969) や Granger=Newbold (1977, pp. 219-224) などを参照されたい．

### 多変量モデルと一変量モデル

任意の定常ベクトル値 ARMA 過程は行列多項式 $\Phi(L)$ を逆転することによって(7.1)の形に書け，さらに

$$y_t = \Phi^{-1}(L)\, \Theta(L)\, \epsilon_t \tag{7.23}$$

と表わすことができる．$\Phi^{-1}(L)$ を $L$ のベキ級数として展開すれば，

$$\Phi^{-1}(L) = \Phi^{\dagger}(L) / |\Phi(L)| \tag{7.24}$$

となるが，ここで $\Phi^{\dagger}(L)$ は $\Phi(L)$ の余因子行列である．さらにこれを(7.23)に代入し整理することによって，自己回帰的最終型(autoregressive final form)

(7.25) $$|\Phi(L)|y_t = \Phi^{\dagger}(L)\,\Theta(L)\,\epsilon_t$$

を得る．

ここで(7.25)において $N$ 個の各方程式の右辺は $p+q$ 次の MA 過程の線型結合から成っている．ところがこのような線型結合は $(N-1)p+q$ 次の一変量 MA 過程として表現することができる(例えば本章 **2.5** 節の例 2 参照)．ゆえに(7.15)の方程式は

(7.26) $$\phi_i^*(L)\,y_t = \theta_i^*(L)\,\eta_{it}, \qquad i = 1,\cdots,N$$

と書くことができる．ここで $\eta_{it}$ はホワイト・ノイズであり $\theta_i^*(L)$ は $(N-1)p+q$ 次のラグ作用素についての MA 多項式であり，$\phi_i^*(L) = |\Phi(L)|$ は $Np$ 次の AR 多項式である．この結果によれば $y_t$ を構成する各変数はそれぞれ一変量 ARMA 過程になる．もちろんこれらの変数を同時に考慮すれば，(7.26)によって各一変量 ARMA モデルの撹乱項は互いに相関を持ち，しかも自己回帰部分は同一になる．もっとも，ある方程式の $\phi_i^*(L)$ と $\theta_i^*(L)$ に共通因子があることによって同一の MA 要素が観測されないことも実際にはあり得る．

## 練習問題

**1.** 次にあげる確率過程は定常であろうか．

(i) $y_t = \alpha\cos\lambda t + \beta\sin\lambda t$,

ただし $\alpha \sim N(0,\sigma^2)$，$\beta \sim N(0,\sigma^2)$ で，かつ $\alpha$ と $\beta$ は互いに独立とする．

(ii) $y_t = y_{t-1} + \epsilon_t$, $\quad \epsilon_t \sim NID(0,\sigma^2)$

(iii) $y_t = \epsilon_{1t} + t\epsilon_{2t}$

ただし

(a) $\epsilon_{it} \sim NID(\mu_i,\sigma_i^2)$, $\quad \mu_i \neq 0, i = 1,2$

(b) $\epsilon_{it} \sim NID(0,\sigma_i^2)$, $\quad i = 1,2$

(iv) $y_t - y_{t-1} + 0.5y_{t-2} = \epsilon_t$.

**2.** 次の確率過程の自己相関関数を定めよ．

$$y_t = 14 + \epsilon_t + 0.4\,\epsilon_{t-1} - 0.2\,\epsilon_{t-2}.$$

**3.** もし AR(2)過程が定常ならば，$\rho^2(1) < [\rho(2)+1]/2$ となることを示せ．

**4.** 次にあげる AR(2)過程において，最初から 5 個の自己相関を計算せよ．

(a) $\phi_1 = 0.6, \phi_2 = -0.2$

(b) $\phi_1 = -0.6, \phi_2 = 0.2$



**5.** (a) 次にあげる確率過程の自己相関関数を図示せよ．

$$y_t = -0.5y_{t-1} + \epsilon_t - 0.8\,\epsilon_{t-1}.$$

(b) その無限次 MA 表現における係数 $\psi_j$ を求めよ．

**6.** モデル

$$y_t = -0.2y_{t-1} + 0.48y_{t-2} + \epsilon_t + 0.6\,\epsilon_{t-1} - 0.16\,\epsilon_{t-2}$$

は母数過剰であるか．

**7.** $x_t$ が MA(1)過程に従うとして，(6.6)における $y_t$ と $x_t$ の相互相関関数の形を決めよ．

**8.** ベクトル MA(1)過程

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix} + \begin{bmatrix} 0.5 & -0.3 \\ 0.7 & 0.6 \end{bmatrix} \begin{bmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \end{bmatrix}$$

において自己相関行列をすべての有限の時差について評価せよ．

**9.** 次にあげる非定常過程を考えよう．

$$y_{1t} = y_{1,t-1} + \epsilon_{1,t} + \theta\epsilon_{1,t-1}, \qquad |\theta| < 1$$

$$y_{2t} = \phi y_{1t} + \epsilon_{2,t}, \qquad |\phi| < 1$$

このとき 1 次の差分ベクトル $(\Delta y_{1t}, \Delta y_{2t})'$ は厳密には反転可能でないベクトル値 MA(1)過程であることを示せ．

# 3章
# 周波数領域

## 3.1 はじめに

周波数領域の分析においては，時系列における種々の周期的な要素の全体への寄与にその興味が注がれる．それらの要素を必ずしも規則的循環(cycle)とみなす必要はない．実際に厳密に規則的循環は，経済時系列においては稀なことである．そこで周期的な要素という場合には，通常それが意味することは，ある周波数を中心とした循環的変動のことである．

やや曖昧に定義された不規則循環という概念について何らかの洞察を得るために，まず規則的循環がいかに分析されるかを調べてみる必要がある．これを**3.2**節において行なうが，そこではフーリエ解析を用いて時系列から周期的トレンド要素をいかにして取り出すかを示す．**3.3**節ではフーリエ解析とスペクトラム解析の関係を確立する．第一段階は，トレンド・モデルにおける係数を簡単に再解釈することによって，周期的トレンド要素を含むモデルを定常過程へ転換することである．第二段階はモデルに無限個の周期的トレンド要素を持ち込むことである．いったんこれがなされれば，前章で説明されたARMA過程の性質についての周波数領域における解釈を与えることが可能となる．

線型フィルターの性質は**3.5**節で吟味される．周波数領域の分析によって，時間領域では必ずしもはっきりしないことについていくつかの重要な洞察が得られる．さらに同節において相互スペクトラム解析の基礎概念が導入される．相互スペクトラムは周波数領域において二つの時系列の関係を性格づけるものである．

スペクトラムの推定では，自己共分散関数を推定する場合には生じない種々の問題が発生する．その問題は **3.6** 節において吟味される．

### スペクトラム

2章(1.23)で記述されるような，どのような非決定的過程のパワー・スペクトラム，またはより簡潔にスペクトラムは，以下のような連続関数によって定義される．

$$f(\lambda)=(2\pi)^{-1}\left[\gamma(0)+2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos\lambda\tau\right]. \tag{1.1}$$

ここで，$\lambda$ はラディアン(角速度)ではかった周波数で $[-\pi,\pi]$ の範囲内の値をとる．しかしながら $f(\lambda)$ は0に対して対称であるので，パワー・スペクトラムの総ての情報は $[0,\pi]$ の範囲内に含まれている．

もし $y_t$ がホワイト・ノイズであれば，$\tau\neq 0$ のとき，$\gamma(\tau)=0$ であり，以下を得る．

$$f(\lambda)=\sigma^2/2\pi. \tag{1.2}$$

ここで $\sigma^2$ は $y_t$ の分散である．ゆえに図3.1に示されているごとくこのスペクトラムは水平なものとなる．したがって，この確率過程はそれぞれの周波数がすべて等しい加重を持つような無限個の周期的要素から成立しているとみなすことができるであろう．実際これが周波数領域におけるホワイト・ノイズの定義である．

上記の(1.2)を見ることにより，$[-\pi,\pi]$ の範囲上でのパワー・スペクトラムの面積はちょうど分散 $\sigma^2$ と等しいということが見てとれる．より一般的には次式が成立する．

$$\int_{-\pi}^{\pi} f(\lambda)\,d\lambda=\gamma(0). \tag{1.3}$$

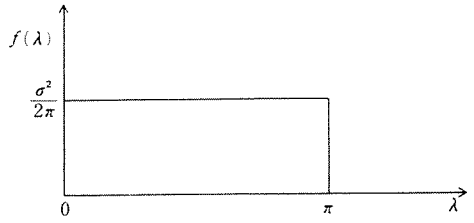



**図3.1** ホワイト・ノイズ過程のパワー・スペクトラム



この結果は **3.3** 節において証明されるが，ここではむしろその解釈が重要である．この(1.3)の表現は，線型過程のパワー・スペクトラムが確率過程の分散を周波数に関して分解したものとみなし得るということを示している．

パワー・スペクトラムはしばしば $\gamma(0)$ で割って標準化される．これは(1.1)における自己共分散を対応する自己相関で置き換えることと同じである．この標準化された関数はスペクトラム密度(spectral density)として知られている．もっとも同じ呼び名がパワー・スペクトラムを意味することに用いられることもある．この混乱に加えてさらに多くの文献では，パワー・スペクトラムとスペクトラム密度を前述の2倍として定義している．そのようにしておくと，$\lambda$ が $[0,\pi]$ 上のみで定義されていても，パワー・スペクトラムおよびスペクトラム密度の面積がそれぞれちょうど $\gamma(0)$ および1になるのである．ただしこの表現法は本書では用いず，ここでの議論は(1.1)に基づく定義によって行なわれる．パワー・スペクトラムとスペクトラム密度の呼び名は交替的に用いられるが，それは実際上の観点から標準化はそれほど重要でないからである．

パワー・スペクトラムがゼロを中心として対称であるときに，$[-\pi,\pi]$ の範囲に定義することについては，たぶん多少の説明が必要であろう．これまでの議論では系列 $y_t$ が実数であるということを暗黙に仮定していた．もし $y_t$ が複素数であれば，パワー・スペクトラムは以下のような複素フーリエ変換によって定義されなければならない．

$$f(\lambda)=(2\pi)^{-1}\sum_{\tau=-\infty}^{\infty}\gamma(\tau)e^{-i\lambda\tau}. \tag{1.4}$$

ここで $\lambda$ は $[-\pi,\pi]$ の範囲内である．もし $y_t$ が実数であるときには，(1.4)は(1.1)に帰着する．しかし一般的には $y_t$ が実数である必要はなく，その場合には $f(\lambda)$ はゼロのまわりで対称であるとは限らない．本書の内容は実数確率過程に制限されているが，他の文献との比較可能性のために，より一般的な定義を用いることが有用と思われる．さらに複素フーリエ変換(1.4)は理論的結果を導くときにはしばしば，より扱いやすいのである．このことが $[-\pi,\pi]$ の範囲で定義されたパワー・スペクトラムを採用するさらに重要な理由である．

**例1**　2章(1.1)で定義されたMA(1)モデルでは $\gamma(0)=(1+\theta^2)\sigma^2$，$\gamma(1)=\theta\sigma^2$，そして $\tau\geq 2$ のとき $\gamma(\tau)=0$ である．2章(1.14)を参照．これらを(1.1)に代入すると以下を得る．

(1.5) $$f(\lambda) = (\sigma^2/2\pi)(1+\theta^2+2\theta\cos\lambda)$$

もし $\theta = 0.5$ であれば以下のようになる．

(1.6) $$f(\lambda) = \sigma^2(5+4\cos\lambda)/8\pi.$$

関数(1.6)が図 3.2 に描かれている．この場合 $y_t$ は現在および遅れのある撹乱項の加重平均なので，この系列はホワイト・ノイズよりもやや滑らかなものとなる．いい換えれば，それはホワイト・ノイズよりもゆっくりと変化する．時間領域においては，それは正の一階の自己共分散に反映されている．他方，周波数領域においては，同様の性質がより低周波数領域における高パワー・スペクトラムの値として現われている．もしこの確率過程の $\theta$ が負であれば，スペクトラムは高周波数においてより大きくなるであろう．そしてそれは時系列がホワイト・ノイズよりも，より不規則に変動することを意味する．

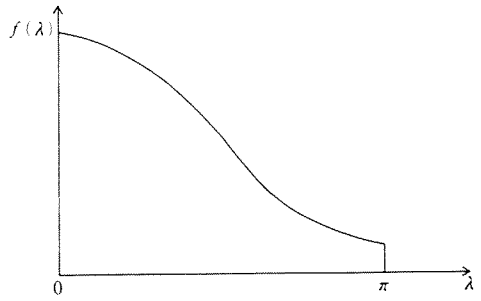



**図 3.2** MA(1)過程パワー・スペクトラム：$\theta = 0.3$ の場合

以上の例から明らかなように，パワー・スペクトラムと自己共分散関数は互いに競合的というより補完的なものである．両者は時系列の性質を異なった方法で際立たせる．ここでパワー・スペクトラムは自己共分散関数に表われていないような情報は全く含んでいないことに留意すべきである．なぜなら，それは単に自己共分散の線型結合にすぎないからである．

## 3.2 周期的トレンド

本節では周期的トレンドの概念を導入する．そしてこのトレンドを回帰分析的手法によってどのようにしてデータに当てはめるかを示す．次にその結果の自然な拡張としてフーリエ解析およびピリオドグラムの定義を導く．この問題を議論するに当っては，三角関数に関するいくつかの標準的な結果を利用する必要がある．その結果は本章末の**付論 B** にまとめられている．



### 周期的関数

余弦(コサイン)関数

(2.1) $$y = \cos x$$

を，ラディアン(角速度)で測った角度 $x$ に関して定義しよう．単一の円は $2\pi$ ラディアンであるから，$x$ がゼロから $2\pi$ に動くときに $y$ はすべての値をとることになる．そして，この形はその後も繰り返される．ゆえにどのような整数 $k$ に関しても $\cos(x+2k\pi) = \cos x$ である．正弦(サイン)関数も同様な性質を持つ．図 3.3 は $\sin x$ および $\cos x$ を $x$ に対して示しているが，余弦関数がゼロに関して対称であるということが見い出されるであろう．これは，偶関数，すなわち $\cos x = \cos(-x)$ であることの反映である．他方，正弦関数は $\sin x = -\sin(-x)$ なる奇関数である．

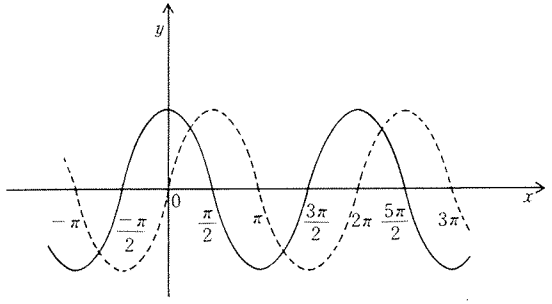



**図 3.3** サイン関数(---)とコサイン関数(——)

ラディアンで測った周波数または角周波数(angular frequency)と呼ばれる母数 $\lambda$ を定義することにより，変数 $y$ を時間に関する周期的関数として表わすことができるであろう．そのとき変数 $x$ は $\lambda t$ によって置き換えられる．$\lambda$ に対して異なった値を割り当てることによって，関数は水平軸 $t$ に沿って拡張したり縮小したりする．$y$ が一連の値を完全にとるのに要する時間，すなわち循環の周期(period)は $2\pi/\lambda$ に等しい．ゆえに 5 単位時間毎に繰り返される三角関数は $2\pi/5$ の周波数を持つであろう．

振幅 $\rho$ と呼ばれる母数を余弦または正弦関数に乗ずることによって，循環的関数の一層の伸縮性を導入することができる．最後に水平軸に関して関数の

位置を固定するという問題が残っているが，これは位相と呼ばれる，やはりラディアンで測られた角度 $\theta$ を導入することによって行なわれる．そこで(2.1)の表現は次のように一般化される．

$$y = \rho\cos(\lambda t - \theta) = \rho\cos\lambda(t-\xi). \tag{2.2}$$

ここで母数 $\xi = \theta/\lambda$ は時間領域に関しての移動を表わす．

**例 1**　5単位時間の周期を持った余弦関数を考えよう．ここでわれわれはそのピークを $t = 0,\ 5,\ 10,\ \cdots$ から $t = 2,\ 7,\ 12,\ \cdots$ に移動させたいと思っている．すると $\xi = 2$ である．したがって $\theta = \lambda\xi = (2\pi/5)\cdot 2 = 4\pi/5 = 2.513$ ラディアンである．この動きが実際に前方への動きなのか，後方への動きなのかはやや問題を含んでいる．なぜなら，$\xi = -3$ とおくことによっても同じ結果を作ることができるからである．これは，$\theta = (2\pi/5)(-3) = -3.770$ ラディアンを意味するであろう．この位相をめぐる不明瞭さは，相互スペクトラム解析においていくつかの重要な問題を提起する．

三角関数の水平方向のシフトは異なった方法によっても導入することができる．正弦または余弦関数に位相を導入することに代えて，正弦関数と余弦関数の組み合わせを考えよう．すなわち(2.2)は以下のようにも表わせる．

$$y = \alpha\cos\lambda t + \beta\sin\lambda t \tag{2.3}$$

ここで $\alpha = \rho\cdot\cos\theta$ そして $\beta = \rho\cdot\sin\theta$ であると考える．すると上の例は以下のように表わせる．

$$y = \cos(\lambda t - 2.513) = -0.809\cos\lambda t + 0.588\sin\lambda t$$

この変換はしばしば逆方向にも行なわれるが，その表現の間には次のような関係が成立している．

$$\rho^2 = \alpha^2 + \beta^2 \tag{2.4 a}$$

$$\theta = \tan^{-1}(\beta/\alpha). \tag{2.4 b}$$

### フーリエ解析

前項の終りにおいて，三角関数の項はその一つを他のものに重ね合わせて，一つのまとまった循環の形を作り出すことができるということを学んだ．今ここで，$T$ 個の点 $y_1,\ \cdots,\ y_T$ が与えられているとする．そしてこれらのすべての点を通る関数 $y(T)$ を作ることを考えよう．この目的を達成する一つの方法は，$y(T)$ を $T$ 個の三角関数の線型関数とすることである．

もし $y_1,\ \cdots,\ y_T$ が時系列の観測値の組としての通常の解釈が許されるのであれば，$T$ 個の三角関数を当てはめる最も簡単な方法はフーリエ解析によるものである．そこでまず $n$ を次のように定義しよう．

$$n = \begin{cases} T/2, & T\text{が偶数ならば} \\ (T-1)/2, & T\text{が奇数ならば}. \end{cases} \tag{2.5}$$

さらに周波数 $\lambda_j$ を以下のように定義する．

$$\lambda_j = 2\pi j/T, \qquad j = 1, \cdots, n. \tag{2.6}$$

適当な関数 $y(t)$ を作るための第一段階は，それぞれの周波数において $\cos\lambda_j t$ と $\sin\lambda_j t$ なる一組の三角関数の項を取り出すことである．もし $T$ が偶数であるならば，$j = n$ に対応する項はただ一つしかない．なぜなら $t$ が整数であるときには，常に $\sin\pi t = 0$ だからである．したがって $T$ が奇数あるいは偶数であるかにかかわらず，ちょうど $T-1$ 個の三角関数の項を与えることになり，一つの定数項を加えることによって $T$ 項の完全な組合せが得られる．

もし $T$ が偶数であるならば，時系列 $y_t$ のフーリエ表現(Fourier representation)は以下のように与えられる．

$$y_t = T^{-1/2}a_0 + \sqrt{2/T}\sum_{j=1}^{n-1}(a_j\cos\lambda_j t + b_j\sin\lambda_j t) + T^{-1/2}a_n(-1)^t. \tag{2.7}$$

もし $T$ が奇数であるならば，(2.7)における最終項は，$\cos\pi t$ に基づくものなので不必要となり，総和は $j = 1,\ 2,\ \cdots,\ n-1$ でなく $j = 1,\ \cdots,\ n$ についてとられることになるが，その他はすべて偶数の場合と同様である．もしこれらの規則を心に入れておけば，奇数の観測値の場合の結果は，$T$ が偶数であるときにもたらされる結果から直接的に得ることができる．いずれの場合においても，周波数が $\pi$ ラディアンよりも大きくなるようには定義されないことに注意しよう．なお $\lambda = \pi$ なる周波数はナイキスト(Nyquist)周波数と呼ばれる．そしてこれはちょうど2単位時間の周期に対応している．これより短い周期を含もうとすることは，その処理に不明確な要素を導入することになる．なぜなら，そのような循環についての意味のある情報は観測値から得られないからである．上式において $T^{-1/2}$ を含む係数の導入はやや恣意的なものに見えよう．この形を用いる理由は後で明らかになるであろう．

上記(2.6)のように周波数を定義するという選択は本質的には計算上の理由からきている．それによって章末付論 **B** に示されている三角関数の直交性の関係を使うことを可能にし，以下で示すようにフーリエ係数，$a_j$ および $b_j$ の簡単な表現が生み出されるのである．まず(2.7)の両辺に $\sin(2\pi it/T)$ を掛け，

$t$ について 1 から $T$ まで足し合わせ，そして(B.3b)および(B.3c)を使うことにより以下の式を得る．

$$(2.8)\qquad \sum_{t=1}^{T} y_t \sin\frac{2\pi i}{T}t = (T/2)^{1/2} b_i, \qquad i = 1,\cdots,n-1.$$

これを整理し，添字 $i$ と $j$ と書き換えることによって以下の表現を得る．

$$(2.9\,\mathrm{a})\quad b_j = (2/T)^{1/2}\sum_{t=1}^{T} y_t \sin(2\pi j/T)\,t, \qquad j = 1,\cdots,n-1.$$

同様に(2.7)の両辺に $\cos(2\pi jt/T)$ を掛けることにより次の表現を得る．

$$(2.9\,\mathrm{b})\quad a_j = (2/T)^{1/2}\sum_{t=1}^{T} y_t \cos(2\pi j/T)\,t, \qquad j = 1,\cdots,n-1,$$

$$(2.9\,\mathrm{c})\qquad a_0 = T^{-1/2}\sum_{t=1}^{T} y_t,$$

かつ

$$(2.9\,\mathrm{d})\qquad a_n = T^{-1/2}\sum_{t=1}^{T} y_t(-1)^t.$$

**例 2**　四つの観測値があり，それらが $y_1 = 1$，$y_2 = 3$，$y_3 = 5$ そして $y_4 = 7$ の値をとると考えよう．フーリエ表現は以下のように与えられる．

$$(2.10)\qquad y_t = T^{-1/2}a_0 + (2/T)^{1/2}a_1\cos\frac{2\pi}{4}t + (2/T)^{1/2}b_1\sin\frac{2\pi}{4}t + T^{-1/2}a_2(-1)^t, \qquad t = 1,\cdots,4.$$

そこで(2.9)からフーリエ係数を計算することにより次のような結果を得る．

$$(2.11)\qquad y_t = 4 + 2\cos\frac{\pi}{2}t - 2\sin\frac{\pi}{2}t + (-1)^t$$

$$= 4 + \sqrt{8}\cos\left(\frac{\pi}{2}t - \frac{1}{4}\pi\right) + (-1)^t.$$

このような状況の下で周期的な表現が意味があるかどうかは，観測された周期的変動が $T$ 個の観測値の以後の時系列においても再び繰り返すものであるかどうかに依存しているが，このことは次式によって表現される．

$$y_{t+4k} = y_t, \qquad t = 1,2,3,4 \quad k = 1,2,\cdots.$$

このとき $y_t$ は図 3.4 に描かれているような形をとる．しかし現実にはこのような形はあまり起りそうでない．そしてこのようなデータに出会ったときのよりもっともらしい処理の仕方は，単に線型のトレンドをあてはめることであるかもしれない．



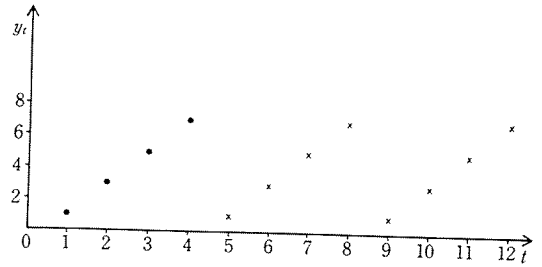



**図 3.4**　例 2 における原系列(•)とフーリエ表現による値(×)

上記(2.7)における正弦項と余弦項の一つの組はまとめられた単一の項によって表わすことができるので，それぞれの周波数の相対的な重要度を単一の量で測ることができる．すなわち，(2.4 a)の表現は次なる量の使用を示唆する．

$$(2.12)\qquad p_j = a_j^2 + b_j^2, \qquad j = 1,\cdots,n-1.$$

他方観測値の偏差の二乗和は，以下のように書くことによってこの量に関して分解することができる．

$$(2.13)\qquad \sum_{t=1}^{T}(y_t - \bar{y})^2 = \sum_{j=1}^{n-1} p_j + p_n,$$

ここで $p_n = a_n^2$ である．これがパーシバルの定理(Parseval's theorem)である．その証明は直交性の関係に基づき非常に自然に与えることができる．

それぞれの $p_j$ を対応する周期 $2\pi/\lambda_j$ に対してプロットした図はピリオドグラム(periodogram)と呼ばれる．しかしながら一般的には，$p_j$ を $\lambda_j$ に対してプロットする方がより便利であろう．これはスペクトログラム(spectrogram)と呼ばれるものである．しかしこの場合にもピリオドグラムという名称が用いられることがあり，実際には $\lambda_j$ に対する $p_j$ をプロットに対してはピリオドグラムと呼ぶ名称が最もよく用いられている．本書においてもこの意味でピリオドグラムを用いることにする．

ピリオドグラムの縦軸はしばしば次のように複素数項に関して定義される．

$$(2.14)\qquad p_j = (2/T)\left|\sum_{t=1}^{T} y_t e^{-i\lambda_j t}\right|^2, \qquad j = 1,\cdots,n-1.$$

しかしながら付論(A.5)の定義式を用いて，指数項を展開することによって虚数部分は消去され，以下のような実数部分が残る．

$$(2.15)\qquad p_j=(2/T)\left[\left(\sum_{t=1}^{T}y_t\cos\lambda_j t\right)^2+\left(\sum_{t=1}^{T}y_t\sin\lambda_j t\right)^2\right].$$

この表現は(2.9)における表現から直ちに導かれるものである．

方程式(2.14)および(2.15)は，$y_t$ を $(y_t-\bar{y})$ に置き換えることによって書き直すことができる．すなわち(2.14)の別の表現は次のように与えられる．

$$(2.16)\qquad p_j=\frac{2}{T}\left|\sum_{t=1}^{T}(y_t-\bar{y})\,e^{-i\lambda_j t}\right|^2,\qquad j=1,\cdots,n-1.$$

このことは以下のような関係式に着目し，

$$\sum_{t=1}^{T}(y_t-\bar{y})\,e^{-i\lambda_j t}=\sum_{t=1}^{T}y_t e^{-i\lambda_j t}-\bar{y}\sum_{t=1}^{T}e^{-i\lambda_j t}.$$

この第二項が付論(B.5)における定義式よりゼロになるということによって直ちに証明される．なお $j=0$ の場合は，(2.16)を用いると $p_0=0$ となるが，これは当然である．なぜなら $p_0$ は平均を反映したピリオドグラムと解釈されるからであり，時系列の周期的な動きを学ぶという観点からは，$p_0$ は興味の対象とはならないのである．

### 循環的トレンド・モデル

フーリエ表現(2.7)の価値は，もとのモデルに関係づけることによってはじめて評価することができよう．そこで次のような循環的トレンド・モデルを考えよう．

$$(2.17)\qquad y_t=T^{-1/2}\alpha_0+\sqrt{2/T}\sum_{j=1}^{n-1}(\alpha_j\cos\lambda_j t+\beta_j\sin\lambda_j t)$$
$$+T^{-1/2}\alpha_n(-1)^t+\epsilon_t,\qquad t=1,\cdots,T.$$

ここで $\epsilon_t\sim NID\ (0,\sigma^2)$ である．この枠組の中でフーリエ係数(2.9)は(2.17)における母数の最小二乗推定量であることが見てとれる．ここで $\gamma=(\alpha_0,\alpha_1,\beta_1,\alpha_2,\cdots,\alpha_{T/2})'$ なる $T\times1$ ベクトルを定義することにより，(2.17)は行列記号で次のように書くことができる．

$$(2.18)\qquad y=Z\gamma+\epsilon.$$

ここで $Z$ は(2.19)に示されている $T\times T$ のフーリエ行列(Fourier matrix)である．付論Bの(B.3)より $Z$ は直交行列である．したがって $\gamma$ の最小二乗推定量は以下のように与えられる．

$$(2.20)\qquad c=(Z'Z)^{-1}Z'y=Z'y.$$



(2.19)

$$Z=\begin{bmatrix}
\frac{1}{\sqrt{T}} & \sqrt{\frac{2}{T}}\cdot\cos\frac{2\pi}{T} & \sqrt{\frac{2}{T}}\cdot\sin\frac{2\pi}{T} & \sqrt{\frac{2}{T}}\cdot\cos\frac{4\pi}{T} & \cdots & -\sqrt{\frac{1}{T}}\\
\frac{1}{\sqrt{T}} & \sqrt{\frac{2}{T}}\cdot\cos\frac{2\pi}{T}2 & \sqrt{\frac{2}{T}}\cdot\sin\frac{2\pi}{T}2 & \sqrt{\frac{2}{T}}\cdot\cos\frac{4\pi}{T}2 & \cdots & \sqrt{\frac{1}{T}}\\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\
\frac{1}{\sqrt{T}} & \sqrt{\frac{2}{T}}\cdot\cos\frac{2\pi}{T}(T-1) & \sqrt{\frac{2}{T}}\cdot\sin\frac{2\pi}{T}(T-1) & \sqrt{\frac{2}{T}}\cdot\cos\frac{4\pi}{T}(T-1) & \cdots & -\sqrt{\frac{1}{T}}\\
\frac{1}{\sqrt{T}} & \sqrt{\frac{2}{T}} & 0 & \sqrt{\frac{2}{T}} & \cdots & \sqrt{\frac{1}{T}}
\end{bmatrix}$$

これは(2.9)の結果そのものであり，回帰モデルの説明変数間の直交性によってそれぞれの最小二乗推定量は互いに独立である．したがって他の母数の推定値に影響を与えることなしに，(2.17)から説明変数を取り除くことが可能である．

フーリエ係数の分布の性質は，標準的な最小二乗理論から直ちに導かれる．すなわち

$$(2.21)\qquad \mathrm{Var}\,(c)\sim\sigma^2(Z'Z)^{-1}=\sigma^2 I$$

であり，$c-\gamma$ の要素は平均ゼロ分散 $\sigma^2$ を持ち独立的に正規分布する．さらにもし $\alpha_j=\beta_j=0$ のときは以下の結果が成り立つ．

$$(2.22)\qquad \sigma^{-2}p_j\sim\chi^2_2,\qquad j=1,\cdots,n-1.$$

このモデルを $T$ 個の説明変数のすべてを用いて当てはめない限り，$\sigma^2$ の推定値は残差から計算することができる．最小二乗法における推定量 $s^2$ に関する通常の形が適用可能であり，統計量 $s^{-2}p_j/2$ が帰無仮説の下でF分布に従うということを用いて検定を行なうことができる．ここで帰無仮説を解釈すると，モデルにおいて周波数 $\lambda_j$ の要素が無いということである．

**例3**　Anderson (1971, pp. 106-7) は3年間にわたる月次データの時系列における季節性の形をとらえるために，周期的トレンドに誤差を加えたモデルを用いた．月次データに対する基本的な季節周波数は $\lambda=2\pi/12$ であり，これは12ヵ月周期に対応している．完全な季節性の形を取り出すための残りの周波数は $\lambda_j=2\pi_j/12$, $j=2,\cdots,6$ であるが，これらは調和(harmonic)周波数として知られている．そのピリオド

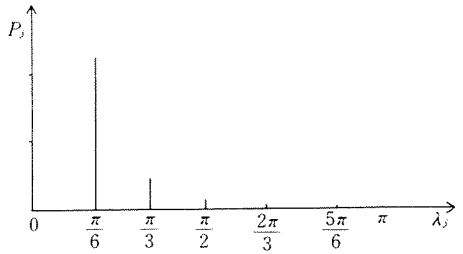



**図 3.5** 例 2.3 における季節的周波数のピリオドグラム

グラムが図 3.5 に示されている．12 ヵ月と 6 ヵ月の期間に対応している二つの低い周波数の要素が季節的な形のほとんどを説明している．残りの要素は比較的に重要でなく，実際に 5% 水準で統計的に有意ではない．

ところで，与えられた季節変動パターンは定数項と 11 の「ダミー変数」によってもまた記述されるかもしれない．しかしながら上の例が示すように，季節的パターンはしばしば限られた数の三角関数の項によって非常にうまく記述することができる．他方，全体の 11 個よりいくつかのダミー変数を除くことは，はるかに不満足な結果を招きがちである．

これまでの議論においては循環的トレンド・モデルに含まれるすべての周波数は，(2.6) によって定義された周波数の集合の中に含まれていると想定されてきた．しかしこれが常に成立すべきであるという理由はない．そこでより一般的には，循環的トレンドに誤差を加えたモデルは次のように書かれるであろう．

$$y_t = \alpha_0 + \sum_j (\alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t) + \epsilon_t. \tag{2.23}$$

ここで $\lambda_j$ はすべての $j$ に対して $0 < \lambda_j \leqq \pi$ であればよい．このモデルにおける母数もまた最小二乗法によって推定することができる．しかしすべての周波数 $\lambda_j$ が (2.6) で定義された集合に含まれない限り，説明変数はもはや直交しないことになる．

## 3.3 確率過程のスペクトラム表現

次のようなモデルを考えよう．

$$y_t = \alpha \cos \lambda t + \beta \sin \lambda t, \qquad t = 1, \cdots, T \tag{3.1}$$



ここで $\alpha$ と $\beta$ は一定の母数であり，$\lambda$ は 0 と $\pi$ の間のある特定の値とする．これは (2.17) で与えられた循環的トレンド・モデルにおける決定論的要素の一つであるとみなすことができよう．このとき $y_t$ は明らかに定常ではない．なぜなら $E(y_t)$ は $t$ の関数だからである．しかしこのような状況は $\alpha$ および $\beta$ を二つの確率変数 (random variable) $u$，$v$ に置き換えることによって修正することができる．ここで $u$ と $v$ は互いに相関がなく平均ゼロ分散 $\sigma^2$ を持つものとしよう．すると次のような結果が得られる．

$$E(y_t) = E(u) \cos \lambda t + E(v) \sin \lambda t = 0 \tag{3.2 a}$$

$$\begin{aligned} \mathrm{Var}(y_t) &= E[(u \cos \lambda t + v \sin \lambda t)^2] \\ &= \sigma^2 (\cos^2 \lambda t + \sin^2 \lambda t) = \sigma^2 \end{aligned} \tag{3.2 b}$$

$$\begin{aligned} \gamma(\tau) = E(y_t y_{t-\tau}) &= \sigma^2 [\cos \lambda t \cos \lambda (t-\tau) + \sin \lambda t \sin \lambda (t-\tau)] \\ &= \sigma^2 \cos \lambda \tau, \qquad \tau = 1, 2, 3, \cdots \end{aligned} \tag{3.2 c}$$

ここで $u$ と $v$ は確率変数であるが特定の実現された時系列においては固定された (fixed) ものと見なされる．したがって，モデルは定常的であるが，やはり決定論的なものである．そして $u$ と $v$ を厳密に決定するのにただ二つの観測値のみが必要である．そしていったん $u$ と $v$ の値が決定されれば，時系列における残りの点は平均二乗誤差ゼロで予測することができる．したがって実際には，非定常モデル (3.1) と対応する定常モデル (3.2) との唯一の違いはモデルの母数の解釈にある．

以上の定常的モデル (3.2) を一つ以上の周波数に対応する項を含むように拡張することができる．もし $J$ 個の周波数 $\lambda_1, \cdots, \lambda_J$ を含めるならば，それは次のようになる．

$$y_t = \sum_{j=1}^{J} (u_j \cos \lambda_j t + v_j \sin \lambda_j t). \tag{3.3}$$

ここでもし

$$E(u_j) = E(v_j) = 0, \qquad j = 1, \cdots, J \tag{3.4 a}$$

$$\mathrm{Var}(u_j) = \mathrm{Var}(v_j) = \sigma_j^2, \qquad j = 1, \cdots, J \tag{3.4 b}$$

$$E(u_i u_j) = E(v_i v_j) = 0, \qquad i \neq j \tag{3.4 c}$$

であり，かつすべての $i, j$ について

$$E(u_i v_j) = 0 \tag{3.4 d}$$

とすれば，時系列の平均はゼロであって，

$$\gamma(0)=\sum_{j=1}^{J}\sigma_j^2, \tag{3.5}$$

かつ

$$\gamma(\tau)=\sum_{j=1}^{J}\sigma_j^2\cos\lambda\tau \tag{3.6}$$

となる．上の(3.5)における表現は，この確率過程の分散が全ての$u_j$と$v_j$の分散の和であるとみなすことができることを意味する．もし(3.3)における周波数の集合が(2.6)によって定義された周波数の集合に対応していれば，この結果は(2.13)と同様なものである．しかしここでの重要な相違点は，(2.13)においてはそれは標本内(in the sample)における分散であり，それぞれ三角関数項の振幅の二乗$\rho_j^2$の推定値$p_j$に分解されていたことである．

ここで2章(1.12)によって定義されたホワイト・ノイズ過程を考えよう．この確率過程は完全な独立系列なので，その周波数要素のすべては等しい重要性を持っている．ここでこの文脈において「すべて」というのは一体何を意味するのかという問題が直ちに生じる．ある特定の周波数からなる集合を考えて，それらが確率過程の分散に対してすべて等しい貢献をしていると主張することは意味を持たない．なぜなら0から$\pi$の範囲には無限個の周波数が存在するからである．そこで(3.3)をこのような状況を含めるように拡張するためには，0から$\pi$の範囲において無限個の三角関数の項を含める必要がある．このことは$J$を無限大にし，和記号を積分に置き換えることによって成し遂げることができよう．そこで$u_j$および$v_j$を$u(\lambda)$および$v(\lambda)$なる$\lambda$の連続(continuous)関数によって置き換えれば，確率過程$y_t$のスペクトラム表現(spectral representation)は次のように与えられる．

$$y_t=\int_0^{\pi}u(\lambda)\cos\lambda t\,d\lambda+\int_0^{\pi}v(\lambda)\sin\lambda t\,d\lambda. \tag{3.7}$$

これはしばしばクラメール表現(Cramér's representation)と呼ばれる．

関数$u(\lambda)$および$v(\lambda)$は確率変数であり，離散型表現(3.3)における$u_j$および$v_j$と同様の性質を持っている．任意の小さな区間$d\lambda$を考え，$[0,\pi]$の範囲における任意の区間$d\lambda$において次式が成立すると仮定する．

$$E[u(\lambda)\,d\lambda]=E[v(\lambda)\,d\lambda]=0. \tag{3.8 a}$$

もしこの仮定が成立すれば，(3.7)に対して期待作用素を適用することによって$E(y_t)=0$を得る．同様にして，



$$\mathrm{Var}[u(\lambda)\,d\lambda]=\mathrm{Var}[v(\lambda)\,d\lambda]=2f(\lambda)\,d\lambda \tag{3.8 b}$$

と仮定する．ここで$f(\lambda)$は連続関数である．さらに

$$E[u(\lambda_1)\,d\lambda_1 u(\lambda_2)\,d\lambda_2]=E[v(\lambda_1)\,d\lambda_1 v(\lambda_2)\,d\lambda_2]=0 \tag{3.8 c}$$

と仮定する．ここで$d\lambda_1$，$d\lambda_2$は$\lambda_1$および$\lambda_2$をそれぞれ中心とする重ならない区間である．そして(3.8 c)は$[0,\pi]$の範囲のすべての重ならない区間において成立するとしよう．最後に$d\lambda_1=d\lambda_2$であるような場合を含めてすべての範囲において

$$E[u(\lambda)\,d\lambda_1 v(\lambda)\,d\lambda_2]=0 \tag{3.8 d}$$

と仮定する．すると(3.7)で与えられた確率過程$y_t$のスペクトラム表現から，

$$\gamma(\tau)=E(y_t y_{t-\tau})=E\left\{\left[\int_0^{\pi}u(\lambda)\cos\lambda t\,d\lambda+\int_0^{\pi}v(\lambda)\sin\lambda t\,d\lambda\right]\right. \tag{3.9}$$
$$\left.\times\left[\int_0^{\pi}u(\lambda)\cos\lambda(t-\tau)\,d\lambda+\int_0^{\pi}v(\lambda)\sin\lambda(t-\tau)\,d\lambda\right]\right\}$$

が与えられる．さらに仮定(3.8 d)を用いてこの表現をときほどくことによって，正弦および余弦関数を含む交差項はすべて消えることになる．残った項の期待値をとり，付論(A.1)(A.2)で与えられた三角関数の和公式を用いることによって以下の結果を得る．

$$\gamma(\tau)=\int_0^{\pi}f(\lambda)\cos\lambda\tau\,d\lambda,\qquad \tau=0,\pm1,\pm2,\cdots \tag{3.10}$$

すべての非決定論的確率過程はスペクトラム表現を持つが，この(3.10)において$\tau=0$と置くことにより次の結果を得る．

$$\gamma(0)=\int_0^{\pi}f(\lambda)\,d\lambda \tag{3.11}$$

すなわちパワー・スペクトラムが，連続的な$[0,\pi]$の範囲における分散の周波数要素への分解であると解釈できることを示している．自己共分散から$f(\lambda)$を決定するためには(3.10)のフーリエ変換が用いられるが，それは(1.1)ですでに与えられている(**付論 C** を参照のこと)．

### 複素スペクトラム表現*

複素スペクトラム表現は(3.7)よりもより簡潔であり，そして数学的にはより扱いやすい．ここで$i$を$i^2=-1$なるものと定義しよう．純粋な非決定論的確率過程は次のように表わされる．

$$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} z(\lambda)\, d\lambda. \tag{3.12}$$

ここで$-\pi \leqq \lambda \leqq \pi$における任意の小さな区間$d\lambda$に関して

$$E[z(\lambda)\, d\lambda] = 0 \tag{3.13 a}$$

であり，かつ

$$E[z(\lambda)\, d\lambda \cdot \overline{z(\lambda)\, d\lambda}] = f(\lambda)\, d\lambda \tag{3.13 b}$$

である．一方$d\lambda_1$，$d\lambda_2$をそれぞれ$\lambda_1$，$\lambda_2$を中心とする重ならない二つの区間とすると，以下が成り立つ．

$$E[z(\lambda_1)\, d\lambda_1 \cdot \overline{z(\lambda_2)\, d\lambda_2}] = 0. \tag{3.13 c}$$

これらの性質は(3.8)と同様のものである．以前と同様に直ちにすべての$t$に対して$E(y_t) = 0$が導かれ，さらに次式が成立する．

$$\gamma(\tau) = E[y_t \overline{y_{t-\tau}}] = E\left[\int_{-\pi}^{\pi} e^{i\lambda t} z(\lambda)\, d\lambda \cdot \int_{-\pi}^{\pi} e^{-i\lambda(t-\tau)} \overline{z(\lambda)\, d\lambda}\right].$$

これは(3.13 b)および(3.13 c)より以下のように表現される．

$$\gamma(\tau) = \int_{-\pi}^{\pi} e^{i\lambda\tau} f(\lambda)\, d\lambda. \tag{3.14}$$

上記(3.14)のフーリエ逆変換はすでに与えられた(1.4)である．

### ウォルドの分解定理*

以上の議論において(3.3)から(3.7)に進むにあたって，決定論的確率過程と非決定論的確率過程の間ではっきりとした区別がなされた．しかしながら，決定論的および非決定論的な二つの要因を同時に含む過程を考えることも可能である．実際にウォルドの分解定理(Wold decomposition theorem)として知られている一般的な結果によると，どのような共分散定常確率過程も次のように表わされる．

$$y_t = y_t^* + y_t^\dagger$$

ここで$y_t^*$は(3.3)の右辺で与えられている決定論的要素である．そして$y_t^\dagger$は2章(1.23)で定義されている線型確率過程である．なお$y_t^*$のスペクトラムは離散型であるので，全体としての過程の周波数領域の性質はスペクトラム密度関数ではなくスペクトラム分布関数(spectral distribution function)に関して記述しなければならない．離散スペクトラムは(3.3)に表われているそれぞれの周波数$\lambda_j$においてのジャンプに対応している．



完全に規則的周期要素は経済時系列においては異例であろう．そのような要素が存在する時系列の場合には，推定されたスペクトラムにおける鋭い峻(スパイク)として現われるであろう．不規則循環は幅広い頂(ピーク)として現われる．図3.8(b)(81ページ)で示されるように，近似的に5年周期を持つ系列のスペクトラムがこの点をうまく例示している．

## 3.4　周波数領域における自己回帰移動平均過程の性質

時間領域においてはARMA過程の性質は自己共分散，または同様なことであるが，自己相関関数によって特徴づけられる．しかしながら，自己共分散関数は確率過程のある側面ははっきり映し出すが，他の側面はぼんやりとしか映し出さない．とくに時系列の周期的性質は自己共分散ではうまく特色づけられず，周期的動きが複雑であればあるほどその欠点は明白となる．このような理由から，自己共分散関数と同様にその確率過程のパワー・スペクトラムを調べることが望ましい．ここで非決定論的確率過程によって生成された定常時系列$x_t$を考えよう．さらに別の時系列$y_t$は$x_{t+s}$，…，$x_t$，…，$x_{t-r}$の加重和によって構成されると考えよう．そこでの加重$w_{t+s}$，…，$w_t$，…，$w_{t-r}$は実数であり$\Sigma w_j^2 < \infty$という条件を満たしていると仮定する．このとき次のように表わすことができよう．

$$y_t = \sum_{j=-r}^{s} w_j x_{t-j}. \tag{4.1}$$

上式(4.1)の操作は線型時間不変(linear time invariant)フィルターと呼ばれる．ここで時間不変という意味は加重(ウェイト)$w_j$が$t$から独立であるということを意味しているにすぎない．加重をすべて加えると1になるとき，(4.1)はしばしば「移動平均」と呼ばれる．ただしこの操作と移動平均過程(process)との混乱を避けるために，その名称の使用については注意が必要である．

上式(4.1)が有用なのは，それが任意のARMA過程のパワー・スペクトラムを求めるにあたっての基礎となるからである．さらに次節において，同じ結果が周波数領域において種々のフィルターをとる操作の効果を調べることにも用いられることが示される．また時系列に対して行なわれる典型的な種々の操作，たとえば季節調整法や平滑法などについて，ある種の洞察を得ることも可能となる．

ここで関数$f_x(\lambda)$を$x_t$のスペクトラムとしよう．そして以下のようにフィルターを定義する．

$$W(\lambda)=\sum_{j=-r}^{s} w_j e^{-i\lambda j} \tag{4.2}$$

すると$y_t$のスペクトラムは次のように伝達される．

$$f_y(\lambda)=|W(\lambda)|^2 f_x(\lambda). \tag{4.3}$$

ここで作用素$|W(\lambda)|^2$は伝達関数(transfer function)として知られている．

この結果の背後に潜む理由づけについては次項で述べる．ただしさらに進む前に，その重要性を2章(1.1)のMA(1)過程に関して以下で例示しよう．上記(4.1)において$w_0=1$，$w_1=\theta$そして他のすべての加重をゼロと置くことにより，

$$W(\lambda)=1+\theta e^{-i\lambda}. \tag{4.4}$$

ゆえに

$$\begin{aligned}|W(\lambda)|^2 &= W(\lambda)\,\overline{W(\lambda)}\\ &=(1+\theta e^{-i\lambda})(1+\theta e^{i\lambda})\\ &=1+\theta^2+\theta(e^{-i\lambda}+e^{i\lambda})\\ &=1+\theta^2+2\theta\cos\lambda\end{aligned}$$

となる．この場合には$x_t$はホワイト・ノイズなので$f_x(\lambda)=\sigma^2/2\pi$である．したがって$y_t$のスペクトラムは以下のように与えられる．

$$f_y(\lambda)=(\sigma^2/2\pi)(1+\theta^2+2\theta\cos\lambda). \tag{4.5}$$

この表現は自己共分散に対してフーリエ変換を適用することによって得られた結果(1.5)の形式と同一である．上記(4.3)の重要性は，伝達関数が知られていれば$y$のスペクトラムを導出するときに，$y_t$の自己共分散関数をいちいち導く必要がない点にある．

### 伝達関数の導出*

複素数表現を用いると$x_t$のスペクトラム表現は以下のように与えられる．

$$x_t=\int_{-\pi}^{\pi} e^{i\lambda t} z_x(\lambda)\,d\lambda. \tag{4.6}$$

ここで

$$E[z_x(\lambda_1)\,d\lambda_1\cdot\overline{z_x(\lambda_2)\,d\lambda_2}]=\begin{cases} f_x(\lambda)\,d\lambda, & \lambda_1=\lambda_2\\ 0, & \lambda_1\neq\lambda_2\end{cases}$$



である．この$x_t$のスペクトラム表現を(4.1)に代入することにより，次の結果を得る．

$$\begin{aligned}y_t &= \sum_{-r}^{s} w_j \int_{-\pi}^{\pi} e^{i(t-j)\lambda} z_x(\lambda)\,d\lambda\\ &=\int_{-\pi}^{\pi} e^{i\lambda t}\left(\sum w_j e^{-i\lambda j}\right) z_x(\lambda)\,d\lambda\\ &=\int_{-\pi}^{\pi} e^{i\lambda t} W(\lambda)\, z_x(\lambda)\,d\lambda.\end{aligned} \tag{4.7}$$

この(4.7)は，$z_y(\lambda)=W(\lambda)z_x(\lambda)$によって示される確率過程$y_t$のスペクトラム表現である．すなわち$\lambda$のある与えられた値について$z_x(\lambda)$に対して周波数応答関数(frequency response function)と呼ばれる$W(\lambda)$を乗じることによって$z_y(\lambda)$が得られる．そこで定義により，

$$f_y(\lambda)\,d\lambda=E[z_y(\lambda)\,d\lambda\cdot\overline{z_y(\lambda)\,d\lambda}]$$

となる．このように$f_x(\lambda)$が与えられれば，(4.3)が直ちに導かれることがわかる．

### 自己回帰過程

定常AR(1)過程は2章(2.5)のように無限の移動平均として表わすことができる．したがって，

$$\begin{aligned}W(\lambda) &= \sum_{j=0}^{\infty} \phi^j e^{-i\lambda j}\\ &=\sum_{j=0}^{\infty}(\phi e^{-i\lambda})^j\\ &=1/(1-\phi e^{-i\lambda}).\end{aligned}$$

ここで上記の和はすべての$\lambda$において$|\phi|\leq 1$であるので可能なのである．そこで$y_t$のスペクトラムは以下のように与えられる．

$$f_y(\lambda)=\left(\frac{\sigma^2}{2\pi}\right)\left(\frac{1}{|1-\phi e^{-i\lambda}|^2}\right)=\left(\frac{\sigma^2}{2\pi}\right)\left(\frac{1}{1+\phi^2-2\phi\cos\lambda}\right). \tag{4.8}$$

ここで(4.8)を導出するもう一つの方法では，$y_t$が連続的なスペクトラムを持つという知識に基づいて，まずモデルを次のように書く．

$$\epsilon_t=y_t-\phi y_{t-1}.$$

そこでMA(1)の場合と同様に議論を進めて次のような結果を導く．

$$f_\epsilon(\lambda) = |1-\phi e^{-i\lambda}|^2 \cdot f_y(\lambda).$$

ここで $f_\epsilon(\lambda) = \sigma^2/2\pi$ と置き整理すれば(4.8)が直ちに導かれる.

この導出方法は高次の過程にも簡単に適用できる. すなわち, AR($p$)モデルについては以下のような伝達関数を考えればよいことになる.

$$W(\lambda) = 1/\left(1-\sum_{j=1}^{p}\phi_j e^{-i\lambda j}\right). \tag{4.9}$$

ところで $\phi>0$ なる AR(1)過程のスペクトラムは, 図3.2に示されている MA(1)のスペクトラムと似ている. 正の系列相関は低周波数における $f(\lambda)$ の高い値に反映されている. $\phi$ が1に近づくに従ってその系列はより「ゆっくり変化する」ようになるので, その確率過程の分散に対する低周波数の貢献がより増大する.

AR(2)過程についてはパワー・スペクトラムは以下のように与えられる.

$$f_y(\lambda) = \left(\frac{\sigma^2}{2\pi}\right)\left(\frac{1}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos\lambda-2\phi_2\cos 2\lambda}\right). \tag{4.10}$$

このスペクトラムの形は母数 $\phi_1$, $\phi_2$ がとる値に決定的に依存している. 例えば AR(1)モデルは AR(2)モデルの特殊ケースであるので, $f(\lambda)$ が単調に増加したりまたは単調に減少するようなスペクトラムも生じうる. ところで2章**2.2**節においてすでに, 固有方程式の根が複素数であるような AR(2)過程が注目された. その場合には自己相関関数は減衰する循環傾向を持ち, さらにそれが時系列におけるある種の循環的変動を示唆していることが指摘された. 実際にはこれらの動きの特徴は周波数領域においてより明瞭に表われている. 例えば図3.6は $\phi_1 = 0.7$ で $\phi_2 = -0.5$ なる AR(2)過程のパワー・スペクトラムを示している. 頂(ピーク)は $\lambda^*$ を中心とする周波数における循環傾向を示している. その動きは規則的ではないので疑似循環変動(pseudo-cyclical behavior)と呼ばれる. たとえば(3.1)のような決定論的循環は鋭い峻(スパイク)となって現れるはずである.

スペクトラムが $0<\lambda<\pi$ の範囲内でピークを持つときには, その正確な位置は(4.10)の微係数をゼロと置くことによって決定される. その式を解くことにより次のような結果を得る.

$$\lambda_0^* = \cos^{-1}[-\phi_1(1-\phi_2)/4\phi_2]. \tag{4.11}$$

図3.6における確率過程については $\lambda_0^* = \cos^{-1}(0.5250) = 1.018$ である. そしてこれは6.172の周期に対応する. これは2章(2.15)から決定される自己相関



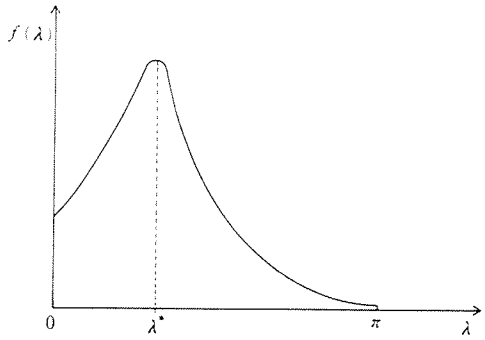



**図3.6**　AR(2)過程のパワー・スペクトラム：$\phi_1 = 0.7$, $\phi_2 = -0.5$ の場合

関数の周期と厳密には一致しないがこれら二つの値はかなり近いものである. やや関連した点であるが, モデルが複素根を持つ場合には一般的には $0<\lambda<\pi$ の範囲内でスペクトラムは必ず頂を持つ. しかしながら必ず $0<\lambda<\pi$ の範囲内に頂が存在すると主張するのは正しくない.

AR(2)過程によって示される疑似循環変動の形は, 経済モデル作成の観点からは魅力的なものである. 季節変動を除けば, 経済学においては規則的循環変動はあまり起り得ない. しかし AR(2)過程によって生成される時系列によって, しばしば実際に観察される種類の変動を簡単に描くことができるのである.

### 混合過程と随伴多項式

すでに示された方法を用いて ARMA($p$, $q$)過程のパワー・スペクトラムが以下のように与えられることを示すことは難しくはない.

$$f_y(\lambda) = \frac{\sigma^2}{2\pi}\cdot\frac{\left|1+\sum_{j=1}^{q}\theta_j e^{-i\lambda j}\right|^2}{\left|1-\sum_{j=1}^{p}\phi_j e^{-i\lambda j}\right|^2}. \tag{4.12}$$

この形は有理スペクトラム(rational spectrum)として知られている.

ここで周波数応答関数とそれに関連して2章(1.26)で定義された多項式の間には, 大変便利な関係があるということを注意しておこう. たとえば MA(1)過程においては, 随伴多項式は以下で与えられる.

$$\theta(L) = 1+\theta L.$$

一方, ラグ作用素を $e^{-i\lambda}$ で置き換えることによって周波数応答関数(4.4)を得

る．より一般的には，もしARMA$(p,q)$過程を随伴多項式によって表現すれば，そのスペクトラムは直ちに以下のように与えられる．

$$f_y(\lambda)=\frac{\sigma^2}{2\pi}\cdot\frac{|\theta(e^{-i\lambda})|^2}{|\phi(e^{-i\lambda})|^2}. \tag{4.13}$$

## 3.5 線型フィルターの性質

線型フィルターの概念は前節で導入され，(4.1)において定義された．線型フィルターまたは「移動平均」フィルターを適用し，データを加工(調整)することは，とくに経済時系列の場合よく用いられる．このようなフィルターは多くの理由によって適用される．最も単純に考えると，時系列はトレンドとその上に加えられたランダムな攪乱項から構成され，トレンドは時間に関して滑らかな関数であるとみなされる．しかし攪乱項のために観測される点のグラフはやや不規則なものとなると考えると，このことから平滑化(smoothing)という概念が導かれる．この方法は各時点におけるトレンドをその点まわりの観測値の加重平均によって表わすという手法である．

平滑化の例として次のような5期間の移動平均を考えよう．

$$y_t=\frac{1}{5}\sum_{j=-2}^{2}x_{t-j},\qquad t=3,\cdots,T-2. \tag{5.1}$$

上式(5.1)における加重(ウェイト)は等しく，合計すると1になる．攪乱項の不規則性は，5期間を通じていくらか互いに相殺されると考えられる．すなわち，攪乱項の影響は完全に取り除かれることは明らかにありえないが，弱められよう．

上記(5.1)を用いてトレンドを表わすことに関しては，$t$以外の時点における観測値を用いることから生じる歪みの問題を提起する．常識的にはより大きな加重が$x_t$とその周辺の値に与えられるべきであろう．これについては実際にトレンドが$t$の近傍において低次の多項式によって表わされると仮定することによって，いくらかの合理づけが与えられる．5期間のフィルターについて，この考え方から例えば次のような形が導かれる．

$$y_t=(-3x_{t-2}+12x_{t-1}+17x_t+12x_{t+1}-3x_{t+2})/35. \tag{5.2}$$

その詳細については，Anderson(1971，3.3節)を参照されたい．

線型フィルターを時系列に適用する一つの動機は，データにおける攪乱要因



の効果を柔らげることである．そしてもう一つは循環的要素を除くことである．たとえば年次データについての(5.1)の効果は，データから5年の循環を取り除くことであろう．しかしながら循環的要素が「除かれた」とか「平滑化された」とかいうことの本当の意味は，周波数領域におけるフィルターの性質を調べることによってはじめて明らかになるであろう．フィルターについての分析によって，フィルターの概念についてより厳密な意味が与えられると同時に，異なった方法間の比較を可能にする．5年の循環は以下の階差の操作によってもまた扱われうる．

$$y_t=x_t-x_{t-5},\qquad t=6,\cdots,T, \tag{5.3}$$

この操作が時系列に対して持つ効果を，(5.1)によってもたらされる効果と対比するのは興味あるところであろう．

季節性の取扱いは，とくに経済時系列においては重要な役割を果たす．線型フィルターの効果を調べることによって，季節調整のいくつかの実際的および潜在的な効果について多くの洞察がもたらされよう．残念ながら季節調整に関して政府の官庁当局によって採用されている方法では，しばしばある種の非線型フィルター操作を行なっている．したがって以下で議論される方法は直接的には適用することはできない．しかしながら，以下で導入される概念は，相互スペクトラム解析を用いて季節調整済み系列および原系列を比較するのに使うことができる．

### ゲインと位相

時系列に対する線型フィルターの効果は二つある．第一の効果は種々の循環的要素の相対的な重要性を変えることである．これはフィルターが適用される前と後の確率過程のスペクトラムを比べることによって明らかになる．前節においてフィルターの効果は伝達関数においてとらえることができるということを示した．そしてその伝達関数の形は，フィルターが与えられれば簡単に求めることができるのである．伝達関数の検討が本節において重要な役を演じている．もっとも，その平方根である$|W(\lambda)|$について調べる方がしばしばより便利であり，この量は$G(\lambda)$と表わされゲイン(gain)と呼ばれる．

フィルターのもう一つの性質として，時系列における時間軸上の位置の移動の大きさに関するものがある．たとえば次のような操作を考えよう．

$$y_t=x_{t-3}. \tag{5.4}$$

このやや自明なフィルター操作の効果は時系列を3期間だけ後へずらすことである．もし$x_t$が，たとえば次のような循環的過程によって生成されているのであれば，

(5.5) $$x_t = \cos \lambda t$$

フィルターをかけられた時系列は以下のようになる．

(5.6) $$y_t = \cos \lambda (t-3) = \cos (\lambda t - 3\lambda).$$

ラディアンで測った位相(phase)の移動は$\mathrm{Ph}(\lambda) = 3\lambda$で与えられる．したがって時間領域での3期間にわたる時間軸上での横への移動は，周波数領域における位相$3\lambda$となる．

$\lambda$に対して$\mathrm{Ph}(\lambda)$をとった図は位相図(phase diagram)と呼ばれる．したがって(5.4)に与えられた純粋の遅れについては，$\mathrm{Ph}(\lambda)$は原点を通る直線で，傾きはちょうど時間の遅れに等しくなる．図3.7を参照されたい．

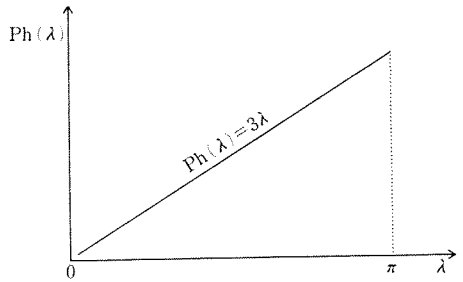



**図3.7** $y_t = x_{t-3}$の位相図

位相は周波数応答関数からも導くことができる．もし$W(\lambda) = W^*(\lambda) + iW^\dagger(\lambda)$であり，ここで$W^*(\lambda)$と$W^\dagger(\lambda)$が双方とも実数であるならば，位相は以下のように与えられる．

(5.7) $$\mathrm{Ph}(\lambda) = \tan^{-1}[-W^\dagger(\lambda)/W^*(\lambda)], \qquad 0 \leqslant \lambda \leqslant \pi.$$

確率過程のクラメール表現を参照することによって，この形についての理論的基礎を最もよく説明することができよう．それについては星印をつけた次項において説明される．さて，逆にある$\mathrm{Ph}(\lambda)$が与えられたときには，時間単位における移動量は$\mathrm{Ph}(\lambda)/\lambda$となる．前例のように，$\mathrm{Ph}(\lambda)$が$\lambda$の線型関数であるとき以外は，時間領域における変化は$\lambda$の異なった値に応じて異なる．

位相の定義は本質的な不明瞭さを含み，そのためそれはしばしば問題が生じる．なぜならある角度から完全な循環を足したり引いたりしても，それは正接(tangent)を変化させないからである．したがってもし$\mathrm{Ph}^*(\lambda)$で示されるような特定の角度が(5.7)の解として生じる場合には，整数$h$に関して$\mathrm{Ph}(\lambda) = \mathrm{Ph}^*(\lambda) \pm 2h\pi$なる関係を満たす角度もまた(5.7)の有効な解となる．すなわち(5.4)のような簡単なフィルターの場合には，一般的解は$\mathrm{Ph}(\lambda) = 3\lambda \pm 2h\pi$となる．ゆえに，たとえば$\lambda = \pi/2$なる周波数に関しては時間単位における変化は3の遅れではなく，$-3+2\pi/(\pi/2) = 1$の進みとも解釈されうる．しかしながらこの場合に$\mathrm{Ph}(\lambda) = 3\lambda$以外の解をとることにすると，それは時間領域における一定の遅れという解釈をもたらさない非線型の位相関数をもたらすことになる．



### 周波数応答関数*

周波数応答関数は一般的には複素数である．それを以下のような極座標形式で書くことが便利である．

(5.8) $$W(\lambda) = G(\lambda)\, e^{-i\mathrm{Ph}(\lambda)}$$

ここで$G(\lambda) = |W(\lambda)|$であり，$\mathrm{Ph}(\lambda)$は(5.7)で定義されている．線型フィルターを乗じることによりもたらされた確率過程のスペクトラム表現はすでに前節で導かれ，(4.7)で与えられている．それに(5.8)で与えられた$W(\lambda)$を代入すると次式が得られる．

(5.9) $$y_t = \int_{-\pi}^{\pi} e^{i\lambda t} G(\lambda)\, e^{-i\mathrm{Ph}(\lambda)} z_x(\lambda) d\lambda$$
$$= \int_{-\pi}^{\pi} e^{i\lambda(t - \mathrm{Ph}(\lambda)/\lambda)} G(\lambda)\, z_x(\lambda)\, d\lambda.$$

ゲインと位相の解釈は今や明らかになったであろう．それぞれの周波数において振幅要素$z_x(\lambda)$は$G(\lambda)$を乗ぜられている．一方，位相移動は$(t - \mathrm{Ph}(\lambda)/\lambda)$の項に表われている．すなわち周波数$\lambda$の周期的要素に関して，$y$は$\mathrm{Ph}(\lambda)/\lambda$期間だけ$x$から遅れることになっている．

### 循環的変動の平滑化および除去

上記(5.1)で与えられた線型フィルターの周波数応答関数は次のように与えられる．

(5.10) $$W(\lambda) = \frac{1}{5} \sum_{j=-2}^{2} e^{-i\lambda j}.$$

そこで付論 A(A.9)および(A.10)を用いることにより以下を得る．

$$W(\lambda)=\frac{\sin(5\lambda/2)}{5\sin(\lambda/2)} \tag{5.11}$$

$W(\lambda)$はこの場合実数となる．これはより一般的な結果を反映している．すなわち，$t$のまわりにおいて対称的な線型フィルターは位相の移動をもたらさない．これは正弦関数が奇関数であり，ゆえに

$$\sum_{j=-s}^{s} w_j e^{-i\lambda j}=\sum_{j=-s}^{s} w_j\cos\lambda j-i\sum_{j=-s}^{s} w_j\sin\lambda j \tag{5.12}$$

$$=w_0+2\sum_{j=1}^{s} w_j\cos\lambda j$$

となるからである．

上記(5.10)は実数であるから，ゲインは単にその絶対値である．時系列の平滑化とその循環(サイクル)に対する効果は図3.8(a)に見ることができる．高周波数におけるゲインが非常に小さいので，時系列における不規則性は減少させられる．したがってフィルターを乗ぜられた時系列のスペクトラムは$\pi$に近い周波数においては比較的小さなパワーしか持たない．このことは今や系列が原系列よりもよりゆるやかに変化することを意味している．

もしデータに5年の循環(サイクル)が存在するならば，もとの系列のスペクトラムは図3.8(b)に示されるようなものになるであろう．周波数$\lambda=2\pi/5$における頂(ピーク)は5年の期間に対応している．これが基本周波数(fundamental frequency)である．$\lambda=4\pi/5$における二番目のピークは最初の調和周波数(harmonic frequency)である．それ以降の調和周波数は図には示されていない．なぜなら二番目の調和周波数$6\pi/5$であり，これはナイキスト周波数$\pi$以上になるからである．フィルターの効果は$\lambda=2\pi/5$と$\lambda=4\pi/5$における循環的要素を完全に取り除くことになる．なぜならこれら周波数におけるゲインはゼロだからである．しかしながら循環的変動は完全に規則的なものではないので，基本周波数および調和周波数の周辺のパワーはいくらか残っている．

さて(5.3)のような階差フィルターは，5年の循環を取り扱うもう一つの方法を提供している．この場合の周波数応答関数は次のように与えられる．

$$W(\lambda)=1-e^{-i5\lambda} \tag{5.13}$$

したがって

$$G^2(\lambda)=(1-e^{-i5\lambda})(1-e^{i5\lambda}) \tag{5.14}$$



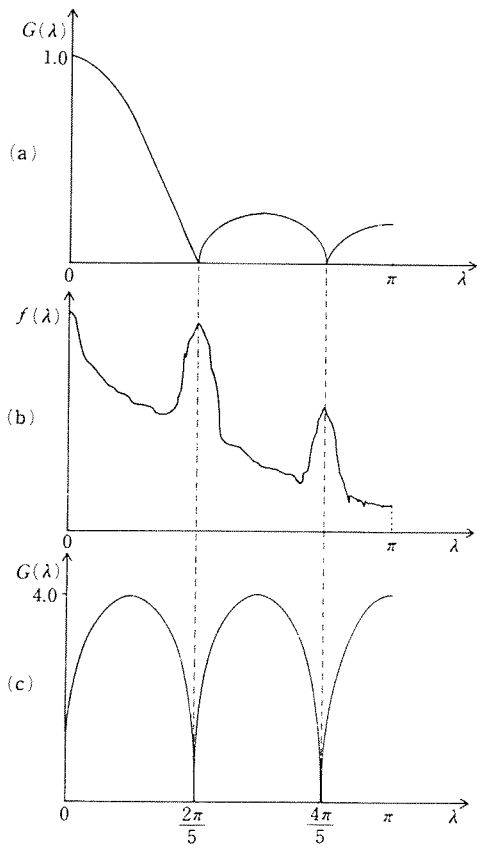



**図 3.8**　(a)年次データの5年の移動平均のゲイン，(b)不規則な5年周期を持つ系列のスペクトラム，(c)5年の階差フィルターのゲイン．

$$=1-e^{-i5\lambda}-e^{i5\lambda}+1$$

$$=2(1-\cos 5\lambda)$$

となる．図3.8(c)に示されたゲインの図は，移動平均フィルターでの対応する図に対して対照的であることは興味深い．双方ともに基本周波数およびその調和周波数におけるゲインはゼロである．しかし移動平均フィルターが高周波数から多くのパワーを取り除いているのに対し，階差フィルターは逆の効果を持っている．とくに$\lambda=0$におけるゲインはゼロである．すなわち移動平均フィルターは，不規則な高周波数要素を取り除いてトレンドとみなされるものの

みを残している．一方，階差操作は根本的にトレンドを除いている．

さらに(5.1)と違って，階差フィルターは位相効果をも持っている．このことは次のようにして示される．

$$\mathrm{Ph}(\lambda)=\tan^{-1}[(-\sin 5\lambda)/(1-\cos 5\lambda)].$$

ここで異なった周波数においてフィルターによってもたらされる位相移動は，対応した位相図を作ることによって調べることができる．

### 季節調整

月次データの時系列を考えよう．これは次のような加重を持つ線型フィルターを適用することによって季節調整を行なうことができよう．

$$w_j=\begin{cases}1/12, & j=0,\pm1,\pm2,\cdots,\pm5\\ 1/24, & j=\pm6.\end{cases} \tag{5.15}$$

この季節調整フィルターの周波数領域での性質は次のように与えられる．

$$W(\lambda)=\frac{\sin 6\lambda\cos(\lambda/2)}{12\sin(\lambda/2)}. \tag{5.16}$$

この表現は(5.11)が(5.10)から導かれたのと同様な方法で導くことができる．もっとも，加重がすべて同一ではないために導出手続きはやや複雑である．前項での議論から推察されるように，季節フィルターは基礎季節周波数 $\pi/6$ と調和周波数 $\lambda_j=\pi_j/12, j=2,\cdots,6$ においてゼロのゲインを持つ．したがってこれらの周波数におけるパワーは取り除かれる．このフィルターはさらに高周波数を弱めることによって系列を平滑化する効果がある．このことについてはFishman(1969, p. 46)を参照されたい．

上記(5.15)で定義されたフィルターを適用することは，季節調整を行なう比較的に簡単な方法である．しかしながらこのフィルターの性質を分析することによって，一般的な季節調整法の本質とそれが成し遂げようとしていることに関していくつかの基本的な問題を提起している．強い季節性を有する時系列のスペクトラムは，季節周波数においてピークを持つ．もし時系列が(5.15)のようなフィルターによって加工されるとすれば，もともと滑らかな形をしているスペクトラムは，季節周波数において谷によって切断させられる．これが季節調整法がなすべきことなのであろうか，またはスペクトラムにおける谷は何らかの過剰加工を意味するのではないのだろうか，この疑問に対して明確な答は存在しない．しかしこの点について，より進んだ議論に興味ある読者はNerlove 他(1979, 8章)，Burman(1980)を参照されたい．



### 見せかけの周期変動

半世紀以上も前に，時系列に対するある種の操作が奇妙な歪みを生じることが観察された．とくに，もしある程度の次数の階差をとり，次に一連の和分操作が適用されると，データに見かけ上の循環的変動をもたらすことが発見された．この現象はユール=スルツキー(Yule=Slutsky)効果として知られている．この現象が生じる理由は，周波数領域でこの問題を分析することによって非常に明確になる．

そのためにはまず多少の準備が必要である．二つの線型フィルターを時系列に対して順番に適用することを考えよう．$W_1(\lambda)$ を最初のフィルターの周波数応答関数を表わすとし，他方，$W_2(\lambda)$ を二番目のフィルターの対応する周波数応答関数と考えよう．二つの操作を組み合わせた全体的な効果は次式によって与えられる．

$$W(\lambda)=W_2(\lambda)\,W_1(\lambda) \tag{5.17}$$

対応する伝達関数は次のように与えられる．

$$\begin{aligned}|W(\lambda)|^2&=|W_2(\lambda)\,W_1(\lambda)|^2\\&=|W_2(\lambda)|^2\,|W_1(\lambda)|^2.\end{aligned} \tag{5.18}$$

したがって，原時系列のスペクトラムへの影響はそれぞれのフィルターの伝達関数の積によって決定される．ここで(5.17)の形式的な証明は必要としない．なぜならこの結果は，最初の原理，たとえば(5.9)からほとんど直ちに得られるからである．

$$|W_1(\lambda)|^2=2(1-\cos\lambda). \tag{5.19}$$

上記(5.18)より，一階の階差操作を $d$ 回適用することによって，次のような伝達関数がもたらされる．

$$\prod_{i=1}^{d}|W_1(\lambda)|^2=|W_1(\lambda)|^{2d}=2^d(1-\cos\lambda)^d \tag{5.20}$$

同様に和分フィルター $1+L$ を $s$ 回適用すると，次のような伝達関数が得られる．

$$|W_2(\lambda)|^{2s}=2^s(1+\cos\lambda)^s \tag{5.21}$$

したがって，$d$ 回の階差操作の後に $s$ 回の和分の操作が続けた全体の効果は以下のようになる．

(5.22) $$|W(\lambda)|^2 = 2^{s+d}(1-\cos\lambda)^d(1+\cos\lambda)^s.$$

階差の操作は低周波数を弱める．他方，和分の操作は高周波数を弱める．そこで複合された効果は中間の周波数を強調することになりうる．この現象は伝達関数がピークを持つ限り生じる．そこで(5.22)を$\lambda$に関して微分すると，それが以下の周波数においてゼロ点を持つことを示される．

(5.23) $$\lambda_0 = \cos^{-1}[(s-d)/(s+d)].$$

さらに二階の微分をとることにより，この点が最大であることが保証される．適当な$d$と$s$の選択により，$\lambda_0$はどのような望ましい周波数にも恣意的に近くすることができる．そして$d$と$s$が大きくなればなるほど，伝達関数のピークはより鋭いものとなる．

ところで，本質的にはユール=スルツキー効果であるような興味ある例がFishman(1969, pp. 45-49)によって提示されている．その例は時系列データを分析するに際して出会うかもしれないいくつかの落し穴を例示しており，重要である．そこで問題としているのはKuznets(1961)の長期景気変動仮説の研究であって経済時系列において，だいたい20年の循環的変動の存在を仮説とするものである．長期的循環変動の証拠のためにデータを調べるに先立って，Kuznetsはまず高周波数の循環的要素を幾分取り除くことを決め，二つのフィルターを適用した．第一は(5.1)の形の簡単な5年の移動平均であり，それは5年の景気変動の効果を弱めることを目的としていると考えられる．このフィルターは(5.11)で与えられたような周波数応答関数を持っている．二番目のフィルターは以下のような式で表される階差操作である．

(5.24) $$y_t = x_{t+5} - x_{t-5}.$$

この操作は$2i\sin 5\lambda$という周波数応答関数を持つ(付論(A.8)を参照のこと)．最後に，(5.18)を用いてこれら二つのフィルター操作の結果をまとめると，全体として次のような伝達関数が与えられる．

(5.25) $$|W(\lambda)|^2 = \left[\frac{2\sin 5\lambda \sin(5\lambda/2)}{5\sin(\lambda/2)}\right]^2.$$

これは20.3年の期間に対応する周波数を中心とした非常に高いピークを持つ伝達関数である．Kuznetsは，彼が観測している循環的変動の平均的期間はだいたい20年であると結論したが，(5.25)が示すように，これらの動きは二つのフィルター操作によってもたらされた見かけ上の循環的変動に対応し得るのである．



## 3.6　スペクトラムの推定

確率過程のスペクトラム密度関数は(1.1)において定義された．そこで$T$個の観測値の組が与えられたときには，次のような推定量が考えられる．

(6.1) $$I(\lambda) = \frac{1}{2\pi}\left[c(0) + 2\sum_{\tau=1}^{T-1} c(\tau)\cos\lambda\tau\right], \qquad 0 \leqslant \lambda \leqslant \pi$$

上式(6.1)においては，(1.1)における理論的自己共分散が標本共分散で置き換えられている．$T$は有限であるので(6.1)における和はもはや無限ではない．自己共分散は$T-1$の遅れについてまでのみ推定することができる．したがって，$c(T-1)$は最初と最後の一組の観測値のみの関数である．

上式(6.1)の表現は標本スペクトラム密度(sample spectral density)を定義しているが，ピリオドグラムと非常に密接な関係を持っている．ピリオドグラムは(2.16)により次のように与えられている．

$$p_j = \frac{2}{T}\left|\sum_{t=1}^{T}(y_t-\bar{y})e^{-i\lambda_j t}\right|^2$$

$$= \frac{2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}(y_t-\bar{y})(y_s-\bar{y})e^{-i\lambda_j(t-s)}.$$

ここで$\tau = t-s$とし，2章の(1.11)における$c(\tau)$の定義を注意することにより，次の結果が与えられる．

(6.2) $$p_j = 2\sum_{\tau=-(T-1)}^{T-1} c(\tau)e^{-i\lambda_j\tau}$$

$$= 2\left[c(0) + 2\sum_{\tau=1}^{T-1} c(\tau)\cos\lambda\tau\right].$$

上記において，最後の等号は$c(\tau)$の対称性からもたらされる．この(6.2)を(6.1)と比較することにより，(2.6)で定義された$\lambda_j$に関して次の関係が成り立つ．

(6.3) $$I(\lambda_j) = (1/4\pi)p_j.$$

もし$T$が偶数であるならば，$I(\lambda_n) = p_n/2\pi$である．

$I(\lambda)$は$[-\pi,\pi]$の範囲内の任意の$\lambda$から計算してもかまわないが，一般的にはピリオドグラムが計算される場合と同一の周波数において定義されると考える方が都合がよいので，このことは以下の議論において暗黙のうちに想定さ

れている．

標本スペクトラム密度関数$I(\lambda)$はピリオドグラムに比例しているので，文献においてそれはしばしばピリオドグラムとも呼ばれる．本書においてもその区別が不必要なときには，この呼び名をしばしば用いる．

### 標本スペクトラム密度の性質

もし観測値$y_1, \cdots, y_t$が正規分布にしたがうホワイト・ノイズ過程から抽出されたものであるならば，標本スペクトラム密度の性質は**3.2**節における結果を用いることにより簡単に導かれる．この場合には(2.17)における母数$\alpha$, $\beta$がすべてゼロであるから，(2.22)と(6.3)より，(2.6)で定義されたすべての$\lambda_j$について次式が成立する．

$$I(\lambda_j) \sim (\sigma^2/4\pi)\cdot\chi_2^2. \tag{6.4}$$

ホワイト・ノイズ過程の理論的なスペクトラム密度は，すべての$\lambda$に関して$\phi(\lambda)=\sigma^2/2\pi$である．自由度$f$の$\chi^2$分布の平均と分散はそれぞれ$f$と$2f$であるので，われわれは以下の結果を得る．

$$E[I(\lambda_j)] = (\sigma^2/4\pi)\,E(\chi_2^2) = \sigma^2/2\pi = f(\lambda) \tag{6.5}$$

一方，

$$\mathrm{Var}[I(\lambda_j)] = \frac{\sigma^4}{16\pi^2}\,\mathrm{Var}(\chi_2^2) = \left(\frac{\sigma^2}{2\pi}\right)^2 = f^2(\lambda). \tag{6.6}$$

与えられた$\lambda_j$について，標本スペクトラム密度は$f(\lambda)$の不偏推定量である．しかしながらその分散は$T$に依存しない．したがって，標本スペクトラム密度は与えられた周波数についてのスペクトラム密度の一致推定量をもたらさない．さらに(2.17)における母数の推定量は互いに独立であるので，$i \neq j$に対して

$$\mathrm{Cov}[I(\lambda_j), I(\lambda_i)] = 0 \tag{6.7}$$

となる．したがって正規分布する観測値について，標本スペクトラム密度の縦軸は互いに独立である．これらの性質の結果，(2.6)で定義された周波数において計算された標本スペクトラム密度のグラフは，ぎざぎざした不規則な形をしている．この様子は図3.9に例示されているがこの傾向はいかに標本が大きくなろうが持続される．

もし観測値が連続的スペクトラム密度関数を持つような定常確率過程によって生成されたものであるならば，$2I(\lambda)/f(\lambda)$は$0 \leqq \lambda \leqq \pi$の範囲において$\chi^2$



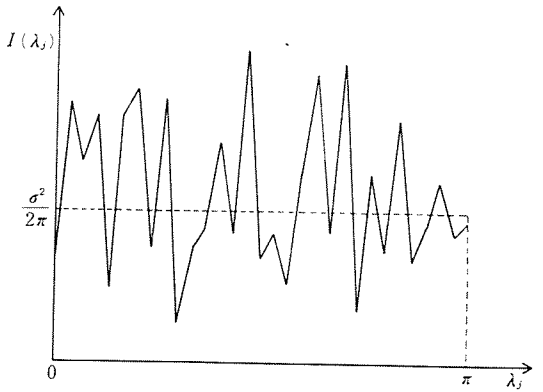



**図3.9**　ホワイト・ノイズ過程のスペクトラム密度

分布を極限分布として持つことが示される．したがって$I(\lambda)$の極限分布は平均$f(\lambda)$および分散$f^2(\lambda)$となる．異なった周波数における標本スペクトラム密度の推定値は漸近的に独立である．このことはホワイト・ノイズ・モデルにおいて述べられたように，$I(\lambda)$が相変わらず不規則な外観を持つことを意味するが，実証的にもたしかにこの通りであることが示されている．

### 標本スペクトラムの平均化

ホワイト・ノイズ過程について，(2.6)で定義される周波数における標本スペクトラム密度の推定量が互いに独立であることから，隣接の周波数についての平均をとることにより，$I(\lambda_j)$を平滑化することが考えられる．そこで周波数$\lambda=\lambda_j$における$f(\lambda)$の以下のような推定量が可能となろう．

$$\hat{f}(\lambda_j) = m^{-1}\sum_{i=-m^*}^{m^*} I(\lambda_{j-i}). \tag{6.8}$$

ここで$m = 2m^*+1$である．$\chi^2$分布の加法性にかんがみ，(6.8)の分布は以下のように与えられる．

$$\hat{f}(\lambda_j) \sim (\sigma^2/4\pi m)\,\chi_{2m}^2. \tag{6.9}$$

$\hat{f}(\lambda)$の期待値は$\sigma^2/2\pi$である．したがって不偏である．さらにその分散は次式で与えられる．

$$\mathrm{Var}[\hat{f}(\lambda_j)] = (\sigma^2/2\pi)^2/m = f^2(\lambda_j)/m. \tag{6.10}$$

ここで(6.10)には$T$は直接的に入ってはいないが，それは$T$が増大するとともに$m$を増大させることにより間接的には$T$に依存させることができる．こ

れを定式化するには定数 $\kappa$ により，$m = \kappa T$ とおけばよい．すると(6.10)の表現は以下のようになる．

$$\mathrm{Var}[\hat{f}(\lambda_j)] = [f^2(\lambda_j)/\kappa]/T. \tag{6.11}$$

すなわち $\hat{f}(\lambda_j)$ の分散のオーダーは $0(T^{-1})$ であるから，この推定量は $f(\lambda_j)$ の一致推定量となる．

上記(2.6)によって定義されたすべての周波数について(6.8)を計算するには簡単な移動平均をとれば便利であり，この方法によって $n$ 点における推定値を得ることができる：もっともその場合には，少なくとも $m$ 点以上離れた推定値は互いに独立となる．

上記(6.8)によって定義された推定量は，矩形ウィンドウ(rectangular window)と呼ばれるものを用いている．「ウィンドウ(窓)」という名称は，その加重和の計算過程において標本スペクトラム推定値のいくつかのみが考慮に入れられ，また「見られ」ているという概念を表わすものである．この場合におけるウィンドウ(窓)は矩形である．なぜなら，考慮される推定値は $\hat{f}(\lambda_j)$ の作成に当って等しい加重(ウェイト)を与えられているからである．

ウィンドウの幅(width)は単純にその基底の長さである．その幅は二つの方法によって測ることができよう．第一には領域(range)であり，それぞれの加重平均において用いられるスペクトラム点の数である．すなわち領域は $m$ に等しい．標本スペクトラムは $n$ 点存在する．したがってこれを $m$ で割った値は完全に独立な推定量の数を与える．第二の量は帯幅(bandwidth)である．それはラジアンで測ったウィンドウの幅である．もし $T = 200$ で $m = 5$ であるなら，$[0, \pi]$の範囲において 20 の独立したスペクトラムの推定量を作ることができる．したがって，それぞれの窓の帯幅は $\pi/20$ である．一般的に帯幅は $2\pi m/T$ に等しい．

ホワイト・ノイズのスペクトラムに関しては，どのような対称な加重平均も $f(\lambda)$ の不偏推定量となる．さらに(6.10)から明らかなように，標本数を固定したときには，$m$ が大きくなるに従ってその推定量の分散は小さくなる．ある時系列がホワイト・ノイズ過程から生成されたか否かにかかわらず，$m$ と推定量の分散の間には逆数の関係がある．しかしながら，いったんもとのスペクトラム密度関数が水平であるという仮説をゆるめると，(6.8)のような推定量がいつも不偏であるとは限らなくなる．さらに $m$ の選択に関して今やトレード・オフ(二者択一)が生じる．なぜなら帯幅が大きければ大きいほど，大きな



偏りを生じる可能性が増すからである．また大きな $m$ の値は，理想的な状況よりもはるかに多くの周波数においてスペクトラムの鋭い頂(ピーク)が「汚れをばらまく」ということになる．したがって明瞭な循環的傾向がはっきりとは現われないかもしれない．広すぎる帯幅に係わる問題は，解像度(resolution)という言葉によって捕えられている．貧弱な解像度は推定に多くの汚れや偏りが混入することを意味している．

これまでの議論は簡単な加重平均(6.8)に関したものであった．他の加重方法もまた考慮することができ，(6.8)は以下のように一般化されるであろう．

$$\hat{f}(\lambda_j) = \sum_{i=-m^*}^{m^*} h_i I(\lambda_{j-i}). \tag{6.12}$$

ここで加重 $h_i$ は和が 1 になるようにとるものとする．一つの方法としては，加重をウィンドウが三角形になるように選ぶことである．このことは推定の対象としている周波数に近い軸により多くの加重を与えることによっている．

上式(6.12)で与えられる推定量は，ホワイト・ノイズ系列のスペクトラムの不偏推定量である．そしてこの場合には，分散を求めるのはそれほど難しくない．この種の推定量の帯幅は，同一の分散を生じる推定量の矩形窓幅と等しくなるように定義される．

### 時間領域における加重和

スペクトラム推定に対するもう一つのアプローチとしては，自己共分散の加重和を用いることが考えられよう．標本スペクトラム密度(6.1)は，$k(0), k(1), \cdots, k(t-1)$ なる加重の系列によって以下のように修正される．

$$\hat{f}(\lambda) = (2\pi)^{-1}\left[k(0)c(0) + 2\sum_{\tau=1}^{T-1} k(\tau)c(\tau)\cos\lambda\tau\right], \quad 0 \leqq \lambda \leqq \pi. \tag{6.13}$$

この加重和はラグ・ウィンドウ(lag window)として知られている．

この種のもっとも簡単な方法として，特定の遅れの長さ $N$ を越える加重を切り捨てることが考えられる．すなわち $|\tau| \leqq N$ の場合は加重は 1 と置かれ，ラグ・ウィンドウは実際上矩形となる．これは直観的には賢明な方法であろう．なぜなら，切り捨てによって最も不安定な高次の標本共分散を取り除くからである．残念ながらこのような方法で作られた推定量は，ある種の望ましくない性質を持っている．このことはこの加重方法が周波数領域で何を意味しているかを調べることによって示すことができる．すなわち時間領域における一つの

ラグ・ウィンドウに対応してスペクトラム・ウィンドウが存在する．これは前節で説明したウィンドウと同一の考え方であるが，ただし今度は，スペクトラム・ウィンドウは$\lambda$の連続関数とみなすことになる．ここで$\lambda$を，興味の対象となる周波数とそれの近接の周波数の差として定義すると便利である．すると次の表現

$$h(\lambda)=(2\pi)^{-1}\left[k(0)+2\sum_{\tau=1}^{T-1}k(\tau)\cos\lambda\tau\right], \qquad -\pi\leqq\lambda\leqq\pi \tag{6.14}$$

は，$\lambda=0$を中心とするスペクトラム・ウィンドウ(spectral window)を定義することになる．ここで$h(\lambda)$は対称であるから正の$\lambda$に関してのみ定義すれば十分であり，通常，正の$\lambda$についてのみ図示される．上式(6.14)の導出は(6.13)における$c(\tau)$をそのフーリエ変換によって置き換え，そして結果の表現をまとめ直すことによって得られる．

端を切り捨てた(最も簡単な“ラグ・ウィンドウ”による)推定量については次の結果が得られる．

$$h(\lambda)=(2\pi)^{-1}\left[1+2\sum_{\tau=1}^{N}\cos\lambda\tau\right] \tag{6.15}$$

$$=\frac{\sin[(N+1/2)\lambda]}{2\pi\sin(\lambda/2)}.$$

最後の等号の導出には付論(A.9)を用いた．図3.10は$N=6$の場合のスペクトラム・ウィンドウを示している．より大きい$N$の値は，この$h(\lambda)$では$\lambda=0$のまわりにより集中させることになる．しかし同時にそれは分散を増大させるという犠牲を払うことにもなる．標本スペクトラム密度を直接的に平均化し

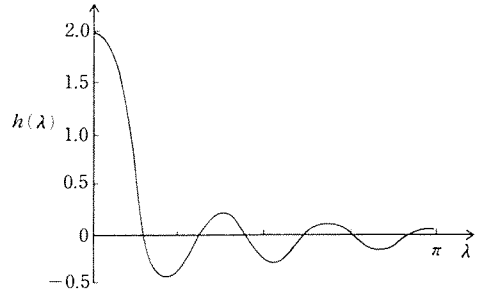



**図3.10** 切りとられた推定量のスペクトラム・ウィンドウ(6.15)：$N=6$の場合



た場合と同様に，偏りと分散の間にはトレード・オフが存在するのである．しかしながら図3.10におけるウィンドウの振動する様子は，たとえば周波数領域における矩形ウィンドウによってもたらされるものとは異なった推定量の偏りをもたらすことを意味している．矩形ウィンドウの場合には，スペクトラムにおける頂(ピーク)と谷は平滑化されがちになる．他方，スペクトラム・ウィンドウの場合はウィンドウには副次的な頂があるので，ある周波数のスペクトラムの推定値は他の周波数における重要な循環(すなわち頂)を反映しているかもしれない．このことは漏れ(leakage)として知られている．もし比較的大きな頂がウィンドウの負値をとる部分に対応する場合には，$\hat{f}(\lambda)$自身も負になるという可能性すら存在するのである．

加重を与える方法を適当に修正することによって，ラグ・ウィンドウに基づく推定量の固有の問題は大部分克服することができる．そこで多くの推定手続きがこれまで提案されている．もっともよく使われる二つの方法はBlackman=Tukeyによるものと，Parzenによるものである．両者とも，ある予め決められた遅れ$N$以上の加重から切り捨ててしまうが，$k(\tau)$の値は$\tau$が増大するに従って減少するようにウィンドウを作っている．高次の自己共分散により小さい加重が与えられており，そしてこれがより良い解像度を持つ推定量を得ることにつながっている．

### 高速フーリエ変換

標本スペクトラム密度を直接的に平均化する方法は，概念的にはスペクトラムの一致推定量を計算する最も直接的な方法である．しかしながらすべての$N$個のピリオドグラムの値を計算する仕事は，一見したところやや時間を浪費するように見える．そしてこのことが歴史的には，時間領域における標本共分散関数の加重和を検討する方向へ重点を向けてきた理由であった．しかしながら，1960年中頃から高速フーリエ変換(fast Fourier transform：FFT)と呼ばれる計算技術が急速に普及するようになった．この方法によってフーリエ係数の計算時間を$0(T^2)$から$0(T\log_2 T)$へと減少することができるようになり，ある程度大きい$T$の値についての計算においては非常に大きな時間的節約がもたらされた．時間領域における加重和の方法も相変わらず広く用いられているが，高速フーリエ変換によって計算手順の利用可能性が広がり，ピリオドグラムの平均化方法が再認識せられてきている．

### 自己回帰スペクトラムの推定

どのような線型確率過程もAR過程によって近似することができる．このことに注目したのがParzen (1969)であって，彼は自己回帰スペクトラム推定(autoregressive spectral estimation)と呼ばれる方法を提案した．パルゼン(Parzen)法では，第一に最小二乗法によってAR($P$)過程の係数を推定する．次にその推定値を(4.12)における母数$\phi_1, \phi_2, \cdots, \phi_P$に置き換え，それによってパワー・スペクトラム$f(\lambda)$の推定量を求める．いい換えると，AR($P$)モデルがデータに当てはめられ，そしてパワー・スペクトラムの推定量は当てはめられた確率過程の理論的なスペクトラムであるとみなされるのである．

自己共分散関数の加重和に基づいて推定されたときに，遅れの長さ$N$が$T$とともに増大するのを許されたと同様に，次数$p$も$T$とともに増大することが許されるとしよう．その場合には十分大きな$T$についてParzenは次のような結果を示した．

$$\mathrm{Var}[\hat{f}(\lambda)] \simeq [2Pf^2(\lambda)]/T. \tag{6.16}$$

分散が$p$に比例しているということは，ARモデルがもととなっているデータの生成過程の適切な近似であるためには，$P$が必要以上に大きくならないように決めなければならないことを示唆している．ここで一つの困難な問題は$P$の値を決定するということにある．なぜなら，もしそれが小さすぎるならば推定されたスペクトラムはひどく偏りを持つであろうし，時系列の周波数領域における特徴を大きく歪めることになる．ゆえに，このトレード・オフは他の形式のスペクトラムの推定において生じる問題と類似している．この問題への一つの解としては，自動的なルール，たとえば$P = \sqrt{T}$などを採用するのではなく，適合度基準によってモデルの次数を決める方法が考えられよう．$P$の値を決定するために採用されている種々の基準の中には1章(3.2)で定義されたAICも含まれている．AICについて詳しくは赤池(Akaike, 1969, 1974)を参照されたい．この方法は，自己回帰スペクトラムの推定に用いられてある程度の成功をおさめている．推定量は理論的なスペクトラムであるので，標準的な方法でもたらされるスペクトラムの推定値よりもはるかに滑らかな特徴を持つ．この方法は滑らかなスペクトラムを与えるにもかかわらず，高い解像度を持つように見える．すなわち，非常に狭い頂(ピーク)を取り出すこともできるようである．この点に関してはたとえばGriffiths=Prieto-Diaz (1977)を参照されたい．この問題は重要である．というのはスペクトラム解析の主な目的は，時系列における周期的変動を説明することにあるからである．いずれにしても$p$の値の決定に多大な注意が払われなければならない．もし設定が低すぎると，強い周期的動きを示すような，狭い頂(ピーク)を示す真のスペクトラムを再現することができないからである．



## 3.7　多変量スペクトラム解析

定常ベクトル値確率過程の時間領域における性質は**2.6**節で吟味した．$N \times N$の自己共分散行列の周波数領域における対応する量は，以下のような$N \times N$の多変量スペクトラム(multivariate spectrum)行列である．

$$F(\lambda) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \Gamma(\tau) e^{-i\lambda\tau}, \qquad -\pi \leqq \lambda \leqq \pi. \tag{7.1}$$

上記の$F(\lambda)$の対角要素は個々の確率過程のパワー・スペクトラムである．さらに$F(\lambda)$の$ij$要素は，$j \neq i$についての$i$番目と$j$番目の変数の間の相互スペクトラム(cross-spectrum)である．この相互スペクトラムは，周波数領域における二つの時系列の間の関係に関するすべての情報を含んでいる．なお$F(\lambda)$の$ji$要素は$ij$要素の単なる共役複素数である．

### ゲイン，位相およびコヒーレンス

二つの時系列の間の関係は，通常ゲインと位相によって特徴づけられる．これら二つの量は相互スペクトラムから導出されるが，いずれも複素数ではなく実数である．ここで$y_t$と$x_t$がベクトル値定常確率過程に従うと想定しよう．そしてそれぞれが$f_y(\lambda)$と$f_x(\lambda)$なる連続なパワー・スペクトラムを持っているとする．このとき$y_t$と$x_t$との間の相互スペクトラムは，$-\pi \leqq \lambda \leqq \pi$の範囲について以下のように定義される．

$$f_{yx}(\lambda) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma_{yx}(\tau) e^{-i\lambda\tau}. \tag{7.2}$$

ゲイン(gain)は以下のように定義される．

$$G(\lambda) = |f_{yx}(\lambda)|/f_x(\lambda), \tag{7.3}$$

さらに，位相(phase)は次のように定義される．

$$\mathrm{Ph}(\lambda) = \tan^{-1}\{-Im[f_{yx}(\lambda)]/\mathrm{Re}[f_{yx}(\lambda)]\}. \tag{7.4}$$

ゲインと位相の概念は，**3.5** 節において線型フィルターの性質を説明するためにすでに導入されたが，相互スペクトラム解析における解釈も同様なものである．ところで(4.1)を修正して，$x_t$ と独立に分布する攪乱項が右辺に加えられたと想定しよう．すると次のようなモデルがもたらされる．

$$y_t = \sum_{j=-r}^{s} w_j x_{t-j} + u_t, \qquad t = 1, \cdots, T. \tag{7.5}$$

ここで簡単化のために，もし $x_t$ の平均をゼロと想定すれば，(7.5)の両辺に $x_{t-\tau}$ を掛け，期待値をとることによって次の式が得られる．

$$\gamma_{yx}(\tau) = \sum_{j=-s}^{r} w_j \gamma_x(\tau - j), \qquad \tau = 0, \pm 1, \pm 2, \cdots \tag{7.6}$$

さらに上式(7.6)を(7.2)に代入することにより次式を得る．

$$f_{yx}(\lambda) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \sum_{j=-r}^{s} w_j \gamma_x(\tau - j)\, e^{-i\lambda\tau}.$$

ここで $k = \tau - j$ とし，和記号の順番を入れ替えることにより次のような結果を得る．

$$f_{yx}(\lambda) = (2\pi)^{-1} \sum_{j=-r}^{s} w_j e^{-i\lambda j} \sum_{k=-\infty}^{\infty} \gamma_x(k)\, e^{-i\lambda k}$$

これを(4.2)と比較すると次のように表わされることがわかるであろう．

$$f_{yx}(\lambda) = W(\lambda) f_x(\lambda). \tag{7.7}$$

このように(7.7)を得ることができたので，(7.3)および(7.4)と，**3.5** 節で与えられた形式とのつながりは明らかであろう．ゲインと位相は $y$ と $x$ の間の関係の特徴を記述しているが，それは(4.1)の線型フィルターの場合と同様である．ただし唯一の違いは，その関係が今や厳密なものではないということである．このことは $y_t$ と $x_t$ の間の関係の強さを，種々の周波数において測定するために何らかの第三の数量が必要であることを意味している．

ここで $z_t$ を $y_t$ の体系的部分を表わすものとしよう．すなわち，

$$z_t = \sum_{j=-r}^{s} w_j x_{t-j}, \qquad t = 1, \cdots, T. \tag{7.8}$$

なお，ここですべての $t$ および $\tau$ について $E(x_t u_\tau) = 0$ であるので，

$$\gamma_y(\tau) = \gamma_z(\tau) + \gamma_u(\tau), \qquad \tau = 0, 1, 2, \cdots \tag{7.9}$$

となる．したがって

$$f_y(\lambda) = f_z(\lambda) + f_u(\lambda) \tag{7.10}$$



である．コヒーレンス(coherence)

$$\mathrm{Coh}(\lambda) = f_z(\lambda)/f_y(\lambda) = 1 - f_u(\lambda)/f_y(\lambda), \qquad 0 \leqq \lambda \leqq \pi \tag{7.11}$$

は $f_y(\lambda)$ のうち，$x$ における動きによって体系的に説明できる部分の大きさである．ここで

$$f_z(\lambda) = |W(\lambda)|^2 f_x(\lambda) \tag{7.12}$$

であるから(たとえば(4.3)を参照)，(7.7)を代入することにより，コヒーレンスは以下のように他の形式で表わすこともできる．すなわち，

$$\mathrm{Coh}(\lambda) = \frac{|f_{yx}(\lambda)|^2}{f_x(\lambda) f_y(\lambda)}. \tag{7.13}$$

**例 1**　2章の(6.6)および(6.8)で定義されたモデルを考えよう．ここで

$$W(\lambda) = \beta e^{-i\lambda}, \tag{7.14}$$

であるから，(5.8)よりすべての周波数において $G(\lambda) = \beta$，$\mathrm{Ph}(\lambda) = \lambda$ であることが直ちにわかる．簡単な計算よりコヒーレンス関数は以下のように表わされる．

$$\mathrm{Coh}(\lambda) = \left[1 + \frac{f_u(\lambda)}{|W(\lambda)|^2 f_x(\lambda)}\right]^{-1} \tag{7.15}$$

したがってこの場合には以下のように求められる．

$$\mathrm{Coh}(\lambda) = [1 + \sigma^2 \beta^{-2} \sigma_\eta^{-2} (1 + \phi^2 - 2\phi \cos\lambda)]^{-1}. \tag{7.16}$$

ここで $\phi = 0.8$，$\beta = 1$ そして $\sigma^2 = \sigma_\eta^2 = 1$ とした場合についての値が図 3.11 に与えられている．この場合 $x_t$ はゆっくり変化するので，そのスペクトラムは低周波数において比較的高いパワーを持っている．そのことは $\lambda = 0$ の周辺における高いコヒーレンスに反映されている．他方，攪乱項は高周波数においてより多くの影響を与えている．そして $x$ と $y$ の間の関係は，その影響でより弱くなっている．ここで $\phi$ がコ

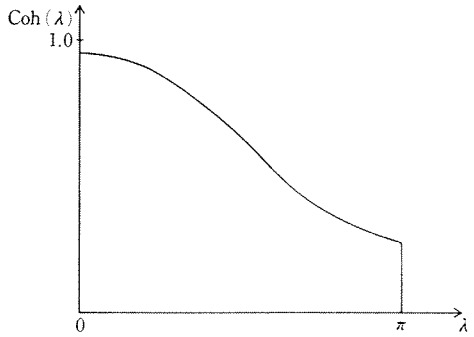



**図 3.11**　例 1 における $x_t$ と $y_t$ のコヒーレンス

ヒーレンス関数の全体的な形を決定する重要な母数であることに注意されたい．$\phi$ が正値である場合の方が，負値の場合よりも現実の時系列の特性としてもっともらしいものであろう．その意味で図 3.11 に示された形はおそらく特異なものではないであろう．

**例 2**　コイック (Koyck) 型または幾何級数分布ラグ (distributed lag) モデル

$$y_t = \beta \sum_{j=0}^{\infty} \alpha^j x_{t-j} + \varepsilon_t, \qquad 0 < \alpha < 1, \tag{7.17}$$

では，ラグ値の係数が指数的に減少する形になっている．すなわち (7.5) に関していえば，$w_j = \beta\alpha^j$ である．この時周波数応答関数は以下のように与えられる．

$$W(\lambda) = \beta \sum_{j=0}^{\infty} \alpha^j e^{-i\lambda j} = \beta/(1-\alpha e^{-i\lambda}). \tag{7.18}$$

もし $x_t$ が 2 章の (6.8) で与えられたような AR(1) 過程によって生成されているときには，そのコヒーレンスは以下のように与えられる．

$$\mathrm{Coh}(\lambda) = [1+\sigma^2\beta^{-2}\sigma_\eta^{-2}(1+\phi^2-2\phi\cos\lambda)(1+\alpha^2-2\alpha\cos\lambda)]^{-1} \tag{7.19}$$

もし $\phi$ が正のときには，図 3.11 の顕著な特徴，すなわち低周波数において比較的高いコヒーレンスが (7.19) によっても再現される．この特徴は $\alpha$ が 1 により近づくに従ってより強調されることになる．

## 推　　定

二つの時系列のデータが与えられれば，ゲイン，位相，そしてコヒーレンス関数をすべて推定することができる．これらの推定値によって，周波数領域において二つの時系列の間の関係が記述される．その際には，これまでの議論で取り上げたような線型関係によって，二つの時系列が関係づけられているというような仮説を必ずしも必要とするわけではない．前述の線型関係の議論は理論的な相互スペクトラムのある側面を例示するために導入されたにすぎない．実際に相互スペクトラム解析は時間領域においてはモデル化しにくいような，循環的関係の特徴づけを許すことにその魅力がある．

相互スペクトラムは複素数であるので，通常二つの実数に分解される．それらはコ・スペクトラム (co-spectrum)，$c(\lambda)$ と，クォドラチュア・スペクトラム (quadrature spectrum)，$q(\lambda)$ と呼ばれ，

$$f_{yx}(\lambda) = c(\lambda) + iq(\lambda) \tag{7.20}$$

であって



$$c(\lambda) = (2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma_{yx}(\tau)\cos\lambda\tau, \tag{7.21 a}$$

$$q(\lambda) = -(2\pi)^{-1} \sum_{\tau=-\infty}^{\infty} \gamma_{yx}(\tau)\sin\lambda\tau \tag{7.21 b}$$

によって定義される．$c(\lambda)$ や $q(\lambda)$ の直接的に解釈する試みが Granger=Hatanaka (1964, pp. 74-76) においてなされている．他方，相互スペクトラムの情報はゲイン，位相そしてコヒーレンスによってより容易に分析され，これらの量は $c(\lambda)$ および $q(\lambda)$ から簡単に求めることができる．

相互スペクトラムを推定するに当って考慮すべき主な点は，本質的にスペクトラムの場合と同様である．標本コ・スペクトラムと標本クォドラチュア・スペクトラムは (7.21) における理論的な相互共分散を，対応した標本相互共分散で置き換えることによって得ることができる．しかしながらゲイン，位相およびコヒーレンスの一致推定量は，何らかの平滑化が行なわれた場合にのみ得ることが可能となる．この平滑化は周波数領域において近接の周波数に関して平均化することによって直接的に行なえるし，また時間領域において標本相互共分散に加重を与える適当なラグ・ウィンドウを用いることによっても行なうことができる．

ゲイン，位相そしてコヒーレンスの標本分布については Fishman (1969, pp. 131-138) において詳細に述べられている．ここで重要な問題は，これらの統計量の大標本分散が母集団のコヒーレンスに依存しているということである．ある周波数において $\mathrm{Coh}(\lambda)$ が 1 に近ければ近いほど，対応する標本統計量の分散は小さくなるのである．他方，コヒーレンスが小さいときには，推定されたゲイン，位相およびコヒーレンスはすべてあまり信頼できないものとなる．しかしながらこのことは，標本コヒーレンスがいろいろな周波数における二つの時系列の間の「相関」の尺度であると解釈することができるということを認識すれば，少しも驚くべきことではないであろう．

## 付論 A　三角関数の定義式

和の公式 (addition formulae) は以下のように与えられる．

$$\sin(x \pm y) = \sin x \cos y \pm \cos x \sin y \tag{A.1}$$

$$\cos(x \pm y) = \cos x \cos y \mp \sin x \sin y. \tag{A.2}$$

これらの和の公式は多くの結果を導く基礎となる．たとえば (A.2) において $x = y$ と

置き，そして以下のような定義式

$$\sin^2 x + \cos^2 x = 1,$$

を用いて，式を整理すれば次の結果を得る．

(A.3) $$\cos^2 x = \frac{1}{2} + \frac{1}{2}\cos 2x.$$

このような結果は，三角関数の計算においてはしばしば有用である．たとえば，以下のように用いられる．

(A.4) $$\int \cos^2 x\, dx = \int\left(\frac{1}{2} + \frac{1}{2}\cos 2x\right)dx$$
$$= \frac{1}{2}x + \frac{1}{4}\sin 2x = \frac{1}{2}\left(x + \sin x \cos x\right).$$

次の定義式

(A.5) $$e^{ix} = \cos x + i \sin x$$

は，スペクトラム解析において特に重要な役割を果す．その主な理由の一つは，指数関数が比較的取り扱いが簡単であるということである．すなわち(A.5)が与えられたときには，最初の和の公式は以下の式

(A.6) $$e^{i(x+y)} = e^{ix} e^{iy}$$

の両辺を直接的に展開し，そして実数部分と虚数部分を等しくすることによって得ることができる．

指数項を扱う便利さによって，正弦関数および余弦関数はしばしば以下のように表わされる．

(A.7) $$\cos x = (e^{ix} + e^{-ix})/2,$$

(A.8) $$\sin x = (e^{ix} - e^{-ix})/2i.$$

双方の結果は(A.5)から直ちに導かれる．

これらの公式を使用する一つの例として，以下の恒等式の証明をあげよう．

(A.9) $$y_n(\lambda) = \frac{1}{2} + \sum_{x=1}^{n} \cos \lambda x = \frac{\sin(n+1/2)\lambda}{2\sin(\lambda/2)}.$$

まず初めに(A.7)を使って以下の式を導こう．

(A.10) $$y_n(\lambda) = \frac{1}{2}\sum_{x=-n}^{n} e^{i\lambda x} = \frac{1}{2}e^{-in\lambda}[1 + e^{i\lambda} + e^{i2\lambda} + \cdots + e^{i2n\lambda}].$$

上式で[ ]の中の項は有限の幾何級数和としてまとめられる．したがって以下の結果を得る．

$$y_n(\lambda) = \frac{1}{2}e^{-in\lambda}\frac{1 - e^{i\lambda(2n+1)}}{1 - e^{i\lambda}} = \frac{1}{2}\frac{e^{-in\lambda} - e^{i(n+1)\lambda}}{1 - e^{i\lambda}}.$$

この表現の分子と分母に$-e^{i\lambda/2}$を掛け，そして(A.8)を使うことにより目ざす結果が



得られる．

## 付論 B　直交関係

正弦関数および余弦関数を組み合わせた三角関数の直交関係は，簡単に以下のような形で表わすことができる．

(B.1) $$\int_{-\pi}^{\pi} e^{imx} e^{-imx}\, dx = \begin{cases} 0, & n \neq m \\ 2\pi, & n = m \end{cases}$$

ここで$n$と$m$は整数である．もし$n = m$のときは，この結果は直ちに導かれる．一方，$n - m = k \neq 0$のときは

(B.2) $$\int_{-\pi}^{\pi} e^{ikx} = \left[\frac{1}{ik}e^{ikx}\right]_{-\pi}^{\pi} = 0$$

である．なぜなら

$$e^{ik\pi} = e^{-ik\pi} = \begin{cases} 1 & k\text{が偶数のとき} \\ -1 & k\text{が奇数のとき}. \end{cases}$$

**3.2**節におけるフーリエ表現を導く際に用いられた直交関係は以下のようなものである．

(B.3 a) $$\sum_{t=1}^{T} \cos\frac{2\pi j}{T}t \cdot \cos\frac{2\pi k}{T}t = \begin{cases} 0, & 0 \leqslant k \neq j \leqslant [T/2] \\ T/2, & 0 < k = j < T/2 \\ T, & k = j = 0, T/2 \end{cases}$$

(B.3 b) $$\sum_{t=1}^{T} \sin\frac{2\pi j}{T}t \cdot \sin\frac{2\pi k}{T}t = \begin{cases} 0, & 0 \leqslant k = j \leqslant [T/2] \\ T/2, & 0 < k \neq j < T/2 \\ 0, & k = j = 0, T/2 \end{cases}$$

(B.3 c) $$\sum_{t=1}^{T} \cos\frac{2\pi j}{T}t \cdot \sin\frac{2\pi k}{T}t = 0, \qquad k, j = 0, 1, \cdots, [T/2].$$

ここで$T$が偶数のときは$[T/2] = T/2$であり，$T$が奇数の時は$[T/2] = (T-1)/2$である．

さらに(B.3 a)および(B.3 c)における$j$を$j = 0$と置くことにより，それぞれ以下の結果を得る．

(B.4 a) $$\sum_{t=1}^{T} \cos\frac{2\pi k}{T}t = 0, \qquad k = 1, \cdots, [T/2]$$

(B.4 b) $$\sum_{t=1}^{T} \sin\frac{2\pi k}{T}t = 0, \qquad k = 0, 1, \cdots, [T/2]$$

これらの結果を(A.5)と共に用いることにより以下の結果が導かれる．

(B.5) $$\sum_{t=1}^{T} e^{i\lambda t} = 0.$$

## 付論C　フーリエ変換

上記(B.1)および(B.2)における結果は，連続関数のフーリエ変換(Fourier transform)の基礎となる．以下の表現

$$f(x)=\frac{\alpha_0}{2}+\sum_{k=1}^{n}(\alpha_k\cos kx+\beta_k\sin kx) \tag{C.1}$$

は，$2n+1$のフーリエ係数$\alpha_0,\alpha_1,\cdots,\alpha_n,\beta_1,\cdots,\beta_n$を伴った次数$n$の三角関数多項式である．このとき$f(x)$は周期$2\pi$の正弦関数および余弦関数の線型結合であるので，$2\pi$の中のどのような範囲$x$に関しても連続関数である．さらに次のように定義することにより

$$\delta_k=(\alpha_k+i\beta_k)/2,\qquad \delta_{-k}=(\alpha_k-i\beta_k)/2\qquad \text{および}\qquad \delta_0=\alpha_0/2.$$

表現(C.1)は複素数表示で以下のように書き表わされる．

$$f(x)=\sum_{k=-n}^{n}\delta_k e^{-ikx}. \tag{C.2}$$

ここで$f(x)$が与えられて，$\delta_k$または同じことであるが$\alpha_k$および$\beta_k$を求めたいという状況を想定しよう．このとき(C.2)の両辺に$e^{ijx}$を掛け，それを$-\pi$から$\pi$の$2\pi$の範囲で積分することにより以下の結果を得る．

$$\int_{-\pi}^{\pi}f(x)e^{ijx}dx=\sum_{k=1}^{n}\delta_k\int_{-\pi}^{\pi}e^{-i(k-j)x},\qquad j=0,1,\cdots,n. \tag{C.3}$$

さらに(B.1)を用いることにより，

$$\delta_k=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)e^{ikx}dx. \tag{C.4}$$

上記の表現(C.2)および(C.4)はしばしばフーリエ変換の対と呼ばれる．
もし$f(x)$が偶関数であるならば，(C.2)は余弦項のみを含み，そして$\delta_k=\delta_{-k}$であり$\delta_k$は実数である．このときにはフーリエ変換の対は次のように書くことができる．

$$f(x)=\delta_0+2\sum_{k=1}^{n}\delta_k\cos kx, \tag{C.5}$$

$$\delta_k=\frac{\alpha_k}{2}=\frac{1}{\pi}\int_0^{\pi}f(x)\cos kx\,dx. \tag{C.6}$$

任意の連続関数は三角関数多項式によって近似することが常に可能である．一つの例として次のような関数を考えよう．

$$f(x)=\begin{cases}1/2b, & |x|\leqslant b\\ 0\ , & |x|>b.\end{cases} \tag{C.7}$$

これに(C.6)を応用することにより以下の結果を得る．



$$2\delta_k=\alpha_k=(2\pi b)^{-1}\int_{-b}^{b}\cos kx\,dx=\frac{\sin bk}{\pi bk} \tag{C.8}$$

ここで$\alpha_0=1/\pi$である．したがって次の結果を得る．

$$f(x)=\frac{1}{2\pi}+\frac{1}{b\pi}\sum_{k=1}^{\infty}\frac{\sin bk}{k}\cos kx. \tag{C.9}$$

### ノート

Fishman (1969) および Granger=Hatanaka (1964) がスペクトラム解析についてのよい入門書である．Fuller (1976)，Anderson (1971) および Hannan (1970) は，種々の重要な定理についての証明を含んでいる．

### 練習問題

**1.** MA (2) 過程のスペクトラムを導け．これは AR (2) 過程によってもたらされる循環的変動を含むことができるであろうか．

**2.** モデル，$y_t=\phi y_{t-12}+\varepsilon_t$ (ここで $|\phi|<1$)，スペクトラムを求めよ．そしてその形についてコメントを加えよ．

**3.** **3.5** 節における (5.1) のフィルターを適用し，フィルターを乗ぜられた時系列を原系列から引いた時系列を考えよ．この時系列のゲインを求め，そしてそれを (5.3) のゲインと比較せよ．

**4.** 位相 Ph$(\lambda)$ の表現をコ・スペクトラムとクォドラチュア・スペクトラムを用いて表現せよ．

**5.** **3.5** 節の (5.16) および (5.23) を導け．

**6.** AR (1) 過程において $\phi=0.5$，$\phi=0.99$ および $\phi=-0.5$ とした時のパワー・スペクトラムを描け．そしてコメントを加えよ．

**7.** ARMA (1, 1) 過程のパワー・スペクトラムの表現を導け．そしてその自己相関関数が図 2.5 に示されているような確率過程のスペクトラムを描け．

**8.** (a) 線型フィルターのゲインおよび位相の意味を注意深く述べよ．

(b) 適当なゲインおよび位相関数を計算することによって，7 年の周期を除くために以下のような形の階差フィルター

$$y_t^*=y_t-y_{t-7},$$

を用いることの有利さ，不利さを簡単な 7 年の移動平均を使うことと対比して論じよ．[ケント大学，1976]

**9.** (a) スペクトラム密度の異なった推定量を評価するために用いられる基準はどのようなものであるか．どのような意味において Blackman=Tukey 推定量が，切り捨て

ラグ・ウィンドウによる推定量より優れているのであろうか.

(b) 時間領域における加重和と周波数領域における加重和を関係づける表現を導け. そして切り捨てラグ・ウィンドウによる推定量のスペクトラム・ウィンドウが以下のように与えられることを示せ.

$$K(\lambda)=\frac{\sin(N+1/2)\lambda}{2\pi\sin(\lambda/2)}$$

ここで$N$は切り捨てられる点を表わす(すなわち, これ以上先のすべての自己共分散はゼロの加重が与えられる) [ケント大学, 1976].

**10.** 以下のモデルを想定しよう.

$$y_t=\alpha+\beta t+u_t,$$
$$u_t=\phi u_{t-4}+\epsilon_t,$$

ここで$\alpha$, $\beta$および$\phi$は母数であって$|\phi|<1$である. このモデルにおいて一階の階差がとられた時に生じる確率過程の自己相関関数およびパワー・スペクトラムの形を求めよ.

**11.** 以下の数値はある経済時系列の四半期観測値である.

| 年 | *Q.1* | *Q.2* | *Q.3* | *Q.4* |
|---|---|---|---|---|
| 1 | 2 | 5 | 9 | 3 |
| 2 | 1 | 3 | 6 | 5 |
| 3 | 3 | 4 | 8 | 6 |
| 4 | 4 | 4 | 7 | 6 |

適当な数の三角関数の項を当てはめることによってこの系列の季節性を推定せよ.

もしモデルが(2.17)の形で与えられたとしたときに, 攪乱項$\epsilon_t$の分散が2.25として知られたとして, それぞれの三角関数の係数がゼロと有意に異なるという仮説の検定を行なえ. その結果の意味についてコメントを加えよ. もし$\sigma^2$が未知であった場合に, そのような検定の手続きをどのように変えればよいのであろうか.



# 4章
# 状態空間モデルとカルマン・フィルター

## 4.1 状態空間

状態空間モデルはもともと制御工学の研究者によって発展させられたものである. その典型的な応用例では, 時間とともに変化する$m$個の状態変数(state variables)に関心が向けられるが、これらの変数はたとえばロケットの位置を表わす信号(signal)などである. ところで信号はしばしば直接的には観測可能ではなく「雑音」(noise)や系統的な歪みなどの影響を受けることがある.

実際に観測される$N$個の変数を$N\times1$ベクトル$y_t$とすると, それは観測方程式(measurement equation)によって状態変数と結びつけられる. 状態変数を$m\times1$ベクトル$\alpha_t$, さらに$Z_t$と$S_t$をそれぞれ$N\times m$, $N\times n$の固定された行列とすれば観測方程式は,

$$y_t=Z_t\alpha_t+S_t\xi_t,\qquad t=1,\cdots,T \tag{1.1}$$

と表現される. ここで$n\times1$攪乱項ベクトル$\xi_t$は平均ゼロ, 共分散行列$H_t$の確率変数である.

状態変数ベクトル$\alpha_t$は直接には観測可能ではないが, その挙動は遷移方程式(transition equation)

$$\alpha_t=T_t\alpha_{t-1}+R_t\eta_t,\qquad t=1,\cdots,T \tag{1.2}$$

によって定められる確率過程に従っていると仮定しよう. ここで$T_t$, $R_t$はそれぞれ$m\times m$, $m\times g$の固定された行列であって$\eta_t$は平均ゼロ, 共分散行列$Q_t$となる$g\times1$の攪乱項ベクトルである.

観測方程式および遷移方程式に表われる攪乱項は系列相関がなく, さらにあ

らゆる時刻において互いに無相関でかつ初期状態変数ベクトル $\alpha_0$ にも無相関であるとしよう．以上の仮定を要約すると，

$$(1.3)\qquad \begin{pmatrix}\xi_t\\ \eta_t\end{pmatrix}\sim \mathrm{WN}\left[0,\begin{pmatrix}H_t & 0\\ 0 & Q_t\end{pmatrix}\right],\qquad t=1,\cdots,T$$

かつ

$$(1.4)\qquad E[\alpha_0\eta_t']=0,\qquad E(\alpha_0\xi_t')=0,\ t=1,\cdots,T$$

と表現されよう．ここで WN とはホワイト・ノイズを意味する（2章(7.2)参照）．

以上の線型動学システムの表現は状態空間形(state space form)として知られている．一見したところこのシステムはこれまでの章で考察してきたどのモデルとも関係が希薄であるようにみえる．しかしながら本節で述べるように，任意の ARMA モデルは状態空間形として表現される．このことは時系列モデルの最尤推定に重要な意味を持つとともに，観測不能な要素(unobservable components)を含む時系列モデルを扱う枠組みを与えることにもなる．さらに回帰モデルと状態空間形の関係を用いることによって時変母数(time-varying parameter)を含むシステムのモデル分析に役立つこととなる．

### カルマン・フィルター

カルマン・フィルターとはある新しい観測値が利用可能になる度ごとに状態変数の推定値を更新する一連の方程式のことである．この更新過程は次の二段階を経て実行される．第一段階ではある時刻に利用可能な情報の下で次の時刻についての最適予測量を構成する．具体的にはそれは予測方程式(prediction equation)によって実行される．そして第二段階では新しい観測値が更新方程式(updating equation)を通じて状態変数ベクトルの推定に組み込まれる．

カルマン・フィルターは予測問題と更新問題へ一つの最適(optimal)解を与える．もし撹乱項が正規分布に従い，さらに状態変数の最新の推定量が達成可能なものの中で最適ならば，予測量と更新された推定量はともに最適となる．正規性を仮定しなければ観測値に関して線型な推定量と予測量のクラスの中でのみ類似の結果が成立する．ところで状態変数ベクトルも確率変数であるからここで「最適性」の解釈には注意する必要がある．通常と違って推定量自体よりもむしろ推定誤差に注意が向けられ，次節で定義する最小平均二乗(線型)推定(minimum mean square (linear) estimator)の概念もその意味においてのみ意



味を持つのである．

カルマン・フィルターの方程式は新しい観測値が一つ利用可能になる度ごとに逐次適用される．その副産物として予測誤差の系列が産み出されるが，最尤推定法におけるその意味は明らかであろう．というのは **1.2** 節で示したように，時間に関して独立でない観測量に対する尤度関数は予測誤差の系列に分解される．この分解は時系列モデルにおいて基本的なものであるが，カルマン・フィルターはその分解の一つの自然なメカニズムを与えているからである．

すべての観測値が用いられた後，今度は平滑化(smoothing)問題を解くために逐次法を逆向きに用いることができよう．更新方程式はある時刻までにおける利用可能なすべての情報をもとにして状態変数の最適な推定量を与える．これに対して平滑化推定量ではすべての情報が用いられている．すなわち，平滑化が観測値から状態変数の推定値を求めるために最適な手法となっている．

### ARMA モデルの状態空間表現

すでに2章(1.25)で与えた ARMA$(p,q)$ モデルを書き直すことによって

$$(1.5)\qquad y_t=\phi_1 y_{t-1}+\cdots+\phi_m y_{t-m}+\epsilon_t+\theta_1\epsilon_{t-1}+\cdots+\theta_{m-1}\epsilon_{t-m+1}$$

となる．ただし $m=\max(p,q+1)$ とする．さて(1.5)のマルコフ表現を得るには，ある $m\times 1$ ベクトル $\alpha_t$ を多変量 AR(1)モデル

$$(1.6)\qquad \alpha_t=\left[\begin{array}{c|c}\phi_1 & \\ \phi_2 & \\ \vdots & I_{m-1}\\ \hline \phi_m & 0'\end{array}\right]\alpha_{t-1}+\begin{bmatrix}1\\ \theta_1\\ \vdots\\ \theta_{m-1}\end{bmatrix}\epsilon_t$$

に従う確率変数と定めればよい．上式(1.6)は遷移方程式(1.2)において $T_t$ 及び $R_t$ は一定で $Q_t=\sigma^2$ であるとみなすことができる．逆に(1.6)より $\alpha_t$ の第一番目の要素を $y_t$ とするとシステムの下から上への代入を繰り返すことによってもとの ARMA モデルを得ることができる．

**例 1**　MA(1)モデルのマルコフ表現は

$$(1.7)\qquad \alpha_t=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\alpha_{t-1}+\begin{bmatrix}1\\ \theta\end{bmatrix}\epsilon_t$$

である．ここで $\alpha_t=(\alpha_{1t},\alpha_{2t})'$ と置けば，$\alpha_{2t}=\theta\epsilon_t$ かつ $\alpha_{1t}=\alpha_{2,t-1}+\epsilon_t=\epsilon_t+\theta\epsilon_{t-1}$ となり，これがもとの MA(1)モデルとなる．

ここで観測方程式について必要なことは状態変数ベクトルの第一要素を取出すことだけである．すなわち(1.1)は単一方程式となり撹乱項 $\xi_t$ は恒等的にゼロとなる．すなわち，

(1.8) $$y_t = z_t'\alpha_t, \qquad t = 1, \cdots, T$$

であって $z_t = (1, 0'_{m-1})$ となる．

ところで ARMA モデルに観測誤差が加わったモデルについても同様の表現を得ることができる．例えば観測値は，

(1.9) $$y_t = w_t + \xi_t, \qquad t = 1, \cdots, T$$

に従っているものとしよう．ここで $w_t \sim \mathrm{ARMA}(p, q)$ でかつ $\xi_t \sim \mathrm{WN}(0, \sigma^2 h)$ とする．このとき遷移方程式は(1.6)の場合と同一であるが $\alpha_t$ の第一要素は $w_t$ となって(1.8)は観測誤差を含むことになっている．

## 4.2 最小平均二乗推定[訳注]

次のモデルを考えよう．

(2.1) $$y = Z\alpha + \xi$$

ただし $y$ は $T\times 1$ の観測値ベクトル，$Z$ は $T\times m$ の固定された行列，$\xi$ は $T\times 1$ の撹乱項ベクトルで平均ゼロかつ共分散行列は $\sigma^2\Omega$ としよう．$\alpha$ は $m\times 1$ ベクトルの未知母数とすれば，(2.1)は一般化回帰モデルの形になっている．しかしながら通常とは違う一つの重要な側面がある．それは $\alpha$ は $y$ の観測値が利用される前にある事前分布よりランダムに抽出される意味で確率変数であるとしていることである．

ここでもし $\alpha$ を固定母数とすると，その一般化最小二乗(GLS)推定量

(2.2) $$\tilde{a} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y$$

は $\alpha$ の最良線型不偏推定量(BLUE)になっている．$\alpha$ が確率変数の場合にもこの $\tilde{a}$ は推定誤差 $\tilde{a}-\alpha$ に関してある最適な統計的性質を持っている．ここで

(2.3) $$(\tilde{a}-\alpha) \sim \mathrm{WS}[0, \sigma^2 (Z'\Omega^{-1}Z)^{-1}]$$

であるから，$\tilde{a}-\alpha$ の性質は通常の GLS 推定の場合と平行して導くことができる．ここで WS$(\mu, V)$ において $\mu$ は平均，$V$ は共分散行列を示し，WS は必ずしも正規性を仮定しない「広義」を意味している．ガウス=マルコフ(Gauss=Markov)定理はここで線型かつ無条件不偏な推定量のクラスの中で $\tilde{a}$ が最適となるように修正されねばならない．ただしある推定量が無条件不偏(unconditionally unbiased，あるいは $u$-unbiased，$u$-不偏と略す)であるとは，その推定誤差の期待値がゼロとなることである．

ここで単純化のために $\Omega = I$ と仮定する(もしそうでなければ $L'L = \Omega^{-1}$ となる $T\times T$ の非特異行列 $L$ で観測ベクトルを変換することによって以下の結果を一般的に示すことができる)．さて $\hat{a} = D^*y$ を $\alpha$ の他の任意の線型推定量としよう．このとき $D = D^* - (Z'Z)^{-1}Z'$ とおくと，

$$\hat{a} - \alpha = DZ\alpha + [D + (Z'Z)^{-1}Z']\xi,$$

となる．ここで $E(\hat{a}-\alpha) = 0$ であるから，$DZ = 0$ が必要である．この条件の下で $\hat{a}-\alpha$ の共分散行列は，

$$\begin{aligned}\mathrm{Var}(\hat{a}-\alpha) &= E[(\hat{a}-\alpha)(\hat{a}-\alpha)'] \\ &= [D + (Z'Z)^{-1}Z']E(\xi\xi')[D' + Z(Z'Z)^{-1}]\end{aligned}$$

となる．さらに $E(\xi\xi') = \sigma^2 I$ かつ $DZ = 0$ であるから，この表現は簡約化されて

$$\mathrm{Var}(\hat{a}-\alpha) = \sigma^2 DD' + \sigma^2 (Z'Z)^{-1}$$

となる．したがって $\hat{a}-\alpha$ の共分散行列は $\tilde{a}-\alpha$ の共分散行列と，ある非負値定符号行列の和となることがわかる．

ある推定量が $u$-不偏のとき，その平均二乗誤差(MSE)行列 $E[(\hat{a}-\alpha)(\hat{a}-\alpha)']$ は推定誤差の共分散行列に等しい．ところがもし推定量の望ましさとして $\hat{a}-\alpha$ の各要素の平均二乗誤差の最小化基準を採用すると推定量は必然的に $u$-不偏とならねばならない．このことを見るために平均二乗誤差を二つの部分に分割して，

(2.4) $$E[(\hat{a}_i-\alpha_i)^2] = [E(\hat{a}_i-\alpha_i)]^2 + \mathrm{Var}(\hat{a}_i-\alpha_i), \qquad i = 1, \cdots, m$$

とする．$\hat{a}_i-\alpha_i$ の分散は $\hat{a}_i$ が $u$-不偏であるか否かに依存しないので(2.4)の最小化は二段階に分けることができる．第一に推定量を $u$-不偏に限ることによって第一項はゼロに等しくなる．次に前述の修正されたガウス=マルコフ定理によって $\hat{a}$ は $\tilde{a}$ に等しくなる．この結果，すなわち $\tilde{a}$ が $\alpha$ の最小平均二乗線型推定量(MMSLE)であることは拡張ガウス=マルコフ定理(the extented Gauss=Markov theorem)として知られている．

訳注　本章での用語，例えば最小平均二乗推定量や最小平均二乗線型推定量等は統計学で通常用いる意味とは異なっている．ARMA$(p, q)$ モデルの状態空間表現から導かれる状態変数の推定量は未知母数に依存するが，本章の推定の最適性の議論ではその母数をあたかも既知であるかのように扱っている．これらの点について読者は注意する必要があろう．

ここでもし攪乱項が正規分布に従っているとすれば，$\tilde{a}$ の最適性は線型推定量のクラスに限定されない．すなわちこの場合には，$\tilde{a}$ は $\alpha$ の最小平均二乗推定量(MMSE)となっている．

ところで最小平均二乗線型推定量と最小平均二乗推定量などの用語は予測量についても用いることができる．というのは予測量とは単に確率変数の将来値の推定量にすぎないからである．

**事前情報と標本情報の結合**

さて $\alpha$ の情報がすでに利用可能であって

$$a_0 - \alpha \sim \mathrm{WS}(0, \sigma^2 P_0) \tag{2.5}$$

の形で表現されているとしよう．ここで $a_0$ はある既知ベクトルで $P_0$ は既知の正定符号行列とする．事前情報を $a_0 - \alpha$ の分布の形で表現しておけば，$\alpha$ の各要素は固定的であろうと確率的であろうとどちらでもよい．そこで(2.5)で表現された事前情報を(2.1)で表現される標本情報に結びつけよう．その第一のステップとして拡張されたモデル，

$$\begin{pmatrix} a_0 \\ y \end{pmatrix} = \begin{pmatrix} I \\ Z \end{pmatrix} \alpha + \begin{pmatrix} a_0 - \alpha \\ \xi \end{pmatrix} \tag{2.6}$$

を考える．このモデルを

$$y^\dagger = Z^\dagger \alpha + \xi^\dagger \tag{2.7}$$

と表わすと，$E(\xi^\dagger) = 0$ でかつ

$$E(\xi^\dagger \xi^{\dagger\prime}) = \sigma^2 V = \sigma^2 \begin{pmatrix} P_0 & 0 \\ 0 & \Omega \end{pmatrix} \tag{2.8}$$

となる．ここで $\alpha$ はその要素の中に固定母数を含み得るように定義されているが，もちろん拡張ガウス=マルコフ定理を適用できる．ゆえに，推定量

$$\tilde{a}^\dagger = (Z^{\dagger\prime} V^{-1} Z^\dagger)^{-1} Z^{\dagger\prime} V^{-1} y^\dagger \tag{2.9}$$

は $\alpha$ の最小平均二乗線型推定量となっている．もとの記号で(2.9)を書き直すと，

$$\tilde{a}^\dagger = P(P_0^{-1} a_0 + Z' \Omega^{-1} y) \tag{2.10}$$

となるが，ここで

$$P = (P_0^{-1} + Z' \Omega^{-1} Z)^{-1} \tag{2.11}$$

としている．したがって，

$$\tilde{a}^\dagger - \alpha \sim \mathrm{WS}(0, \sigma^2 P) \tag{2.12}$$



となっている．

もし $\alpha$ の総ての要素が固定母数であれば，(2.1)は一般化回帰モデルである．事前情報(2.5)を活用する方法は Theil=Goldberger (1961) によって導入され，Theil (1970, pp.346-352) に記述されている混合推定法 (mixed estimation) の特殊な場合である．他方，もし $\alpha$ が確率変数の場合にはベイズ的 (Bayesian) 解釈ができよう．すなわち(2.5)は $\alpha$ の事前分布 (prior distribution)，(2.12)は事後分布 (posterior distribution) である．ただしその場合には両者ともに未知母数 $\sigma^2$ についての条件付分布として考えねばならない．

## 4.3 更新方程式と予測方程式

各時刻において観測値が一個のみ利用可能なことがしばしば見られる．その場合には(1.1)は単一の観測方程式に還元され，

$$y_t = z_t' \alpha_t + \xi_t, \qquad t = 1, \cdots, T \tag{3.1}$$

および $\xi_t \sim \mathrm{WN}(0, \sigma^2 h_t)$ となる．遷移方程式も(1.2)よりも単純であり，例えば $T_t$，$R_t$，$Q_t$ が時間に関して不変なことがしばしばある．その場合には，

$$\alpha_t = T \alpha_{t-1} + R \eta_t, \qquad t = 1, \cdots, T \tag{3.2}$$

であって，$\eta_t \sim \mathrm{WN}(0, \sigma^2 Q)$ となる．ここで $\mathrm{Var}(\xi_t)$ と $\mathrm{Var}(\eta_t)$ に表われるスケール母数 $\sigma^2$ は必ずしも必要ではないが，$N = 1$ の場合の多くのモデルではこのような表現が自然となる．$\sigma^2$ を導入することによって一般性を失わない上に更新方程式と予測方程式が $\sigma^2$ と独立に表現され，しかも方程式の導出上に多少の簡便さをもたらすことになる．すなわち，カルマン・フィルターは $T$，$R$，$Q$，$h_t$ などの母数に依存するが，$\sigma^2$ とは無関係に表現される．

以下ではカルマン・フィルターをまず(3.1)と(3.2)のシステムについて導出し，次にそれを一般状態空間モデルに拡張する．ここで記号 $a_t$ をすべての過去と今期の観測値を含んだ情報に基づく $\alpha_t$ の最小平均二乗線型推定量としよう．すなわち，$a_t$ は時刻 $t$ における $\alpha_t$ の最小平均二乗線型推定量であり，$a_{t/t-1}$ は時刻 $t-1$ における $\alpha_t$ の最小平均二乗線型推定量を表わすものとする．

**予　　測**

時刻 $t-1$ において利用可能な情報は $\alpha_{t-1}$ の最小平均二乗線型推定量である $a_{t-1}$ に縮約されている．$a_{t-1}$ の共分散行列は $P_{t-1}$ を既知の行列として $\sigma^2 P_{t-1}$

と書くことができる．

さて(3.2)で表わされる遷移方程式より時刻$t-1$における$\alpha_t$の最小平均二乗線型推定量は，

$$a_{t/t-1} = Ta_{t-1} \tag{3.3}$$

となる．そこで(3.3)の両辺より$\alpha_t$を減じると，

$$a_{t/t-1} - \alpha_t = T(a_{t-1} - \alpha_{t-1}) - R\eta_t \tag{3.4}$$

を得る．この(3.4)の右辺の期待値はゼロであるから$a_{t/t-1}$は$u$-不偏となる．推定誤差の共分散行列は(3.4)より直接求めると，

$$E[(a_{t/t-1} - \alpha_t)(a_{t/t-1} - \alpha_t)'] = \sigma^2 TP_{t-1}T' + \sigma^2 RQR'$$

となる．したがって，

$$a_{t/t-1} - \alpha_t \sim \mathrm{WS}(0, \sigma^2 P_{t/t-1}) \tag{3.5}$$

であって，

$$P_{t/t-1} = TP_{t-1}T' + RQR' \tag{3.6}$$

と表現される．時刻$t-1$における$\alpha_t$の最小平均二乗線型推定量である$a_{t/t-1}$が与えられると，時刻$t-1$における$y_t$の最小二乗線型推定量は

$$\tilde{y}_{t/t-1} = z_t' a_{t/t-1} \tag{3.7}$$

となり，その予測誤差が

$$v_t = y_t - \tilde{y}_{t/t-1} = z_t'(\alpha_t - a_{t/t-1}) + \xi_t \tag{3.8}$$

で与えられることは明らかであろう．$\alpha_t - a_{t/t-1}$及び$\xi_t$はともに期待値がゼロであるから，$E(v_t) = 0$となる．したがって

$$\begin{aligned}\mathrm{Var}(v_t) = E(v_t^2) &= E[z_t'(\alpha_t - a_{t/t-1})(\alpha_t - a_{t/t-1})'z_t] \\ &\quad + E(\xi_t^2) + 2E[z_t'(\alpha_t - a_{t/t-1})\xi_t]\end{aligned}$$

であるが，右辺第三項の期待値がゼロとなることから，

$$\mathrm{Var}(v_t) = \sigma^2 z_t' P_{t/t-1} z_t + \sigma^2 h_t = \sigma^2 f_t \tag{3.9}$$

となる．

以上を要約すると(3.3)と(3.6)は状態変数ベクトルとその共分散行列に対する予測方程式(prediction equation)であり，(3.8)は時刻$t-1$において$y_t$を予測した時の予測誤差である．

### 更　　新

最適予測量$a_{t/t-1}$に縮約される，すでに利用可能な情報と新しく利用できる$y_t$の情報を結びつけるのが更新方程式の役割である．この問題は前節で述べた



事前情報(2.5)を標本情報(2.1)と結びつけることに類似している．この場合には事前情報は(2.5)で表わされ，標本は遷移方程式(3.1)より発生した単一の観測値である．したがって拡張されたモデルは，

$$\begin{pmatrix} a_{t/t-1} \\ y_t \end{pmatrix} = \begin{pmatrix} I \\ z_t' \end{pmatrix} \alpha_t + \begin{pmatrix} a_{t/t-1} - \alpha_t \\ \xi_t \end{pmatrix} \tag{3.10}$$

と表現される．攪乱項の期待値はゼロでありその共分散行列は，

$$\begin{aligned} &E\left[ \begin{pmatrix} a_{t/t-1} - \alpha_t \\ \xi_t \end{pmatrix} (a_{t/t-1}' - \alpha_t', \xi_t) \right] \\ &\qquad = \sigma^2 \begin{bmatrix} P_{t/t-1} & 0 \\ 0 & h_t \end{bmatrix} \end{aligned} \tag{3.11}$$

となる．

さらに(2.10)と(2.11)より$\alpha_t$の最小平均二乗線型推定量は，

$$a_t = P_t(P_{t/t-1}^{-1} a_{t/t-1} + z_t y_t / h_t) \tag{3.12}$$

で与えられるが，ここで

$$P_t = (P_{t/t-1}^{-1} + z_t z_t' / h_t)^{-1} \tag{3.13}$$

である．したがって，

$$a_t - \alpha_t \sim \mathrm{WS}(0, \sigma^2 P_t) \tag{3.14}$$

となる．

さて，(3.12)で与えられる更新公式は本章の付論に与えた行列の逆転公式を用いると別の形で表現される．ここで付論(A.1)の記号を用いて$P_t = D$，$P_{t/t-1}^{-1} = A$，$z_t = B$，$h_t = C$とおけば(3.13)と付論(A.2)より

$$P_t = P_{t/t-1} - P_{t/t-1} z_t z_t' P_{t/t-1} / f_t \tag{3.15}$$

$$f_t = z_t' P_{t/t-1} z_t + h_t \tag{3.16}$$

を得る((3.9)を参照)．

この(3.15)によって行列の逆転操作を含まない更新公式を得ることができる．すなわち，ここでさらに(3.15)を(3.12)に代入すれば，

$$\begin{aligned} a_t &= (P_{t/t-1} - P_{t/t-1} z_t z_t' P_{t/t-1} / f_t)(P_{t/t-1}^{-1} a_{t/t-1} + z_t y_t / h_t) \\ &= a_{t/t-1} + P_{t/t-1} z_t (y_t / h_t - z_t' a_{t/t-1} / f_t - z_t' P_{t/t-1} z_t y_t / h_t f_t) \\ &= a_{t/t-1} + f_t^{-1} P_{t/t-1} z_t (y_t f_t / h_t - z_t' a_{t/t-1} - z_t' P_{t/t-1} z_t y_t / h_t) \end{aligned}$$

となる．そこで(3.16)での$f_t$の定義を用いて整理すると，状態変数の更新方程式

$$a_t = a_{t/t-1} + P_{t/t-1} z_t (y_t - z_t' a_{t/t-1}) / f_t \tag{3.17}$$

を得る．ここで予測誤差 $y_t - z_t' a_{t/t-1}$ の演じている役割は興味深いものである．すなわち $y_t$ の新しい情報を総て含みカルマン・ゲイン(Kalman gain)を通じて $a_{t/t-1}$ の更新に用いられている．カルマン・ゲインとは $m \times 1$ ベクトル $P_{t/t-1} z_t / f_t$ のことである．この量は $P_t$ や $P_{t/t-1}$ と同様 $y_t$ とは独立であって時刻 $t$ において事前に計算しておくことができる．

一連の方程式(3.15)～(3.17)が基本的な更新方程式(updating equations)である．その導出に際しては $h_t > 0$ を仮定している．もしそうでないと行列(3.11)は特異になる．しかしながらタイル(1971, pp. 282-287)の結果を用いるとこの更新方程式は $h_t = 0$ の場合にも適用可能なことがわかる．モデルの状態空間表現の中には観測方程式に攪乱項を含まない場合もあるので，以上の点は重要であろう．例えば(1.8)の ARMA 過程に対する方程式はこの場合にあたる．

カルマン・フィルターについての一般的な注意点としては状態変数ベクトルの中に固定母数が存在する場合にも妥当することであろう．このことは前節の議論から明らかである．

### カルマン・フィルターの一般形

一般状態空間モデル(1.1)及び(1.2)に対する予測方程式と更新方程式はこれまでに述べた特殊な場合と同様にして導かれる．時刻 $t-1$ における $\alpha_{t-1}$ の最小平均二乗線型推定量である $a_{t-1}$ が，

(3.18) $$a_{t-1} - \alpha_{t-1} \sim \mathrm{WS}(0, P_{t-1})$$

として与えられると，予測方程式は

(3.19 a) $$a_{t/t-1} = T_t a_{t-1}$$

および

(3.19 b) $$P_{t/t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t', \qquad t = 1, \cdots, T$$

となる．そこで更新方程式は

(3.20 a) $$a_t = a_{t/t-1} + P_{t/t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t/t-1}),$$

(3.20 b) $$P_t = P_{t/t-1} - P_{t/t-1} Z_t' F_t^{-1} Z_t P_{t/t-1},$$

(3.20 c) $$F_t = Z_t P_{t/t-1} Z_t' + S_t H_t S_t', \qquad t = 1, \cdots, T$$

で与えられる．予測誤差

(3.21) $$v_t = y_t - Z_t a_{t/t-1}, \qquad t = 1, \cdots, T$$

はこの場合には $N \times 1$ ベクトルとなり，その平均はゼロで共分散行列は $F_t$，すなわち $E(v_t) = 0$，$E(v_t v_t') = F_t$ である．単一の方程式の場合と同様に予測誤差は状態変数の更新に際して重要な役割を演じる．すなわち(3.20 a)における $a_{t/t-1}$ の修正項はカルマン・ゲイン $P_{t/t-1}^{-1} Z_t' F_t^{-1}$ と $v_t$ の積になっている．

もし事前情報が $a_0$ と $P_0$ を既知として

(3.22) $$\alpha_0 \sim \mathrm{WS}(a_0, P_0)$$

で与えられていれば，カルマン・フィルターは $T$ 個のあらゆる観測値に基づいて $\alpha_T$ の最小平均二乗線型推定量をもたらす．事前情報(3.22)についての適切な定式化を得ると云う問題については次節で吟味する．



## 4.4 最尤推定

さて(3.1)と(3.2)で与えられる状態空間モデルにおいて $\alpha_t$ のすべての要素が確率変数であるとしよう．さらに攪乱項 $\xi_t$，$\eta_t$ は正規分布に従い，初期時点において $a_0$，$P_0$ を既知として，

(4.1) $$\alpha_0 \sim N(a_0, \sigma^2 P_0)$$

であるとしよう．以上の仮定の下で $T \times 1$ 観測値ベクトル $y$ は平均 $\mu$，共分散行列 $\sigma^2 V$ の多変量正規分布に従っている．初期値母数 $a_0$，$P_0$ は遷移方程式と観測方程式に表われる母数とともに $\mu$ と $V$ の形を決定する．それにもかかわらず，一般にこれら $\mu$ や $V$ を具体的に評価する必要はない．というのはカルマン・フィルターを適用すると，すべての $t = 1, 2, \cdots, T$ について前期までの観測値を与えた時の $y_t$ の最小平均二乗推定量とその予測誤差の系列が導かれる．したがって1章(2.15)で与えられる尤度関数は予測誤差分解(1章(2.21))から直接的に得ることができる．

以上の議論は通常逆の方向で行なわれることが多い．すなわち，ある時系列データを一つの時系列モデルから発生していると仮定し，そのモデルから1章(2.15)で表現される尤度関数を導く．次にモデルの状態空間表現を求め予測誤差分解によって尤度関数を評価する．ここで解決されるべき残された問題は適切な状態空間表現を決定し，初期値を構成することである．

ところで，状態空間モデルの中でも状態変数ベクトルが定常的であるか否かによって二つのカテゴリーを区別すべきであろう．定常的な場合には状態変数の初期値の分布は自動的に与えられる．以下では議論を(3.1)と(3.2)で与えられている状態空間表現を持つモデルに限定する．むろん多変量における予測分

解をもたらすカルマン・フィルターによって同様の議論を(1.1)と(1.2)で与えられる一般的な場合に拡張することができる.

### 定常モデル

もし(3.2)において $\alpha_t$ が定常であるならば, 遷移行列 $T$ は1章(1.15)における多変量AR(1)モデルにおける母数 $\Phi$ の条件を満たさねばならない. なお(3.2)には定数項がないので $E(\alpha_t)=0$ となるが, この条件はあまり制約を課したことにならない. 行列 $\sigma^2 P_0$ は $\alpha_t$ の無条件の共分散行列であり, $P_0$ は **2.7** 節の結果を使って導くことができる. 2章の(7.12)の記号を用いると, $\Omega=\sigma^2 RQR'$ となる. したがって,

$$\text{vec}(P_0)=[I-T\otimes T]^{-1}\,\text{vec}(RQR') \tag{4.2}$$

を得る.

次に与える例によって母数 $\sigma^2$ を導入した理由が明らかとなろう. 最尤推定の立場からは $\sigma^2$ について尤度関数を集約することには意味がある.

**例 1** MA(1)モデルについて遷移方程式は(1.7)で与えられる. 状態変数の初期値を $a_0=a_{1/0}$ とする. $\alpha_t=(y_t,\ \theta\epsilon_t)'$ であるから初期行列は直ちに求めることができ,

$$P_{1/0}=P_0=\sigma^{-2}E(\alpha_t\alpha_t')=\begin{bmatrix}1+\theta^2 & \theta\\ \theta & \theta^2\end{bmatrix} \tag{4.3}$$

となる. 第1期の予測誤差は $v_1=y_1$ でかつ $f_1=1+\theta^2$ である. 更新方程式を用いると,

$$a_1=\begin{pmatrix}y_1\\ \theta y_1/(1+\theta^2)\end{pmatrix}\quad\text{and}\quad P_1=\begin{pmatrix}0 & 0\\ 0 & \theta^4/(1+\theta^2)\end{pmatrix} \tag{4.4}$$

となり, $\alpha_2$ に対する予測方程式は

$$a_{2/1}=\begin{pmatrix}y_1\,\theta/(1+\theta^2)\\ 0\end{pmatrix} \tag{4.5}$$

及び

$$\begin{aligned}P_{2/1}&=\begin{pmatrix}\theta^4/(1+\theta^2) & 0\\ 0 & 0\end{pmatrix}+\begin{pmatrix}1 & \theta\\ \theta & \theta^2\end{pmatrix}\\ &=\begin{pmatrix}(1+\theta^2+\theta^4)/(1+\theta^2) & \theta\\ \theta & \theta^2\end{pmatrix}\end{aligned} \tag{4.6}$$

であり, かつ

$$v_2=y_2-\theta y_1/(1+\theta^2)\quad\text{and}\quad f_2=(1+\theta^2+\theta^4)/(1+\theta^2) \tag{4.7}$$



となる. この操作を繰り返し, カルマン・フィルターの漸化式によって予測誤差を計算すると,

$$v_t=y_t-\theta v_{t-1}/f_{t-1},\qquad t=1,\cdots,T \tag{4.8}$$

及び

$$f_t=1+\theta^{2t}/[1+\theta^2+\cdots+\theta^{2(t-1)}] \tag{4.9}$$

となる. ただしここで $v_0=0$ としておくことにする.

**例 2** 観測誤差を伴ったARMAモデル(1.9)の場合, モデルの母数と $h$ が与えられればカルマン・フィルターを用いて尤度関数を構成することができる. 初期条件は観測誤差なしのARMAモデルと全く同様に求められる. この問題への別の解法としては $y_t$ についてより一般のARMAモデルを求めることが考えられる. 2章 **2.5** 節で述べたように観測誤差を伴ったAR($p$)モデルはARMA($p$, $q$)モデルと同等である. ただ観測誤差を明示的に扱う利点としては次の二点を挙げることができよう. 第一に未知母数の数をより節約できること. 第二には(1.9)における変数 $w_t$ の情報を観測値 $y_t$ より抽出できることであろう. それは次節で述べる平滑化の方法によって行なわれる.

**例 3** $y_t$ を二つの観測不能な要素(unobserved components) $w_{1t}$, $w_{2t}$ の和であってそれぞれ $w_{it}$ は互いに独立なARMA過程に従うとしよう. この時観測方程式は

$$y_t=w_{1t}+w_{2t} \tag{4.10}$$

であって状態変数ベクトル $(\alpha_{1t}',\alpha_{2t}')'$ はそれぞれのARMA過程に対する適当な状態変数ベクトルとなる. 初期条件は

$$a_0=\begin{pmatrix}0\\ 0\end{pmatrix},\qquad P_0=\begin{pmatrix}P_0^{(1)} & 0\\ 0 & P_0^{(2)}\end{pmatrix} \tag{4.11}$$

と表現されるが, その意味は明らかであろう.

### 非定常モデル

もし $\alpha_t$ が非定常的である時は一般的には初期値は観測値自体から構成され得る. $N=1$ の場合にはしばしば一つの簡便な方法として最初の $m$ 個の観測値より初期値が設定される. 漸化式は $t=0$ 時点より $a_0=0$, $P_0=\kappa I$ (ただし $\kappa$ はある大きな数値)として始めることができるが, その場合尤度関数を作る際には $t>m$ における予測誤差のみを用いるべきであろう. この方法の応用例としては, **7.3** 節で述べる未知母数が酔歩(random walk)過程に従う回帰モデルを挙げることができよう.

## 4.5 予測，平滑化および定常状態

一期先予測量の構成はカルマン・フィルターによる推定問題の解法の一部分であった．ところでいったん時刻 $t=T$ における推定値を得ると，しばしば次に一期先以上の将来の予測が必要となる．すなわち一般多期間予測 (general multi-period prediction) の問題とは，$a_T$ と $P_T$ が与えられたときの $y_{T+l}$ $(l \geq 1)$ の最適推定量の導出である．これに対して平滑化とは，観測値がすべて与えられたときに $t=1, 2, \cdots, T$ に対する $\alpha_t$ の最適推定量を求める問題であるが，これは $a_T$, $P_{T/T}$ より始めて逆方向への漸化式を導くことによって解くことができる．本節の最後でさらに定常解の概念について述べる．本節を通して，他に特にことわりがなければ (1.1) と (1.2) で表現される一般状態空間モデルについて扱うことにしよう．

### 多期間予測

多期間予測の公式は (3.19) で与えられた単純予測の方程式を拡張することによって求められる．更新方程式を無視することとして $l$ 期先の状態変数ベクトルの最小平均二乗線型推定量は漸化式

$$a_{T+l/T} = T_{T+l} a_{T+l-1/T}, \qquad l=1,2,\cdots \tag{5.1}$$

で与えられる．ただし $a_{T/T}=a_T$ としよう．平均二乗誤差 (MSE) 行列は $P_{T/T}=P_T$ とすれば，

$$P_{T+l/T} = T_{T+l} P_{T+l-1/T} T'_{T+l} + R_{T+l} Q_{T+l} R'_{T+l}, \qquad l=1,2,\cdots \tag{5.2}$$

によって求められる．$y_{T+l}$ の最小平均二乗線型推定量は

$$\tilde{y}_{T+l/T} = Z_{T+l} a_{T+l/T} \tag{5.3}$$

で与えられ，予測誤差分散は

$$\mathrm{Var}(v_{T+l}) = \mathrm{MSE}(\tilde{y}_{T+l/T}) = Z_{T+l} P_{T+l/T} Z'_{T+l} + S_{T+l} H_{T+l} S'_{T+l} \tag{5.4}$$

となる．

このように $y_t$ の将来値の最適予測量及びその平均二乗誤差は容易に求めることができる．したがって実質的には新しく計算をつけ加える必要はない．予測問題を解く理論上の展開はそれ自体としてはほとんどなく，カルマン・フィルターによる推定問題の解法によってすでに与えられているといえよう．



**例 1**　一変量時系列 AR(1) モデルにおいては，$T_t=\phi$, $y_t=\alpha_t$ である．したがって (5.1) の表現を用いると，

$$\tilde{y}_{T+l/T} = \phi^l y_T, \qquad l=1,2,\cdots \tag{5.5}$$

となる．ここで $\sigma^2=1$ とおくと平均二乗誤差は漸化式

$$P_{T+l/T} = \phi^2 P_{T+l-1/T} + 1, \qquad l=1,2,\cdots \tag{5.6}$$

によって求められる．一期先予測誤差は単に $\epsilon_{T+1}=y_{T+1}-\phi y_T$ であって $P_{T+1/T}=1$ となる．(5.6) を繰り返し用いると，

$$P_{T+2/T} = \phi^2+1,$$
$$P_{T+3/T} = \phi^2(\phi^2+1)+1 = \phi^4+\phi^2+1$$

などとなる．一般には

$$P_{T+l/T} = \phi^{2(l-1)} + \phi^{2(l-2)} + \cdots + \phi^2 + 1, \qquad l=1,2,\cdots \tag{5.7}$$

である．$l \to \infty$ となるにつれて平均二乗誤差は $\sigma^2/(1-\phi^2)$ に近づくが，これは確率過程の分散に他ならない．したがって $l$ が大きければ将来の予測値は現在の観測値から隔ってゆき，有益な情報をもたらさないこととなる．

観測値に一部分欠落がある場合 (missing observation) には多期間予測は標本期間内においても有用となろう．その場合には一つの尤度関数が予測誤差分解を用いて作られる．例えば時刻 $t$ における AR(1) モデルの観測値が欠落しているとしよう．時刻 $t$ における予測誤差は尤度関数に入らないが，$v_{t+1}=y_{t+1}-\phi^2 y_{t-1}$ より $f_{t+1}=1+\phi^2$ となる．この点について詳しくは Harvey=Pierra (1980) 及び Jones (1980) を参照されたい．

### 平　滑　化

カルマン・フィルターでは各時点において利用可能な情報の下で $\alpha_t$ の最小平均二乗線型推定量を構成する．したがって標本情報のすべてを利用する推定量は最終時刻における状態変数の推定量 $a_T$ のみである．そこで平滑化方程式は $a_T$ と $P_T$ より始め，逆方向に作用させる．

ここで $a_{t/T}$ と $P_{t/T}$ をそれぞれ時刻 $t$ における平滑化推定量とその共分散行列とすると平滑化方程式は，

$$a_{t/T} = a_t + P^*_t (a_{t+1/T} - T_{t+1} a_t), \tag{5.8 a}$$
$$P_{t/T} = P_t + P^*_t (P_{t+1/T} - P_{t+1/t}) P^{*\prime}_t \tag{5.8 b}$$

と表現される．ただしここで

$$P^*_t = P_t T'_{t+1} P^{-1}_{t+1/t}, \qquad t=T-1,\cdots,1 \tag{5.8 c}$$

であって，しかも $a_{T/T}=a_T$, $P_{T/T}=P_T$ とする．平滑化は時刻 $t=2, \cdots, T$

のどの時点からでも実用可能であるが，統計的な応用では時刻 $t=T$ からの平滑化が主たる関心事となろう．

平滑化推定量が得られると直接残差(direct residual)とでも呼ぶべきものが得られる．この残差は $N\times 1$ ベクトル

(5.9) $$e_t = y_t - Z_t a_{t/T}, \qquad t=1,\cdots,T$$

で与えられるが，平均はゼロで共分散行列は $Z_t P_{t/T} Z_t' + S_t H_t S_t'$ となる．この直接残差は(3.21)で与えられる予測誤差からの残差(prediction error residuals)と区別されるべきものである．予測残差はもし状態空間モデルの母数が既知であれば時間について互いに無相関，すなわち $t \neq s$ に対して $E(v_t v_s') = 0$ である．直接残差に関しては一般にこの性質がない．

**例2** 次の二式によって定められるモデルを考えよう．

(5.10 a) $$y_t = \alpha_t + \xi_t, \qquad \xi_t \sim \mathrm{WN}(0,\sigma^2),$$

(5.10 b) $$\alpha_t = \alpha_{t-1} + \eta_t, \qquad \eta_t \sim \mathrm{WN}(0,\sigma^2 q).$$

ここで状態変数 $\alpha_t$ と観測値 $y_t$ はともにスカラーである．状態変数は酔歩(ランダム・ウォーク)過程に従っているが観測不能な攪乱項 $\xi_t$ が加わり直接観測されない．すなわち，これは工学研究者の間で「記号・雑音の加法モデル」と呼ばれるものの中での簡単な例である．ここで信号と雑音の分散比は既知としておこう．このモデルにおける予測方程式は

(5.11 a) $$a_{t/t-1} = a_{t-1},$$

(5.11 b) $$P_{t/t-1} = P_{t-1} + q$$

となり，更新方程式は

(5.12 a) $$a_t = a_{t/t-1} + P_{t/t-1}(y_t - a_{t/t-1})/(P_{t/t-1}+1)$$

及び

(5.12 b) $$P_t = P_{t/t-1} - P_{t/t-1}^2/(P_{t/t-1}+1)$$

で与えられる．ここで $T=4$ の観測値 $y_1$, $y_2$, $y_3$, $y_4$ が与えられ，しかも $\alpha_0$ の平均は $a_0$，共分散行列は $\sigma^2 P_0$ としよう．$a_0=4$, $P_0=12$, $q=4$ としたときの観測値が表4.1に与えられている．$y_1$ が与えられた時の $\alpha_1$ の最小平均二乗線型推定値は，

$$a_1 = 4 + (12+4)(4.4-4.0)/(12+4+1) = 4.376$$

であって，しかも

$$P_1 = 16 - 16^2/17 = 0.941$$

となる．観測方程式において $z_t=1$ であるから $y_t$ の最小平均二乗線型推定量はすべての $t$ に対して $a_{t/t-1}$ と等しくなる．したがって $\tilde{y}_{2/1} = a_1 = 4.376$ となる．この計算



**表4.1** 平滑化推定値と残差

| $t$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $y_t$ | 4.4 | 4.0 | 3.5 | 4.6 |
| $a_t$ | 4.376 | 4.063 | 3.597 | 4.428 |
| $P_t$ | 0.941 | 0.832 | 0.829 | 0.828 |
| $v_t$ | 0.400 | −0.376 | −0.563 | 1.003 |
| $a_t/T$ | 4.306 | 4.007 | 3.739 | 4.428 |
| $P_t/T$ | 0.785 | 0.710 | 0.711 | 0.828 |
| $e_t$ | 0.094 | 0.007 | −0.239 | 0.172 |

を $t=2$, 3, 4 について繰り返した結果が表4.1に与えられている．

最終期の推定値は $a_4 = 4.428$, $P_4 = 0.828$ である．次にこの値を平滑化アルゴリズム(5.8)の初期値とする．その結果(5.8 a)は

$$a_{t/T} = a_t + P_t P_{t+1/t}^{-1}(a_{t+1/T} - a_t), \qquad t = T-1,\cdots,1$$

となるので，

$$a_{3/4} = 3.597 + 0.829(4.428-3.597)/4.829 = 3.739$$

かつ

$$P_{3/4} = 0.829 + (0.828-4.829)(0.829/4.829)^2 = 0.711$$

となる．直接残差と予測残差はともに表4.1に与えられているが，この例では(5.9)と(3.21)はそれぞれ $e_t = y_t - a_{t/T}$ と $v_t = y_t - a_{t-1}$ と簡単化される．

## 定常解

ところでもし $Z_t, S_t, T_t, R_t, H_t, Q_t$ などの母数が時間について不変であるならば，カルマン・フィルターはしばしば $P_t$ と $P_{t/t-1}$ が時間に関して不変となるある定常状態へと近づいてゆく．その収束についての一般的条件については Jazwinski (1970) 及び Anderson=Moore (1979) を参照するとよい．ここでは次に二つの例をあげておくことのみにとどめよう．

**例3** 表4.1において $P_t$ の動きに注意すると(5.10)で表わされているシステムはある定常状態へ急速に近づいている．(5.11 b)と(5.12 b)の予測方程式と更新方程式を組み合わせると，

(5.13) $$P_t = P_{t-1} + q - (P_{t-1}+q)^2/(P_{t-1}+q+1)$$

となる．ここで $P_t = P_{t/t-1} = \bar{P}$ と置き，二次方程式を解くと $\bar{P} = 0.828$ となるが，これは $P_4$ の数値でもある．一度定常状態に到達すると $P_t$ に関する更新方程式と予測方程式はともに不要となる．

**例 4**　本章 **4.4** 節の例 1 で考察した MA(1) モデルにおいて，行列 $P_{t|t-1}$ は

$$P_{t|t-1}=\begin{bmatrix} f_t & \theta \\ \theta & \theta^2 \end{bmatrix}, \qquad t=1,2,\cdots \tag{5.14}$$

によって与えられる．(4.3) と (4.6) を参照のこと．もし $|\theta|\leq 1$ ならば (4.9) より明らかに

$$\lim_{t\to\infty} f_t = 1 \tag{5.15}$$

が成立する．したがって漸化式はある定常状態へ近づくことになる．その意味については次章であらためて吟味されることになる．

## 付論：行列の逆転公式

$A$ と $C$ をそれぞれ $n\times n$ と $m\times m$ の非特異行列，$B$ を $n\times m$ 行列としよう．さらに $n\times n$ 行列 $D$ を

$$D=[A+BCB']^{-1} \tag{A.1}$$

としよう．このとき

$$D=A^{-1}-A^{-1}B[C^{-1}+B'A^{-1}B]^{-1}B'A^{-1} \tag{A.2}$$

が成立する．このことは $DD^{-1}=I$ となることを直接示すことによって証明が与えられる．すなわち $E^{-1}=C^{-1}+B'A^{-1}B$ と置けば，

$$\begin{aligned} D^{-1}D &= I-BE^{-1}B'A^{-1}+BCB'A^{-1}-BCB'A^{-1}BE^{-1}B'A^{-1} \\ &= I+[-BE^{-1}+BC-BCB'A^{-1}BE^{-1}]B'A^{-1} \\ &= I+BC[-C^{-1}E^{-1}+I-B'A^{-1}BE^{-1}]B'A^{-1} \\ &= I+BC[I-EE^{-1}]B'A^{-1}=I \end{aligned}$$

である．ところで行列 $B$ が $n\times 1$ ベクトル $b$ であって，しかも $C=1$ とすると公式

$$(A+bb')^{-1}=A^{-1}-\frac{A^{-1}bb'A^{-1}}{1+b'A^{-1}b} \tag{A.3}$$

が導かれる．行列の逆転についての種々の有用な公式は Jazwinski (1970, pp.261-262) に与えられている．

## ノート

状態空間モデルについての基本論文は Kalman (1960) であるが，**4.2** 節および **4.3** 節の議論は Duncan=Horn (1972) に基づくものである．予測誤差分解と最尤推定については特に Schweppe (1965) 及び Harvey=Phillips (1976 a, 1979) を参照されたい．カルマン・フィルターに基づくベイズ予測への興味深いアプローチとしては Harrison=Stevens (1976) がある．



## 練習問題

**1.** 次のモデルを考えよう．

$$\begin{aligned} y_t &= 4+\alpha_t+\xi_t, & \xi_t &\sim \mathrm{WN}(0,\sigma^2), \\ \alpha_t &= 0.5\,\alpha_{t-1}+\eta_t, & \eta_t &\sim \mathrm{WN}(0,4\sigma^2). \end{aligned}$$

$T=4$ であって観測値が表 4.1 で与えられ，しかも $a_0=4$，$P_0=12$ とする．このとき $a_t$，$P_t$，$v_t$ $(t=1,2,3,4)$ を計算し，尤度関数を評価せよ．

**2.** (3.2) で与えられる遷移方程式を修正して

$$\alpha_t=T_t\alpha_{t-1}+B_t x_t+\eta_t, \qquad t=1,\cdots,T$$

としよう．ただし $x_t$ は既知の変数ベクトルであり $B_t$ は既知の行列とする．このときカルマン・フィルターの方程式はどのような影響を受けるであろうか．

**3.** 観測方程式が (5.10 a) で与えられ，遷移方程式が

$$\alpha_t=(-1)^{2t-1}\alpha_{t-1}, \qquad t=1,2,3,\cdots$$

で与えられる状態空間モデルを考える．ただし $\alpha_0\neq 0$ としよう．この時定常状態における行列 $P_t=\bar{P}$ はゼロ行列であることを示し，その結果についてコメントを加えよ．

# 5章
# 自己回帰移動平均モデルの推定

## 5.1 はじめに

本節では2章(1.25)の形で与えられるARMA$(p, q)$モデルの推定を取り扱う．モデルは$p$と$q$が既知であるという意味において正しく定式化されていると想定する．適当な$p$および$q$の値を決定すること，および誤った定式化をした場合の結果の評価にまつわる問題は次章で考察される．

本章を通じて，観測量は正規分布に従うと想定される．したがって最尤推定の原理によって，本質的には残差二乗和関数を最小にする方法に基づく手続きが導かれることになる．これは推定の立場からは非常に便利なことであるが，もし正規性の仮説が疑わしい場合には，結果としてもたらされる推定量は頑健性(robustness)を欠くという批判にさらされることになる．この点については例えばHarvey(1981 a，3.7節)の議論を参照されたい．

**5.4**節において推定量の大標本特性を議論するに当っては，正しいモデルが定常的であり，かつ反転可能であると想定される．これらの仮定は，**5.5**節において検定手続きが導かれるときにも想定される．しかしながら，MA多項式の根のいくつかが単位円上にあるようなモデルは意味があり，実際に階差のとりすぎが起ったときに生じる可能性があるということに注意する必要があろう．簡単な例として$y_t = \epsilon_t$の場合を考えてみよう．ここで$\epsilon_t \sim NID(0, \sigma^2)$である．このとき以下の式が成り立つ．

$$\Delta y_t = \epsilon_t + \theta \epsilon_{t-1}. \tag{1.1}$$

ここで$\theta = -1$である．したがって$\Delta y_t$は厳密にいって反転不能のMA(1)過

程から生成されることになる．すなわちAR母数が定常性の条件を満たすように要求することは意味のあることであるけれども，同じことは反転可能性については必ずしも正しくないのである．この問題については**5.3**節でより深く論じる．他方，モデル選択におけるその意味については次章で取り上げられるであろう．

厳密な尤度関数は**1.2**節における予測誤差分解から構成されるが，これは通常の慣行からは少し異なっている．しかしながら予測誤差分解は，理論的観点からは自然なアプローチであり，この方法によれば厳密な尤度関数の構造が明らかになるばかりか，条件付尤度関数に組み込まれている近似の本質が完全なまでに明らかになる．条件付尤度関数は，特定の初期の攪乱項または観測値を固定することによって得られ，条件付尤度関数の最大化は対応した二乗和関数の最小化に等しくなる．結果として得られる推定量は，最小二乗推定量として知られているものであるが，厳密最尤推定量の適切な近似としての役割を果すと一般的に考えられている．

尤度関数についての別のアプローチはコレログラムについての推定値に基づくものである．純粋のARモデルに関しては，コレログラムから導かれた推定値はきわめて有力なものである．このことはMAモデルや混合モデルの場合にはもはや成立しないが，コレログラムから求められる推定値は逐次最適化法にとっての便利な初期推定値として用いることができる．

最後に本章で考慮の対象となるモデルについて，平均ゼロと仮定することについてであるが，これは特に制約的なものではない．ゼロでない平均の導入は，$\phi$および$\theta$の推定量の作成またはそれらの大標本特性にとって何らの重要な影響も及ぼさない．さらに標本平均$\bar{y}$は漸近的に平均$\mu$の有効推定量である．このことは次節においてAR(1)モデルについてはっきりと示される．そこでの最終的な結論とは，任意のARMAモデルにおける$\phi$や$\theta$は，観測値を平均からの偏差の形に置くことによって推定できるということである．

## 5.2 自己回帰モデル

AR(1)過程についての尤度関数は1章(2.26)において導かれた．AR($p$)過程に関しても同じ方法を用いることができ，その第一段階は以下のような予測誤差の分解を行なうことである．



$$\log L(y)=\sum_{t=p+1}^{T}\log(y_t/y_{t-1},\cdots,y_1)+\log L(\boldsymbol{y}_p). \tag{2.1}$$

この表現は1章(2.19)と同様なものである．上式(2.1)の右辺の第一項は，$\epsilon_{p+1},\cdots,\epsilon_T$の結合分布の対数であると解釈することができる．他方，$L(\boldsymbol{y}_p)$は最初の$p$観測値$\boldsymbol{y}_p=(y_1,\cdots,y_p)'$の結合分布である．もし$\boldsymbol{y}_p$の共分散行列が$\sigma^2V_p$で表わされるならば，対数尤度関数は次のように表わされる．

$$\log L=-(1/2)T\log 2\pi-(1/2)T\log\sigma^2-(1/2)\log|V_p| \\ -(1/2)\sigma^{-2}\left[\boldsymbol{y}_p'V_p^{-1}\boldsymbol{y}_p+\sum_{t=p+1}^{T}(y_t-\phi_1y_{t-1}-\cdots-\phi_py_{t-p})^2\right]. \tag{2.2}$$

### 厳密最尤推定

母数$\sigma^2$については尤度を集約化することにより尤度関数から取り除くことができる．そのようにして得られる尤度関数は，$\phi$に関して依然として非線型である．したがって最尤推定は数値的最適化によってなされなければならない．行列$V_p^{-1}$を直接的に求めるためにはいくつかの方法が利用可能である．この点についてはたとえばBox=Jenkins(1976)またはGalbraith=Galbraith(1974)を参照されたい．

AR(1)モデルに関しては，$\partial\log L/\partial\phi=0$なる式は$\phi$についての三次関数となる．これは$[-1,1]$の範囲で単一の実数根を持つ．この根を直接的に計算する方法はBeach=Mackinnon(1978)に与えられている．

### 回 帰 法

ARモデルの推定は，最初の$p$個の観測値$y_1,\cdots,y_p$を固定されたものとみなすことによって非常に簡単化される．この考えによって(2.1)における最後の項を落とすことの理論的正当化が行なわれよう．そして尤度関数を最大化することが，以下の二乗和関数を最小化するのと等価であることになる．

$$S(\phi)=\sum_{t=p+1}^{T}(y_t-\phi_1y_{t-1}-\cdots-\phi_py_{t-p})^2. \tag{2.3}$$

すなわち$\phi$の最尤推定量は，$y_t$をそのラグ付き変数$y_{t-1},\cdots,y_{t-p}$上に最小二乗回帰したものによって得ることができる．

大標本においては推定が厳密最尤法でなされるか最小二乗法でなされるかによってはほとんど差がない．AR(1)モデルにおいては，たとえば1章(2.26)か

らわかるように，厳密な尤度関数の顕著な特徴は尤度に $y_1^2$ および $\log(1-\phi^2)$ なる項を含んでいるということのみであり，この二項は尤度関数の残りの項によって支配される．そしてこの二項が取り除かれたとしても，$\phi$ および $\sigma^2$ の推定量の漸近分布は影響を受けない．

AR(1)モデルに1章(1.1)のようにゼロでない平均を導入することによって，以下のような二乗和関数が得られる．

$$S(\phi)=\sum_{t=2}^{T}[y_t-\mu-\phi(y_{t-1}-\mu)]^2.$$

これを $\mu$ に関して微分することにより以下の結果を得る．

$$\tilde{\mu}=\frac{\sum_{t=2}^{T}y_t-\tilde{\phi}\sum_{t=1}^{T-1}y_t}{(T-1)(1-\tilde{\phi})}\simeq\bar{y}.$$

このようにして $\mu$ は $\phi$ とは独立に推定することができる．

### ユール=ウォーカー方程式

母数 $\phi$ の推定値はまたコレログラムからも得ることができる．純粋な AR($p$)モデルに関しては，自己相関関数は2章(2.37)の $p$ 次の定差方程式で与えられる．この式を $\tau=1,\cdots,p$ について書き出すことにより，以下の表現を得る．

$$(2.4)\qquad \begin{pmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \cdots & \rho(p-2) \\ \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{pmatrix}\begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{pmatrix}=\begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(p) \end{pmatrix}$$

この表現はユール=ウォーカー(Yule=Walker)方程式として知られている．上式において，$\rho(\tau)$ を $r(\tau)$ に置き換えることにより線型方程式の組が得られ，それを解くと $\phi_1,\cdots,\phi_p$ の推定値が直接的に求められる．

**例1** AR(2)モデルについては以下のように求められる．

$$(2.5)\qquad \hat{\phi}_1=\frac{r(1)[1-r(2)]}{1-r^2(1)},\quad \hat{\phi}_2=\frac{r(2)-r^2(1)}{1-r^2(1)}.$$

「末端の観測値の影響」を除いては，このようにして得られた推定量は $y_t$ を $p$ 個のそれ自身の過去の値に回帰して得られたものと同一である．したがってユール=ウォーカー方程式から得られた推定量は，最尤推定量と同じ漸近分布を持つことになる．



## 5.3 移動平均過程および混合過程

任意の ARMA($p,q$)過程の尤度関数は予測誤差分解から作り出すことができる．もし必要ならばモデルを状態空間形式に当てはめ，予測誤差はカルマン・フィルターによって計算することができる．MA(1)モデルに対してこの手法を適用した例が，**4.4**節に与えられている．しかしながら，尤度関数を計算する方法を見い出すことは計算手順の第一段階にすぎない．なぜなら尤度関数はその後にその要素である $\phi$ および $\theta$ に関して最大化されなければならないからである．この最大化はとくに，もし $p+q$ が大きい場合には非常に時間を浪費するものとなりかねないであろう．

もし攪乱項または観測値の初期値についてある特定の仮定を立てられるのであれば，条件付(conditional)尤度関数が得られる．純粋な AR 過程の場合においては，仮定の下では $\phi$ についての線型な最尤推定量が導かれる．MA 過程や混合過程の場合には，最尤推定量は相変わらず非線型であるが，尤度関数の計算はかなり簡単になる．さらに解析的な微係数が簡単に利用可能となり，それが最適化手続きの効率性を改善するのに重要な役を果たすことになる．

### MA(1) モデルの条件付二乗和

MA(1)モデル

$$(3.1)\qquad y_1=\epsilon_t+\theta\epsilon_{t-1},\qquad t=1,\cdots,T$$

は，純粋に自己回帰型ではないようなモデルを推定する場合において用いられる方法をもっとも簡単に例示する．ここで当分の間 $\theta$ が既知であると想定しよう．もし時点 $t-1$ において $\epsilon_{t-1}$ が既知であれば，$y_t$ の最小平均二乗推定量は $\hat{y}_{t/t-1}=\theta\epsilon_{t-1}$ となる．そして対応する予測誤差は以下のようになる．

$$(3.2)\qquad \epsilon_t=y_t-\theta\epsilon_{t-1}.$$

将来の予測誤差はそれぞれ対応する攪乱項と同じものであるが，(3.2)の漸化式から計算することができる．しかしながらこの方法には一つの困難な問題が存在する．すなわち予測誤差の系列 $\epsilon_1,\cdots,\epsilon_T$ を計算するためには，最初の攪乱項 $\epsilon_0$ が既知であると想定する必要があることである．この想定の下では予

測誤差はすべて計算することができる．ここでさらに$\epsilon_0$がすべての実現系列においてゼロに固定されると想定すれば，対数尤度関数は1章(2.22)のような形をとる．多くの実際的目的に関する限り，仮定$\epsilon_0=0$が正しいか否かはほとんど差を生じない．しかしながら，もし$\epsilon_0$が確率変数であるなら，条件付尤度関数に暗黙のうちに用いられている近似は確率過程が反転可能である，すなわち$|\theta|<1$のときのみ有効であるということに注意を要する．というのは，反転可能である限りにおいては，厳密な尤度関数と条件付尤度関数の差は，$T$が増大するに従って無視し得るものとなるからである．

条件付尤度関数を最大化することは，条件付二乗和(conditional sum of squares：CSS)関数$S(\theta)=\sum\epsilon_t^2$を$\theta$に関して最小化することと同値である．ここで$\epsilon_t$はもはや攪乱項ではなく，$\theta$に依存した残差であることに注意を要する．記号$\epsilon_t(\theta)$はこの点を強調している．もっとも$\epsilon_t$の解釈は文章の内容から明らかなので，通常はこのような修正を行なう必要はない．

既知のある値$T$について，$T$個の残差の一組が次のような漸化式から$\epsilon_0=0$を出発として与えられる．

$$\epsilon_t=y_t-\theta\epsilon_{t-1},\qquad t=1,\cdots,T. \tag{3.3}$$

これを用いて$S(\theta)$を作り出すことができる．残された唯一の問題は，$\theta$に関してこの関数を最小化にする適切な方法を探すことである．この場合ただ一つの母数が含まれるので，$[-1,1]$の範囲の上を格子状に探索することができる．しかしながらより一般的な確率過程にとってはこの方法は有効ではなく，用いるべき周知のアルゴリズムは，ガウス=ニュートン(Gauss=Newton)法かまたはそれの修正法である．MA(1)モデルについては，(3.3)を微分すれば以下の結果を得る．

$$\frac{\partial\epsilon_t}{\partial\theta}=-\theta\frac{\partial\epsilon_{t-1}}{\partial\theta}-\epsilon_{t-1},\qquad t=1,\cdots,T. \tag{3.4}$$

なお$\epsilon_0=0$であるから，$\partial\epsilon_0/\partial\theta=0$となる．このように微係数についても，その初期値設定と同様なやり方で処理することによって，(3.3)と同様な漸化式によって作り出すことができる．このアルゴリズムでは，$\theta$のある推定値が与えられたときにまず$\epsilon_t$および$\partial\epsilon_t/\partial\theta$を計算し，$\epsilon_t$の$\partial\epsilon_t/\partial\theta$上への回帰により推定値を更新する手続きが行なわれる．



### 一般の場合の条件付二乗和

高次のMAモデルに関しては，条件付尤度関数は$\epsilon_{1-q},\cdots,\epsilon_0$をすべての実現値についてゼロと置くことによって与えられる．次に，条件付二乗和関数を計算するために用いる残差は，以下の漸化式によってもたらされる．

$$\epsilon_t=y_t-\theta_1\epsilon_{t-1}-\cdots-\theta_q\epsilon_{t-q},\qquad t=1,\cdots,T. \tag{3.5}$$

ここで$\epsilon_{1-q}=\epsilon_{2-q}=\cdots=\epsilon_0=0$である．

同様の手続きは混合モデルの場合にも適用できる．もっとも混合モデルの場合には，初期の観測値の取り扱いという追加的問題が生じる．ここで2章(4.1)で与えられたARMA(1.1)モデルを考えよう．もし$y_1$が固定されたものとみなし得るならば，予測誤差は以下の漸化式から$\epsilon_1=0$として計算することができ，

$$\epsilon_t=y_t-\phi y_{t-1}-\theta\epsilon_{t-1},\qquad t=2,\cdots,T. \tag{3.6}$$

もう一つの方法は$y_0$および$\epsilon_0$をゼロと置いて，逐次式を$t=1$からスタートさせる方法である．この方法によれば$T-1$でなく$T$個の残差をもたらすけれども，この方法を推薦することはできない．なぜなら任意に$y_0=0$と置くことは計算に歪みを持ちこむことになるからである．一般にARMA$(p,q)$モデルについての適切な方法は，以下のような漸化式から$T-p$の予測誤差を計算することである．

$$\epsilon_t=y_t-\phi_1y_{t-1}-\cdots-\phi_py_{t-p}-\theta_1\epsilon_{t-1}-\cdots-\theta_q\epsilon_{t-q},\qquad t=p+1,\cdots,T. \tag{3.7}$$

ただしここで$\epsilon_p=\epsilon_{p-1}=\cdots=\epsilon_{p-q+1}=0$と置くことにする．

ガウス=ニュートン法を用いるのに必要な微係数もやはり漸化式によって計算することができる．ARMA(1,1)モデルについては，それらは以下の式で与えられる．

$$\frac{\partial\epsilon_t}{\partial\phi}=-y_{t-1}-\theta\frac{\partial\epsilon_{t-1}}{\partial\phi},\qquad t=2,\cdots,T, \tag{3.8 a}$$

$$\frac{\partial\epsilon_t}{\partial\theta}=-\theta\frac{\partial\epsilon_{t-1}}{\partial\theta}-\epsilon_{t-1},\qquad t=2,\cdots,T. \tag{3.8 b}$$

ここで$\partial\epsilon_1/\partial\phi=\partial\epsilon_1/\partial\theta=0$とする．一般的には$p+q$個の微係数が必要となるであろう．これらの値は二度の漸化式に基づいた計算により作ることができる．一度はAR部分についてであり，もう一度はMA部分についてである．

この方法について詳しくは Box=Jenkins (1976, p. 237) を参照されたい.

### 厳密最尤推定*

MA (1) モデルについての条件付二乗和を計算するに当っては，最初の撹乱項 $\epsilon_0$ は定義上ゼロに設定される．この場合の厳密な尤度関数を評価する伝統的アプローチでは，$\epsilon_0$ をデータから推定するが，この方法は少なくとも観測値を通る予測と逆予測 (backcasting) なる二つの経路を必要とする．もう一つのアプローチとしては厳密な尤度関数を予測誤差分解を基礎にして得る方法がある．モデルを状態空間形式に当てはめることにより，尤度関数をカルマン・フィルターに基づく単一の経路から計算することができる．

MA (1) 過程の状態空間表現は 4 章 (1.7) および (1.8) で与えられた．さらに尤度関数の評価については **4.4** 節で詳しく論じた．4 章において (4.4) から (4.9) にかけて詳しく吟味したことは教育的意味があって，条件付尤度と厳密な尤度の間の関係についての洞察を得ることができた．どちらの方法においても最初の予測誤差は $y_1$ である．条件付二乗和においては $\epsilon_0 = 0$ という仮定から $\mathrm{Var}(y_1) = \mathrm{Var}(\epsilon_1) = \sigma^2$ を意味する．他方 $\epsilon_0$ が確率変数であると考えるならば，$\mathrm{Var}(y_1) = \sigma^2(1-\theta^2)$ となって，$f_1$ に反映されている．$\epsilon_0$ の分布を考えることは，二番目の予測誤差，すなわち 4 章 (4.7) と，それに伴った分散についてもまたその影響が現われる．4 章 (4.8) で与えられた漸化式に関していえば，条件付二乗和における簡単化の仮定は，$f_{t-1}$ をすべての $t$ について 1 と置くことを意味している．

4 章 (5.14) において $|\theta| \leq 1$ なる MA (1) モデルについては，$t$ が増大するにつれて $f_t$ は 1 に近づくことが観察された．したがって十分に大きな $t$ の値については，4 章 (4.8) で与えられた予測誤差の漸化式は，条件付二乗和の漸化式 (3.3) によって近似されよう．この性質を Gardner=Harvey=Phillips (1980) による方法では利用している．すなわちその方法では $f_t$ に注目し，いったんそれが 1 に十分近くなったならば，アルゴリズムは (3.3) の形の漸化式によって予測誤差を評価するように切り換えるのである．さてここでいわれている意味で十分に近いことを，小さな数 $\delta$ で表わすとしよう．すると切り換えは $f_t < 1+\delta$ のときに行なわれることになる．$\delta$ をより小さくすることは近似の誤差を小さくする効果を持つけれども，切り換えを行なう時点 $t$ の値 $t^*$ が増大するという損失をもたらす．この $\delta$ を選択するためのある手掛りは Gardner=Harvey=Phillips (1980) に与えられている．同論文の表 I は $t^*$ と近似の誤差の間のトレード・オフを示し，さらに $\delta = 0.01$ と置くことが適当な妥協点であると結論されている．



4 章 (4.9) から生じるもう一つの注目すべき点は，$|\theta| > 1$ のときには 4 章 (5.14) はもはや成立しないことである．なぜなら以下のような不等号が成立するからである．

$$\lim_{t\to\infty} f_t = \theta^2 > 1, \qquad |\theta| > 1. \tag{3.9}$$

したがって $f_t$ に注目することによって反転不能な解を見つけ出すことができる．もし $|\theta| = 1$ である場合にはその確率過程は厳密にいって反転不能であるが，$f_t$ は $t \to \infty$ のときに 1 に収束する．$|\theta| = 1$ は MA 過程として完全に有効であるから，これはもっともなことである．この問題については，たとえば本章 (1.1) を参照されたい．

カルマン・フィルターのアルゴリズムは，任意の ARMA $(p, q)$ モデルに適用可能である．しかしながら厳密な尤度関数を計算することは，条件付二乗和を計算するよりも必ずより時間浪費的である．さらに条件付二乗和の場合には，解析的な微係数が簡単に利用可能であることは，その数値的最適化が一般的により効率的になされることを意味している．ただし，とくに MA 母数が反転可能領域の境界の近くまたはその上にある場合には，厳密な最尤推定の方がいくつかの統計的利点を持つかもしれない．この主張についての若干の証拠は **5.6** 節に示されている．

### 初期推定値

ガウス=ニュートン反復法には出発値 (初期推定値) が必要である．MA (1) モデルでは，$\hat{\theta} = 0$ と置いて始めることが可能であろう．しかし同様の戦略は，混合モデルにおいて採用することはできない．例えば ARMA (1, 1) モデルにおいて $\hat{\phi} = \hat{\theta} = 0$ と置くことは，二つの微係数を等しく置くこととなり，$\epsilon_t$ の $\partial\epsilon_t/\partial\phi$ と $\partial\epsilon_t/\partial\theta$ 上への回帰に際して完全な多重共線性が生じることによってアルゴリズムは直ちに停止してしまう．

反復を始めるより有力な方法は母数の一致推定量を用いて始めることであり，コレログラムから得た推定値を用いることが一つの可能性として考えられる．例えば，MA (1) モデルを考えよう．2 章 (1.15) において，$\rho(1)$ を $r(1)$ で置き換

えれば，$\theta$ の推定値は以下の二次方程式を解くことによって得られる．

$$\hat{\theta}^2 r(1) - \hat{\theta} + r(1) = 0. \tag{3.10}$$

この方程式は二つの解を持つ．しかしながら

$$r^{-1}(1) = \hat{\theta} + \hat{\theta}^{-1}$$

であるから，一つの根は明らかにもう一つの逆数となっている．いずれの根を選べばよいかを決める問題は，反転可能性の条件によって解決することができる．したがって $\theta$ の初期推定量は以下のように与えられる．

$$\hat{\theta}_c = \{1 - [1 - 4r^2(1)]^{1/2}\}/2r(1). \tag{3.11}$$

ここでもう一つの解 $\hat{\theta}_c^{-1}$ は除外されているが，それは絶対値で 1 より大きい値を持つからである．

なお (3.10) が相異なる解を持たない唯一の場合は，$r(1) = \pm 0.5$ となる場合であり，そのとき推定量は $\hat{\theta}_c = \pm 1$ である．もし $|r(1)| > 0.5$ であるなら，(3.10) に対する実数の解は存在しない．なぜなら MA (1) モデルについての理論的な一次の自己相関は 0.5 を越えることはないからである．小標本においては，このような値が標本変動のために MA (1) 過程からも生じ得るが，もしこのような値の $r(1)$ を得たときには MA (1) モデルを用いるべきでないことを示唆していると通常は考えられている．

同様の方法を ARMA (1, 1) モデルの母数を推定するためにも用いることができよう．自己相関関数は 2 章 (4.13) で与えられている．したがって $\phi$ を以下のような方法で推定することが示唆される．

$$\hat{\phi}_c = r(2)/r(1). \tag{3.12}$$

もし 2 章 (4.13 a) における $\rho(1)$ および $\phi$ が $r(1)$ および $\hat{\phi}_c$ で置き換えられるならば，推定量は再び二次方程式の解として与えられる．徐々に複雑になるけれども，このアプローチは高次のモデルへも一般化が可能である．詳しくは Godolphin (1976) を参照されたい．初期値を得る他の方法は Wilson によって開発されたが，その方法は Box=Jenkins (1976, pp. 201-205) によって説明されている．

一般にコレログラムから一致推定量を得ることができるが，もしモデルに MA 部分が含まれている場合には，それらは漸近的に有効な推定量ではない．MA (1) の場合においては，たとえば $\theta = \pm 0.5$ のとき推定効率は $\mathrm{Eff}(\hat{\theta}_c) = 0.49$ である．$|\theta|$ が 1 に近づくとき，$\mathrm{Eff}(\hat{\theta}_c)$ は 0 に近づく．このことは，高次の標本自己相関が $r(1)$ に現われない情報を含んでいるということを意味している．



### 定常性と反転可能性*

AR (1) 過程については 1 章 (2.26) から明らかなように，$|\phi| \geq 1$ の場合には尤度関数が定義できない．結果として AR (1) 過程については厳密最尤推定量は常に単位円の中にくる．これに対してもし推定量が他の方法で計算された場合には，その値が定常性の条件を満たしているという保証はもはやない．たとえば Harvey=Phillips (1977) は，$\phi = 0.9$ の場合について最小二乗推定量は 200 回の繰り返しのうち，5 回ほど 1 より大きな値を生じることを報告している．

高次の AR モデルや混合モデルの場合には，厳密最尤法が用いられる時にも推定値に定常性の条件を満たすとは限らない．そこで AR 母数のどのような推定値も定常性の条件を満たしているか否かチェックされるべきである．もし反復最適化の方法が用いられるのであれば，各段階において計算される推定値に目を光らせるのが望ましいであろう．他の方法は制約付最適化法を用いることである．

MA 母数に関する反転可能条件が犯されることはやや異なった意味を持つ．反転不能な MA 過程または ARMA 過程は，相変わらず尤度関数がうまく定義できる定常確率過程である．しかしながら反転不能な推定値を許すということは，尤度方程式が複数の解を持つことを意味し，そこで母数識別の問題が生じる．ところで，このことは MA (1) 過程によってうまく例示される．この場合には単位円周外の $\theta$ の値に対して，同じ自己相関関数を持つ単位周円内の $\theta$ が存在する．これは 2 章 (1.15) が，$\theta$ がその逆数 $1/\theta$ に置き換えられても同じ値を持つことからきている．このことは一般に厳密な尤度関数は二つの最大値 (一つは $\tilde{\theta}$ において，もう一つは $1/\tilde{\theta}$ において) を持つことを意味しているが，この法則についての唯一の例外は，尤度の最大値がちょうど単位円周上，すなわち $\tilde{\theta} = \pm 1$ で生じる場合である．$\theta$ の推定値が反転可能の条件を満たすように制約を加えることによって，多数根から生じる解の不明瞭さは排除される．なお反転不能な推定値を取り除く根本的な理由は，その推定値が非効率な予測をもたらすからである．この問題に関しては **6.4** 節を参照されたい．

厳密最尤推定を行なうに当って，反転可能性を取り扱う問題については，基本的に二つのアプローチが考えられる．第一の方法は制約なしの最適化法を採用することである．反転可能領域の外で最大値が探し出された場合には，反転

可能領域の中での対応する最大値を決定しなければならない．しかしながらこのようなアプローチが採用される場合には，もし反転不能な推定値が繰り返し計算の途中に生じたときには，丸めの誤差の影響を受けないような方法によって尤度関数を計算しなければならない．これについてはOsborn (1976) を参照されたい．カルマン・フィルターのアルゴリズムおよびAnsley (1979) によって提唱された関連する方法は，この問題については満足すべき解であると信じられている．第二の方法は，最初から適当な制約付きの非線型探索法を適用したり，または母数の変換を行なうことによって推定値に反転可能の制約を課すことである．この方法については，例えばHarvey (1981 a, 4.2節) を参照されたい．ところで変換の選択に当っては，いくつかの注意が必要であろう．なぜなら厳密最尤推定値は，しばしば反転可能領域の境界上に存在することがあるからである．この問題については**5.6**節で述べる．厳密最尤法以外の方法では，反転可能条件を課することの意味はいっそう強い．とくに条件付二乗和関数は，反転可能領域の外のMA母数についてはあまり意味あるものとはならないことがわかっている．

## 5.4 漸近的特性

ARMAモデルにおける最尤推定量の漸近特性は，通常対応する条件付二乗和推定量(最小二乗推定量)を調べることによって得られる．モデルが定常的でかつ反転可能である場合には，大標本における厳密な尤度関数と条件付尤度関数の差は無視し得るものとなる．したがって二つの推定量の漸近特性は同じものとなる．純粋なARモデルについては，条件付二乗和を最小にすることによって直ちに回帰推定量が導かれる．したがってマン=ワルド(Mann=Wald)の定理が適用可能である．このことについては，例えばHarvey (1981 a, 1.4節) を参照されたい．本書ではMAモデルや混合モデルについては，最尤推定量が漸近的に正規分布に従い，その共分散行列が情報量行列の逆行列に等しいという主張の厳密な理由づけは行なわない．この問題については，Fuller (1976), Hannan (1970) またはCrowder (1976) を参照されたい．

### 自己回帰過程

AR(1)モデルについては，情報量行列は$\log L$の二階微分の期待値をとることにより直接評価される．このことによって以下の結果が得られる．

$$\text{(4.1)}\qquad \mathrm{Avar}(\tilde{\phi},\tilde{\sigma}^2)=T^{-1}IA^{-1}(\phi,\sigma^2)=\begin{bmatrix}(1-\phi^2)/T & 0\\ 0 & 2\sigma^4/T\end{bmatrix}.$$

この結果は高次のモデルにも拡張される．そしてすべての場合において，$\sigma^2$の推定量は大標本では分散$2\sigma^4/T$を持ち，$\tilde{\phi}$と漸近的に独立に分布する．また$\tilde{\phi}$の漸近分散は以下のように与えられる．

$$\text{(4.2)}\qquad \mathrm{Avar}(\tilde{\phi})=T^{-1}V_p^{-1}.$$

ここで$\sigma^2V_p$は$(y_1,\cdots,y_p)'$の共分散行列である．

### MAおよび混合過程

純粋なARモデルの場合と同様に，$\sigma^2$の最尤推定量は大標本において分散$2\sigma^4/T$を持ち，他の推定量と漸近的に独立に分布する．ARMA母数$\psi=(\phi',\theta')'$の漸近共分散行列の表現は以下の行列

$$\text{(4.3)}\qquad IA(\psi)=\sigma^{-2}\,\mathrm{plim}\,T^{-1}\sum z_t z_t'$$

の逆行列を評価し，それを$T$で割ることによって得られる．ここで$z_t=-\partial\epsilon_t/\partial\psi$である．

MA(1)モデルに関しては$\partial\epsilon_t/\partial\theta$は(3.4)の漸化式に従う．したがって$z_t$は以下のAR(1)過程に従う．

$$\text{(4.4)}\qquad z_t=(-\theta)z_{t-1}+\epsilon_{t-1},\qquad t=1,\cdots,T.$$

そこで

$$\text{(4.5)}\qquad \mathrm{plim}\,T^{-1}\sum_{t=1}^{T}z_t^2=\mathrm{Var}(z_t)=\sigma^2/(1-\theta^2),$$

であり，$\tilde{\theta}$は平均$\theta$で以下のような分散を持って漸近的に正規分布する．

$$\text{(4.6)}\qquad \mathrm{Avar}(\tilde{\theta})=(1-\theta^2)/T.$$

混合モデルも同様に扱うことができる．ARMA(1,1)の場合は，(3.8 b)より$z_{2t}=-\partial\epsilon_t/\partial\theta$は(4.4)と同じようなAR(1)過程によって作られる．他方$z_{1t}=-\partial\epsilon_t/\partial\phi$と置くことによって，

$$z_{1t}=-\theta z_{1,t-1}+y_{t-1}.$$

これは以下のように書き改められる．

$$z_{1t}=\frac{y_{t-1}}{1+\theta L}=\frac{1}{1+\theta L}\frac{1+\theta L}{1-\phi L}\epsilon_{t-1}=\frac{\epsilon_{t-1}}{1-\phi L}.$$

すなわち$z_{1t}$もやはりAR(1)過程に従う．

$$z_{1t} = \phi z_{1,t-1} + \epsilon_{t-1}.$$

ここで $z_{1t}$ は $z_{2t}$ と同じ攪乱項を持っている．

表現(4.3)は(4.5)を考慮することにより評価することができる．そこで以下のように書くことができる．

$$\begin{aligned} E(z_{1t}z_{2t}) &= E\left[\left(\frac{\epsilon_{t-1}}{1-\phi L}\right)\left(\frac{\epsilon_{t-1}}{1+\theta L}\right)\right] \\ &= E\{[\epsilon_{t-1}+\phi\epsilon_{t-2}+\phi^2\epsilon_{t-3}+\cdots] \\ &\quad \times[\epsilon_{t-1}+(-\theta)\epsilon_{t-2}+(-\theta)^2\epsilon_{t-3}+\cdots]\} \\ &= \sigma^2[1+(-\phi\theta)+(-\phi\theta)^2+\cdots] = \sigma^2/(1+\phi\theta). \end{aligned}$$

したがって $\phi$ および $\theta$ の最尤推定量の漸近的共分散行列は以下のように与えられる．

(4.7)

$$\mathrm{Avar}(\tilde{\phi},\tilde{\theta}) = \frac{1}{T}\frac{1+\phi\theta}{(\phi+\theta)^2}\begin{pmatrix} (1-\phi^2)(1+\phi\theta) & -(1-\phi^2)(1-\theta^2) \\ -(1-\phi^2)(1-\theta^2) & (1-\theta^2)(1+\phi\theta) \end{pmatrix}.$$

## 5.5 仮説検定と信頼区間

ARMA 母数に関する仮説検定の手続きは **1.3** 節で述べられた原則から作ることができる．標準的な検定の枠組では，帰無仮説がより一般的なモデルの中に入れ子(nested)になっているということを想定している．すなわちこの枠組では制約のないモデルは，母数が $\psi=(\phi', \theta')'$ であるような ARMA$(p, q)$ 過程である．他方，制約されたモデルにおいては，帰無仮説 $H_0: \psi = \psi_0$ が課せられる．したがって，AR(1)モデルを MA(1)モデルに対して検定を行なうことはできない．他方，AR(1)モデルを ARMA(1, 1)に対して行なう検定はきわめて正当なものである．ところで非入れ子型(non-nested)の確率過程の間で区別を行なうことは，モデル選択の重要な役割である．赤池の情報量基準に基づいてそのような問題を扱う方法については，**6.3** 節で説明される．

次に，ARMA モデルにおいて検定することのできる仮説に関して，やや自明でない制限が課されなければならないことを述べることにしよう．ここで自己回帰および移動平均項が同時にあるモデルに加えられたとしてみよう．すると例えば制約されないモデルは ARMA(2, 2)過程であり，それが帰無仮説の下では ARMA(1, 1)過程となる．もし帰無仮説が正しければそれは ARMA(2, 2)



モデルでは $\phi_2 = \theta_2 = 0$ となり，それは識別可能ではない．このことは，モデルに対して共通因子の制約を課す仮定は標準的な手続きでは検定することができない，ということを意味している．

### ワルド検定および尤度比検定

ワルド(Wald)検定の統計量は 1 章(3.4)によって定義された．もしモデルがガウス=ニュートン法によって推定されたならば，$I^{-1}(\tilde{\psi})$ のごく自然な大標本推定量は以下のように与えられる．

(5.1)
$$\mathrm{Avar}(\tilde{\psi}) = \tilde{\sigma}^2(\Sigma z_t z_t')^{-1}.$$

ここで $z_t = -\partial\epsilon_t/\partial\psi$ は $\psi=\tilde{\psi}$ で評価されたものである．ガウス=ニュートン法の回帰的手法によって，ワルド検定統計量より $\epsilon_t(\tilde{\psi})$ を $z_t(\tilde{\psi})$ の上に回帰した場合における伝統的な $F$ 統計量へ修正することが示唆されるかもしれない．ここで $\sigma^2$ の推定量では $T-p-q$ を用い，ちょうど古典的な線型回帰における推定量 $s^2$ に対応している．一方 $W$ を帰無仮説 $H_0$ の下での制約の数 $m$ で割ることにより，$\chi^2$ 統計量を $F$ 統計量に転換することができる．$F$ 分布の使用については，この検定が厳密なものであるということを意味すると考えるべきではない．古典的回帰との類似から，$F$検定の方がある場合にはより正確であろうと連想されるけれども，$F$ 検定統計量は主にその便利さから利用されるのである．モデルにおける単一の母数についての検定は，古典的 $t$ 統計量に基づいて行なうことができる．この検定もやはり厳密なものではない．「漸近的 $t$ 比」(asymptotic $t$-ratio)および「漸近的 $F$ 比」(asymptotic $F$-ratio)という名称はこのような意味においてしばしば用いられている．

**例 1**　AR(1)モデルにおいて $\phi$ の条件付二乗和推定量(最小二乗推定量)は，$y_t$ を $y_{t-1}$ 上に回帰することによりただちに得られる．漸近的 $t$ 比は以下のように与えられる．

(5.2)
$$\tilde{\phi}/(s/\sqrt{\Sigma y_{t-1}^2}).$$

ここで $s^2$ は古典的回帰における $\sigma^2$ の不偏推定量に対応している．上記の(5.2)を用いた検定は，標準誤差の推定値として $(1-\tilde{\phi})^2/T$ の平方根を用いた検定と漸近的に同等である．それは以下のような関係によるものである．

(5.3)
$$\mathrm{plim}\,(s^{-2}\Sigma y_{t-1}^2/T) = 1/(1-\phi^2).$$

**例 2**　ガウス=ニュートン法によって MA(1)モデルを推定するとき，$\theta$ に関しての検定は推定値が収束したときに行なうことができる．これは $\partial\epsilon_t/\partial\theta$ の係数の $t$ 統計

量に基づいて行なわれる．

さて尤度比検定統計量は以下のように定義される．

(5.4) $$\mathrm{LR} = T\log(\tilde{\sigma}_0^2/\tilde{\sigma}^2) = T\log(SSE_0/SSE).$$

ここで $SSE_0$ と $SSE$ はそれぞれ制約されたモデルおよび制約されないモデルの残差二乗和 $\sum\epsilon_t^2$ である．古典的線型回帰との類似性から，次のような検定の修正が提唱されよう．すなわち統計量

(5.5) $$\mathrm{LR}^* = \frac{(SSE_0 - SSE)/m}{SSE/(T-p-q)}$$

を，帰無仮説 $H_0$ の下で，自由度 $(m, T-p-q)$ の $F$ 分布に従うと解釈して検定を行なうことが考えられる．

### ラグランジュ乗数検定

制約されたモデルが推定されたと想定しよう．帰無仮説 $H_0$ のラグランジュ乗数検定は，ガウス=ニュートン法を1回行なうことにより帰着される．これは1章(3.6)における $TR^2$ 統計量による方法であって，この統計量が帰無仮説の下で $\chi_m^2$ 変量となることによって検定することができる．

**例3** AR(1)モデルが当てはめられ，そしてわれわれは追加されたMA部分の検定を行なうという状況を想定しよう．いい換えれば，制約のないモデルはARMA(1, 1)で，他方帰無仮説は $H_0: \psi = \psi_0$ であり，ここで $\psi_0 = (\phi, 0)'$ である．ARMA(1, 1)過程の残差関係は以下のように与えられる．

(5.6) $$\epsilon_t = y_t - \phi y_{t-1} - \theta\epsilon_{t-1}.$$

さらに $\epsilon_t$ の $\phi$ および $\theta$ に関する微分は，(3.8 a)および(3.8 b)の漸化表現によって与えられている．ラグランジュ乗数法は $\epsilon_t(\tilde{\psi}_0)$ のそれらの微係数の $\tilde{\psi}$ で評価されたものへの回帰に基づいている．帰無仮説 $H_0$ の下で $\theta = 0$ であるので，この方法は $\epsilon_t$ の $y_{t-1}$ および $\epsilon_{t-1}$ 上への回帰と簡単化される．検定は $TR^2$ が $\chi_1^2$ 変数であると扱うことによって行なうことができよう．この場合修正ラグランジュ乗数統計量は，$\epsilon_{t-1}$ の係数に関する $t$ 検定の形をとる．

**例4** もし上記の例における帰無仮説が $H_0: \phi = \theta = 0$ であるならば，検定を行なうことはできない．なぜなら $\phi = \theta = 0$ のときには $\partial\epsilon_t/\partial\theta = -\epsilon_{t-1} = -y_{t-1}$ であり，かつ $\partial\epsilon_t/\partial\phi = -y_{t-1}$ であるから，$\epsilon_t$ の偏微係数に対する回帰における二つの説明変数は完全な共線性を持っている．ラグランジュ乗数統計量がこの場合に計算できないのは実は，本節の最初において示された共通因子についての制限に関する一般的な注意点の反映にすぎないのである．



### 信頼区間

単一の母数についての信頼区間を計算する最も直接的な方法は，標準的な回帰手法に基づくものである．母数の漸近的な標準誤差の推定値はガウス=ニュートン法の最適化アルゴリズムの一部として求められる．したがって近似的な信頼区間は $t$ 分布に基づいて作られるであろう．

もう一つのアプローチは，二乗和関数に基づいている．単一の母数については，これは先に述べた方法と同じ解を与えるとは限らないが，差があるとしてもそれは非常に小さいものであろう．二乗和によるアプローチの有利な点は，いくつかの母数についての信頼領域を作ることに用いることができるからである．すなわち $M$ 個の母数が考慮されているとし，$\chi_m^2(\alpha)$ が有意水準 $\alpha$ の検定のための $\chi^2$ 分布の臨界値であると想定しよう．すると漸近的な $1-\alpha$ の信頼域は，以下のような二乗和関数の等高線によって境界づけられる．

(5.7) $$SSE(\psi) = SSE(\tilde{\psi})[1 + \chi_m^2(\alpha)/T - m].$$

この問題についてより詳しくは Box=Jenkins (1976, pp. 228-231) を参照されたい．

## 5.6 小標本特性*

純粋の AR の場合には小標本特性についていくつかの研究がなされているが，ARMA モデルの推定量の分布に関してはほとんど解析的な結果は得られていない．したがって推定量の小標本特性について何らかの証拠を提供するためには，モンテ・カルロ実験に依存する必要がある．本節では種々の文献に報告されている結果のいくつかをまとめることを試みる．ここでの展望は Ansley=Newbold (1980), Davidson (1981), Dent=Min (1978), Harvey=Phillips (1977), Kang (1975), Nelson (1974 ; 1976) および Nelson=O'Shea (1979) 等の研究に基づいている．

### 自己回帰過程

AR(1)過程に関しては，$T$ が非常に小さい，たとえば20以下でない限り，厳密最尤法と最小二乗推定量の差はほとんど無視し得るものであるという左証

がある．いずれの場合にも，推定量には下方の偏りがあり，それはだいたい $2\phi/T$ である．詳しくは Shenton and Johnson (1965) を参照されたい[訳注]．

AR (1) の場合においては種々の漸近的に等価な推定量の間で選択を行なうことにはほとんど意味はないが，そのことは高次のモデルにおいては当てはまらないようである．$T$ が 100 で，$p = 2$ および 3 の場合において，Dent=Min (1978) は，ある場合において最尤推定量の標本平均二乗誤差は，他の方法によって得られた平均二乗誤差よりはるかに小さいということを報告している．容易に類推されるように，種々の方法の間の差は真の母数が定常性領域の境界に近いほどより顕著になる．

信頼区間および $t$ 統計量は，AR モデルが回帰により推定されたときには非常に簡単に作ることができる．例えば本章例 1 を参照されたい．真の母数が定常性の境界の近くにない限り，これらの統計量は十分に信頼し得るもののようである．このことについては，さらに Nelson (1976) および Ansley=Newbold (1980) を参照されたい．

### 移動平均過程と混合過程

厳密最尤法についての反転可能性の意味は **5.4** 節において詳しく議論した．そこでは制約を課すことなしに最適化が行なわれ得ることを取り上げた．MA (1) の場合には，単位円周の外側の $\theta$ の最尤推定値は，単にその逆数をとることによって反転可能な推定値へ簡単に変換することができる．他方，条件付二乗和関数は単位円周外の $\theta$ の値に関して意味を持たない．ところで小標本においては，反転可能領域の中で条件付二乗和関数が最小値を持たないことが時々生じ，制約を課さない最適化方法は収束しないか，またはもし収束したとしても意味のない推定値をもたらすことがある．このようなことは $|\theta|$ が 1 に近くかつ $T$ が小さいときにのみ起るということは強調されなければならないが，さらに制約付非線型探索法を適用することの重要性をも指摘しているといえよう．このことから「ボックスの計算法」(Box complex algorithm) が Dent=Min (1978) により推奨されている．

厳密な尤度関数を制約を課すことなしに最大化することにより，比較的興味ある結果が得られている．すなわち小標本でかつ $\theta$ の真の値が単位円に近いときに，$\theta$ が厳密に $+1$ または $-1$ と等しい値において全域的最大値が比較的多く見い出されるのである．この効果は図 5.1 に示されているが，それは Harvey=Phillips (1977) の表 2 から作られたものである．$T = 25$，$\theta = 0.9$ の場合について 200 回の繰り返し中，ほぼ半数の推定値はほぼ $\theta = 0.85$ を中心とした典型的な釣鐘型を形成している．残りの 103 の推定値はちょうど 1 となっている．さらにこのことについては例えば，Ansley=Newbold (1980) における図 1 を参照されたい．推定量がある特定の値をゼロでない確率でとり得るという考えは，やや受け入れにくい．しかしながら最近，Sargan and Bhargava (1980) がこの問題についての解析的な証明を与えた．彼らは，$\theta$ および $T$ の与えられた値について $\hat{\theta} = \pm 1$ となる確率を評価する方法を提示したのである．

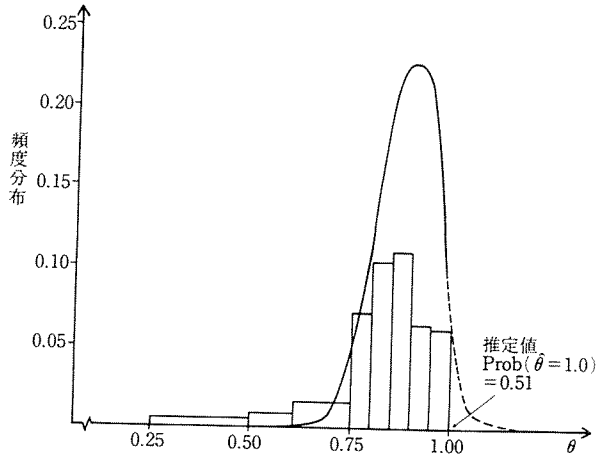



**図 5.1** MA (1) 過程における $\theta$ の最尤推定量の理論的および実験的頻度分布：$\theta = 0.9$，$T = 25$ の場合

先に述べたモンテ・カルロ実験の同じ繰り返しの組において，Harvey=Phillips は条件付二乗和推定値 (最小二乗推定値) も計算した．32 の場合，$\theta = 1$ において最小値が得られた．しかしこれは単位円周外の推定値はすべて認められないものとしたためである．すなわち最尤法の場合と異なり，これが最適化問題についての全域的な解を意味しているわけではない．

最尤推定量の分布についてこのような結果が得られたのであるから，標本の平均二乗誤差に基づく推定量の間の比較においては注意しなければならない．なぜなら標本平均二乗誤差は，最尤推定量の分布における $\hat{\theta}$ の以上のような

訳注　Yamamoto T. and N. Kunitomo, "Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models", *Annals of the Institute of Statistical Mathematics*, 1984 も参照.

やや特異な二つの山を捕えてはいないからである．しかしながら以下で言及される結果の多くは $T = 100$ の場合に基づいており，そのような大きさの標本数に関しては前述のような影響は，$\theta$ が単位円に近くない限りそれほど重要ではない．ただし階差の過剰によって $\theta$ の真の値が $-1$ になり得るということは，異なった推定量を評価するときに心に留めておくべきことであろう．

Nelson (1974)，Davidson (1981) そして Dent=Min (1978) は，$\theta$ の種々の値についていくつかの推定法を比較している．残念ながら Nelson の論文は無条件二乗和を尤度のもっともらしい近似として採用し，彼の結果には最尤推定量は含まれていない．無条件二乗和推定量は，厳密な尤度から行列式の項 $\log|V|$ を除いた関数を最大化することによって得られる．$T$ が十分大きければ，この推定量は厳密最尤推定量と同様なものである．しかしもし $T$ が小さく $\theta$ が単位円周上に近いときは，両者は非常に異なるであろう．Kang (1975) や Davidson (1981) によって示された証拠によると，この方法は推薦する理由はないこととなる．以上三つの研究の主な結論は以下の通りである．第一に，$T = 30$ の場合，$\theta = \pm 0.9$ の時には条件付二乗和推定量は明らかにゼロへの偏りを持つ．しかしこの偏りの程度は $T$ が 100 のときは大幅に減少する．第二に，$\theta$ の真の値が単位円周上に近いときには，平均二乗誤差基準の下では最尤推定量が条件付二乗和推定量よりも好ましい．ただし，もし $T = 100$ の場合において，$|\theta|$ が 0.8 より小さいときには，二つの推定量の間にはほとんど差がないということを Dent=Min の結果が示している．実際には，どちらかと言えば条件付二乗和推定量の方がより小さい平均二乗誤差を持つ．これを裏づける証拠が Davidson によっても示されている．

Nelson は推論の他の側面について，すなわち標準誤差の推定について調べている．この推定値はガウス=ニュートン法によって生じる交差積行列から計算できるし，ヘッセ行列を数値的に評価することによっても得ることができる．条件付二乗和推定量については，推定された標準誤差はすべての繰り返しから与えられる対応した平均二乗誤差よりもかなり小さくなる傾向があることを見い出している．このことは帰無仮説が必要以上に棄却される可能性を意味する．さらに $T = 30$ の場合に，$\theta$ の多くの場合について，名目上の検定の有意水準 0.05 が実際の検定では 0.10 の大きさをもたらすことを Nelson の結果が示している．標本数が増大すればいくらかの改善は見られる．しかしながら容易に想像されるように，この点についても $\theta$ の値が単位円周上に近い場合は問題を



生じる．例えば $T = 100$ のとき，$\theta = 0.9$ の場合には，0.05 の名目上の検定の有意水準は実際にはだいたい 0.12 の大きさをもたらす．Nelson はこれらの結果から MA モデルにおける $t$ 比 ($t$-ratios) は小標本においては当てにならないと結論している．これを裏づける証拠は Nelson=O'Shea (1979) および Ansley=Newbold (1980) によってさらに示されている．

最尤法と条件付二乗和以外の推定量の性質も，いくつかの論文において報告されている．ここで特に興味をそそるのは，(3.11) で定義されたコレログラム推定量 $\hat{\theta}_c$ である．もし $\theta$ がゼロに近いときは，$\hat{\theta}_c$ は条件付二乗和推定量に近い平均二乗誤差を持つ．しかしその相対的有効性は，$|\theta|$ が 1 に近づくに従って比較的早く下落する．$T$ が増大するに従ってその平均二乗誤差は減少するけれども，それを制限付二乗和推定量と比べると，相対効率は実際にはより悪くなる．Nelson の研究においては，$\theta = 0.5$ で $T = 30$ のときには，$\hat{\theta}_c$ の標本平均二乗誤差は対応する制限付二乗和推定量の平均二乗誤差の約 2 倍となっている．そして $T$ が 100 の場合には，それは約 3 倍となる．さらに $\theta = 0.5$ の場合，その $\hat{\theta}_c$ の漸近的有効性は 0.49 である．Dent=Min によって与えられた $T = 100$ の場合の結果も非常に似たものである．全体的な結論としては，$\hat{\theta}_c$ は最終的な推定量としては受け入れ難いが，繰り返し手法における出発値としては十分な推定量といえる．Nelson (1974, p. 127) によれば，「もし $\hat{\theta}_c$ が最初の推定値として用いられるときには，ガウス=ニュートン法は，制限付二乗和関数の全域的最小値に確実にかつ素早く導いてくれる」とのことである．

MA (2) 過程に関する Dent=Min の結果は，母数が反転可能領域の境界上に近くない限りは，最尤推定量と制限付二乗和推定量の差にほとんど区別がつかないということを再び示している．境界上に近い場合には，最尤推定量を用いることに利益があろう．Dent=Min (1978, p. 38) が到達した「純粋の移動平均モデルについては，条件付最小二乗推定量が最尤推定量よりもわずかながら優れているかもしれない」という結論は，彼ら自身の MA (2) 過程についての結果からは必ずしも十分に支持されたとはいい難い．条件付最小二乗推定量に関して言及されている平均二乗誤差は，研究されたモデルの半数以下の場合においてのみ対応する最尤推定量の平均二乗誤差より小さく，かつその差は非常に小さい．したがって 100 回の繰り返し実験のみを基礎に，確固たる結論を導き出すのは難かしいといえよう．

ARMA (1, 1) モデルについては，Dent=Min の結果は，最尤推定量と条件付き

最小二乗推定量は同じような良好な性質を持つことを示している．一方が他方よりも優れているという明確な傾向は生じていない．他方，Ansley=Newbold (1980)の結果によると，$\phi$および$\theta$の両方にとって厳密最尤推定量を用いることによって明らかに利得が得られている．MA(1)モデルの場合のように平均二乗誤差を比べる際にはどのような場合にも，単位円周上に真の値が近い場合の$\theta$の推定値の「積み重ね」効果は心に留めておかなければならない．たとえば$\theta = 0.9$かつ$\phi = 0.9$の場合にHarvey=Phillips (1977)は，$T = 25$のときに40%から50%の$\theta$の最尤推定値がちょうど1になることを報告している．同様の比率が$\phi$を0.5としたときにも観察された．

Ansley=Newbold (1980)はまた混合モデルの信頼区間の信頼性について研究を行なっている．彼らは，そこに何かしらでも余分の母数があるような場合には，重大な問題があることを報告している．すなわちそのような場合には，計算される信頼区間は極めて狭すぎるようになることがある．

## 5.7 多変量モデル

一般的な平均ゼロのベクトル値ARMA$(p, q)$過程は以下のように表現される．

$$y_t = \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \epsilon_t + \Theta_1 \epsilon_{t-1} + \cdots + \Theta_q \epsilon_{t-q}, \quad t = 1, \cdots, T. \tag{7.1}$$

ここで$y_t$は観測値からなる$N \times 1$のベクトルで，$\Phi_1, \cdots, \Phi_p$，$\Theta_1, \cdots, \Theta_q$はARMA母数の$N \times N$行列である．攪乱項ベクトル$\epsilon_t$は正規分布していると仮定する．すなわち$\epsilon_t \sim NID(0, \Omega)$とする．ここで$\Omega$は$N \times N$の正値定符号行列である．

この形式のモデルを推定するために用いることが可能な方法は，本質的に一変量の場合に開発された手法の一般化である．原理的には厳密最尤推定量は以前と同様に求めることができる．しかしながら，複数の時系列を結びつけるために導入される母数のために，計算上の問題がより複雑になり，この問題のために厳密最尤法が実際的な方法として考慮の対象外となることが多い．このために比較的高い関心が条件付二乗和法(最小二乗法)の一般化にはらわれ，多変量のガウス=ニュートン法のアルゴリズムが一般の場合に用いられる．一変量の場合と同様に，微係数の解析的表現が比較的容易に計算できるという事実によって，このアプローチの有効性は極めて高くなっている．



以下での議論はベクトル値AR(1)モデルおよびベクトル値MA(1)モデルにしぼって行なわれる．二つのモデルのうち，前者はそれが推定の視点から最も簡単な多変量確率過程であるという意味において重要なモデルである．いったんこのモデルを扱うことができるのであれば，高次の自己回帰モデルへの拡張は比較的容易である．他方，ベクトル値MA(1)過程は一般的な多変量ARMA$(p, q)$過程を推定するに当って含まれる困難な問題をほとんどを例示している．

### ベクトルAR(1)過程の推定

比較的簡単であることのために一次のベクトル値自己回帰過程

$$y_t = \Phi y_{t-1} + \epsilon_t, \quad t = 1, \cdots, T \tag{7.2}$$

は，とくに計量経済学の文献においてしばしば注目されている．ところでこのモデルは多数の未知母数を含んでいる．自己回帰行列$\Phi$は$N^2$の母数を含み，他方，対称な共分散行列$\Omega$は，さらに$N(N+1)/2$の異なった母数を持っている．

モデル(7.2)の尤度関数は，一変量の場合と全く同様に導くことができる．ベクトル$y_{t-1}$が与えられたときには，$y_t$の分布は平均$\Phi y_{t-1}$および共分散行列$\Omega$の多変量正規分布となる．そこで

$$\log L(y_T, y_{T-1}, \cdots, y_2 / y_1) = -\frac{N(T-1)}{2} \log 2\pi - \frac{1}{2}(T-1) \log |\Omega| - \frac{1}{2} \sum_{t=2}^{T} (y_t - \Phi y_{t-1})' \Omega^{-1} (y_t - \Phi y_{t-1}) \tag{7.3}$$

とすると，これは$t = 1$の観測値$y_1$に条件付けられた対数尤度関数である．厳密な対数尤度関数は，$y_1$の無条件分布を考慮に入れることによって得ることができる．$y_1$は平均ゼロ，そして以下のように定義される共分散行列$\Sigma$を持つ多変量正規分布である．

$$\Sigma = \Phi \Sigma \Phi' + \Omega. \tag{7.4}$$

上式については例えば2章(7.11)を参照されたい．$y_1$の分布を(7.3)と結合することにより，以下のような厳密な対数尤度関数が得られる．

$$\log L(y_T, \cdots, y_1) = -\frac{NT}{2} \log 2\pi - \frac{1}{2} \log |\Sigma| - \frac{(T-1)}{2} \log |\Omega| - \frac{1}{2} \sum_{t=2}^{T} (y_t - \Phi y_{t-1})' \Omega^{-1} (y_t - \Phi y_{t-1}) - \frac{1}{2} y_1' \Sigma^{-1} y_1. \tag{7.5}$$

上記(7.5)の $\Omega$ および $\Phi$ に関する最大化は，一般的な数値的最適化の方法によって行なうことができる．しかしながら $y_1$ を固定されたものとして扱い，条件付対数尤度関数(7.3)を扱った方がはるかに容易であろう．その場合，$\Omega$ が与えられたときには，$\Phi$ の最尤推定量は以下を最小化することより得られる．

$$S = \sum_{t=2}^{T} (y_t - \Phi y_{t-1})' \Omega^{-1} (y_t - \Phi y_{t-1}). \tag{7.6}$$

ところでこれは，多変量最小二乗推定量によって最小化される基準関数に他ならない．((Harvey 1981 a, 2.10節)を参照のこと)したがって，$\Phi$ の最尤推定量は $\Omega$ に依存しない．$\tilde{\Phi}$ の $i$ 番目の行は，$y_t$ の $i$ 番目の要素を $y_{t-1}$ ベクトルの上へ最小二乗回帰を行なうことによって得られる．すなわち，

$$\tilde{\Phi}' = \left[ \sum_{t=2}^{T} y_{t-1} y_{t-1}' \right]^{-1} \sum_{t=2}^{T} y_{t-1} y_t' \tag{7.7}$$

となる．

共分散行列 $\Omega$ の最尤推定量は，残差ベクトル $e_t = y_t - \tilde{\Phi} y_{t-1}, t = 2, \cdots, T$ から直接的に計算される．すなわち

$$\tilde{\Omega} = T^{-1} \sum_{t=2}^{T} e_t e_t' \tag{7.8}$$

である．

### 最尤推定量の漸近的分布と関連する検定統計量

ここで $\phi = \mathrm{vec}(\Phi)$ としよう．すると $\phi$ の最尤推定量は漸近的に平均 $\phi$，そして以下のような共分散行列を持つ正規分布をする．

$$\mathrm{Avar}(\tilde{\phi}) = T^{-1}(\Sigma^{-1} \otimes \Omega). \tag{7.9}$$

ここで $\Sigma$ を推定する簡単な方法としては以下の推定量が考えられよう．

$$\tilde{\Sigma} = T^{-1} \sum_{t=2}^{T} y_t y_t'. \tag{7.10}$$

母数行列 $\Phi$ に関する仮説検定は，ワルド検定の原理を用いて非常に簡単に行なうことができる．すなわち帰無仮説 $H_0 : \Phi = 0$ を対立仮説 $H_1 : \Phi \neq 0$ に対して検定するための適切な統計量は，以下のように与えられる．

$$W = \tilde{\phi}' [\mathrm{Avar}(\tilde{\phi})]^{-1} \tilde{\phi}. \tag{7.11}$$

もし $H_0$ が正しければ，$W$ は大標本において自由度 $N^2$ の $\chi^2$ 分布に従う．



仮説検定へのもう一つのアプローチは，尤度比統計量を用いることである．すなわち $\Phi = 0$ の検定について以下のような統計量を考えればよい．

$$\mathrm{LR} = T \log(|\tilde{\Sigma}|/|\tilde{\Omega}|). \tag{7.12}$$

ワルド統計量および尤度比統計量は，$\Phi$ が対角であるかどうかの検定にも使うことができる．実際にも，まず $\Phi$ の対角化について検定することは有用である．そしてもしこの仮説が棄却されなければ，さらに $\Omega$ の対角要素がゼロであるかを検定することが考えられよう．仮説は入れ子型となっている．したがってもし $W_1$ が $\Phi$ の対角性を検定するワルド統計量であり，$W$ が(7.11)の検定統計量であるならば，$W_1$ および $W - W_1$ は漸近的に独立にそれぞれ自由度 $N(N-1)$ および $N$ の $\chi^2$ 分布として分布する．これについては，例えばHarvey(1981 a, 5.8節)を参照されたい．ここでもし $\Phi$ が対角であるとみなされれば，$\Phi$ の最尤推定値はもはや $\Omega$ から独立ではないということに注意を要する．すなわち一連のワルド検定は，対応する一連の尤度比検定よりも少ない計算量ですむことになる．

### ベクトル値 MA(1) 過程

ベクトル値 MA(1)過程は2章(7.4)で定義された．もし $\epsilon_t \sim NID(0, \Omega)$ でかつ $\epsilon_0 = 0$ であるならば，対数尤度関数は1章(2.28)の形をとる．$\Theta$ が与えられたという条件下で，$\epsilon_t$ は以下のような漸化式から得ることができる．

$$\epsilon_t = y_t - \Theta \epsilon_{t-1}, \qquad t = 1, \cdots, T. \tag{7.13}$$

他方 $\Omega$ は(7.8)と同様の形式から推定される．尤度関数は多変量ガウス=ニュートンのアルゴリズムにより，$\Theta$ に関して最大化される．これについてはHarvey(1981 a, 4.4節)およびWilson(1973)を参照されたい．いま $\hat{\theta}$ を $N^2 \times 1$ ベクトル $\theta = \mathrm{vec}(\Theta)'$ の初期推定量であるとしよう．$\Omega$ が与えられれば，これはガウス=ニュートン法の一般化によって以下のように更新される．

$$\theta^* = \hat{\theta} + \left[ \sum_{t=1}^{T} Z_t \Omega^{-1} Z_t' \right]^{-1} \sum_{t=1}^{T} Z_t \Omega^{-1} \epsilon_t. \tag{7.14}$$

ここで $Z_t$ は $Z_t = -\partial \epsilon_t' / \partial \theta$ で定義される $N^2 \times N$ の微係数行列である．更新される $\Omega$ の推定量を用いて，これを繰り返して収束するまで続けられる．

微係数はちょうど一変量の場合と同様に逐次的に計算することができる．もしモデルが以下の形で書かれるならば，

$$y_t = \epsilon_t + (I \otimes \epsilon_{t-1}') \theta. \tag{7.15}$$

残差の漸化式(7.13)は以下のようになる.

$$\epsilon_t = y_t - [I \otimes \epsilon'_{t-1}]\,\theta. \tag{7.16}$$

したがって，$Z_0 = 0$とすると以下の結果を得る.

$$Z'_t = [I \otimes \epsilon'_{t-1}] + [I \otimes Z'_{t-1}]\,\theta, \qquad t = 1, \cdots, T. \tag{7.17}$$

最尤推定量の漸近的共分散行列の推定量は，多変量ガウス=ニュートン法のアルゴリズムから自動的に与えられる．それは以下のように求められる.

$$\mathrm{Avar}(\tilde{\theta}) = \left[\sum_{t=1}^{T} Z_t \tilde{\Omega}^{-1} Z'_t\right]^{-1}. \tag{7.18}$$

ここで$Z_t$のすべての要素は$\theta = \tilde{\theta}$で評価される．そこでワルド検定を，ちょうどベクトル値AR(1)モデルの場合と同様に行なうことができる.

### 見かけ上無関係なベクトル値 ARMA 過程*

対角の係数行列としてモデルを定式化し，二段階の推定方法を用いることによって，MA 過程および混合過程は大きく簡素化される．ARMA 母数の一致推定値は，$N$個の過程をそれぞれ別々に一変量時系列として扱うことによって得られる．他方 $\Omega$ の推定値はその推定残差から作ることができる．多変量ガウス=ニュートン法を一回行なうと，ARMA 母数全体の漸近的に有効な推定量が得られる．こうして得られる推定量は，見かけ上無関係な多変量回帰方程式(seemingly unrelated regression equations)体系における一般化最小二乗推定量の型をしている.

Nelson (1976)が報告したモンテ・カルロ実験は，ARMA 過程におけるそれぞれの系列を一つの体系として扱うことによって向上が期待される小標本効率を調べている．以下のような$N = 2$の場合のベクトル値AR(1)過程について，

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 \\ 0 & \phi_2 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \tag{7.19}$$

母数の一つ，例えば$\phi_1$の一変量からの推定量との相対効率は，大標本に関しては解析的に求めることができる．もし$\hat{\phi}_1$および$\tilde{\phi}_1$をそれぞれ一変量時系列および多変量時系列の最尤推定量であるとし，$\rho$を$\epsilon_{1t}$と$\epsilon_{2t}$の相関とすると，$\hat{\phi}_1$の漸近効率は以下のように与えられる.

$$\mathrm{Eff}(\hat{\phi}) = \frac{\mathrm{Avar}(\tilde{\phi})}{\mathrm{Avar}(\hat{\phi})} = (1-\rho^2)\left[1 - \frac{\rho^4(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)^2}\right]^{-1}. \tag{7.20}$$

標本数が30程度あれば，(7.20)は実際に達成される効率性の改善についての良い指標を与えている．期待された通りに時系列間の同時相関が高ければ高いほど，母数を同時に推定することから得られる効率についての利益は大きくなる．さらに(7.20)を吟味することにより，二つの母数の値に大きな差があるほど，効率性の利益が大きくなることが見てとれる．すなわち$|\phi_1 - \phi_2|$がその最大値である2に近づくに従って，相対効率$\mathrm{Eff}(\hat{\phi})$はその最小値である$1-\rho^2$へ近づく.



同様の分析が二つのMA(1)過程についても行なうことができる．再びNelsonの結果によれば，得ることのできる効率性の改善は漸近理論によって予想されるものとだいたい等しいことを示している．もっともMA母数の小標本の分布は，漸近理論が示唆するよりもはるか分散が大きくなっている.

### 練習問題

**1.** 帰無仮説がARMA(1, 1)なるモデルであり，対立仮説がARMA(1, 3)であるような仮説検定を行なうラグランジュ乗数検定を導け．またどのようにして，$F$分布に基づく修正されたラグランジュ乗数検定を行なうことができるであろうか.

**2.** ARMA(1, 1)モデルにおける$\theta = 0$のラグランジュ乗数検定は，AR(1)モデルの残差から計算された一階の自己相関$r(1)$に基づく検定と同値であることを示せ．ここで$r(1)$は平均ゼロおよび分散$\phi^2/T$に従う漸近的正規変量であるとみなすことにする.

**3.** 一組の観測値3.6, 4.8, 5.0, 6.6, 4.7より$\theta_1 = -0.5$および$\theta_2 = -0.2$なるMA(2)過程における条件付二乗和を計算せよ.

**4.** もしある時系列の標本自己相関が$r(1) = 0.8$および$r(2) = 0.5$であるとき，AR(2)モデルにおける母数の漸近的に有効な推定値を見つけよ．さらに観測値の分散が2.0であるとき，撹乱項$\epsilon_t$の分散を計算せよ．母数$\phi_1$，$\phi_2$の推定値の漸近的標準誤差を計算せよ.

**5.** 問4におけるAR(2)過程のパワー・スペクトルを描け．そしてその最大値を探しコメントを加えよ.

**6.** AR(2)モデルにおけるユール=ウォーカー方程式は，$y_t$を$y_{t-1}$と$y_{t-2}$に回帰したときの$\phi_1$および$\phi_2$とほぼ同じ推定量を与えることを示せ．これらの推定量の漸近的共分散行列の$\phi_1$および$\phi_2$に関する表現を求めよ.

**7.** 標本数100から$r(1) = 0.60, r(2) = 0.32, r(3) = 0.18$および$\bar{y} = 7.2$という情報が与えられたとして，平均がゼロでないAR(1)モデルにおける漸近的に有効な推定値を計算せよ．もし$c(0) = 4.0$であるとき，情報量行列を導くことによりこれらの推定値の漸近的標準誤差を計算せよ.

**8.** 以下のような検定がベクトル AR(1)モデル，すなわち(7.2)における $\Phi = \Phi_0$ なる仮説を検定するために提唱されてきた．

$$\mathrm{tr}\,(\hat{\Phi}-\Phi_0)\Big[\sum_{t=2}^{T} y_t y_t'\Big](\hat{\Phi}-\Phi_0)\hat{\Omega}^{-1}.$$

これは自由度 $N^2$ の $\chi^2$ 変量として検定に用いられる．この検定は本文の中で議論されたものと関係があるであろうか．

**9.** モデル $y_t = \phi y_{t-1} + \theta_0 + \epsilon_t + \theta_1 \epsilon_{t-1}$ における $\theta_0 = 0$ という仮説についての尤度比検定，ワルド検定およびラグランジュ乗数検定をどのようにすれば構成できるだろうか．



# 6章
# 時系列モデルの作成手順と予測

## 6.1 はじめに

一変量時系列データを扱う基礎として，前章までに述べた自己回帰移動平均(ARMA)モデルは有用である．ARMA 過程は主として定常時系列を扱うモデルとして考えられたが，広範囲な非定常的変動を説明するモデルのクラスへと容易に拡張することができる．通常それは階差(差分)をとることによって行なわれ，ARMA モデルはもとの観測値そのものではなく一回あるいは二回の階差に対してフィットされる．その結果として生じるモデルのことを自己回帰和分(integrated)移動平均(ARIMA)過程という．

ARIMA モデルをデータにあてはめる方法は主としてボックス=ジェンキンズによって発展させられた．そこでしばしばそのアプローチ全体はボックス=ジェンキンズ法(Box=Jenkins method)と呼ばれている．その中心部分については **6.3** 節と **6.7** 節の最後の部分で解説する．基本的にはモデル選択の手順は三つの部分から構成されている．第一段階では仮のモデルが選択される．このことは適当な階差の次数とともに次数 $p, q$ についてある選択がなされることを意味している．ボックス=ジェンキンズはこれをモデルの識別(model identification)と呼んでいる．第二段階は推定であり，この問題についてはすでに前章で述べた．最後の段階ではモデルの診断(diagnostic checking)が行なわれる．モデルの識別・推定・診断は満足のゆくモデルが得られるまで行なわれる．以下ではこの手順を説明するがその際，Box=Jenkins (1976) の方法についていくつかの修正と改善を示唆するので，もとの方法と必ずしも同一ではない．

ARIMA モデルを作成する理由づけについては **6.5** 節で述べる．階差をとる考え方は確率的トレンドの概念とともに局所的トレンドと大域的トレンドの関係とも結びついているのでその点について詳しく解説する．**6.6** 節では非確率的季節要素のモデル分析法とともに徐々に変化する季節変動要素が階差操作によってうまく扱われることを示す．**6.7** 節では一つの一般的モデルのクラスが導入される．

ARMA モデルを用いた予測方法は **6.4** 節で説明される．さらに最小平均二乗誤差予測量が考察されその方法が **6.5** 節と **6.6** 節で導入される季節 ARIMA モデルに直ちに拡張される．多変量 ARIMA モデルの作成手順は最後の節で解説される．

## 6.2 独立性検定

有用なモデルはデータのシステマティクな動きをとらえることを目ざしている．したがって，もしこの目的が達成されればそこから生じる残差は本質的に独立な確率変数となるべきであろう．いい換えると，残差にはもはや何のシステマティクで予測可能な要素は含まれないことになる．本節では観測される時系列の独立性を評価する基本的問題を扱い，次節でこの考え方をあてはめられた ARMA モデルからの残差の検定の問題へと拡張する．

さて $y_1, y_2, \cdots, y_T$ を平均 $\mu$，分散 $\sigma^2$ のホワイト・ノイズ過程の実現値としよう．この過程では理論的な自己相関はゼロである．ところが標本自己相関は標本誤差のために必ずしもゼロとはならない．そこでもし独立性の評価が標本相関関数に基づいてなされる場合には，その標本分布を調べる必要が生じる．幸いにも大標本では非常に単純な結果，すなわち $\tau \neq 0$ に対し $r(\tau)$ は平均ゼロ，分散 $1/T$ の互いに独立な正規分布に従うことが知られている．このことは $r(\tau)$ は近似的に $y_t$ を $y_{t-\tau}$ に回帰したときの $y_{t-\tau}$ の係数に等しくなるので，マン=ワルド (Man=Wald) の定理 (Harvey (1981 a) pp. 48-49) より直ちに導かれる．時系列がホワイト・ノイズ過程の場合には $\sqrt{T}$ で基準化した極限正規分布は平均ゼロで分散

$$\sigma^2 (\text{plim}\, T^{-1} \sum y_{t-\tau}^2)^{-1} = 1$$

となる．

この結果を用いると仮説 $\rho(\tau) = 0$ は任意の $\tau$ に対して $T^{1/2} r(\tau)$ を標準正規



変数とみなして検定を行なうことができる．5%有意水準では $|T^{1/2} r(\tau)| > 1.96 \simeq 2.00$ ならば帰無仮説は棄却される．ただしこのような検定は前もって検定に用いるラグ $\tau$ をあらかじめ特定化したときにのみ妥当する．すなわち，このことは時系列の性質についての事前 (a priori) 知識が必要であることを意味している．例えば，四半期データでは $r(4)$ の有意性検定に意味があるのは明らかであろう．ただし季節要因の場合を除くとこのような事前情報があるのは例外的であり，一般的な検定方法としては一階の標本自己相関係数 $r(1)$ に限定することが多い．このような理由でこれまで $\rho(1)$ に関する仮説についてのいくつかの精密な検定方法の開発に力がそそがれてきた．そこで，その方法については以下で述べることにしよう．

ところで以上述べたような問題はあるが，コレログラム上に横軸上に高さ $2/\sqrt{T}$ の直線と下方に同じ高さの直線の二本の直線をプロットすることの有用性についてつけ加えて述べておこう．これらの直線は独立性からの乖離を評価する基準として用いることができる．もし時系列がホワイト・ノイズであるならばほとんどの標本自己相関は二本の直線間に入るであろう．そしてその外側にとび出ることは何らかの非独立性の左証となるであろう．もっともここでホワイト・ノイズ過程が正しいモデルの場合にも，標本自己相関係数の中で約 20 個に 1 個は二直線の範囲外に出ることにも注意を払うべきである．

### フォン・ノイマン比

フォン・ノイマン比は $\bar{y}$ を標本平均とするとき，

$$\text{VNR} = \frac{T}{T-1}\left[\sum_{t=2}^{T} (y_t - y_{t-1})^2 \Big/ \sum_{t=1}^{T} (y_t - \bar{y})^2\right] \tag{2.1}$$

によって定義されるが，この量は一階の標本自己相関 $r(1)$ と密接に関係している．$y_t - y_{t-1} = y_t - \bar{y} - (y_{t-1} - \bar{y})$ と書き直して (2.1) に代入すると，

$$\text{VNR} \simeq 2[1 - r(1)] \tag{2.2}$$

となる．ここで近似は最終観測期の扱いのために生じるのであるが，標本が小さくなければその影響は無視されよう．

もし観測量がホワイト・ノイズで正規分布に従っている，すなわち $y_t \sim NID(0, \sigma^2)$ ならば (2.1) の精密分布は知られている．この場合には Hart (1942) の有意点を用いて精密な検定が可能である．さて (2.2) より $r(1) = 0$ のときフォン・ノイマン比は近似的に 2 となり，$r(1)$ が 1 に近づくにつれてフォン・ノ

イマン比はゼロに近づいてゆくことがわかる．そこで正の系列相関に対する片側検定では，ノイマン比が適当な有意点よりも小さければ帰無仮説を棄却することになる．逆に負の系列相関についての検定では，分布の上側に棄却域を考える．ある片側対立仮説に対してフォン・ノイマン比が一様最強力不偏検定であることを Anderson (1971, 6章) が示している．

もし $T$ が十分大きければノイマン比の分布は近似的に平均 $2T/(T-1)$，分散 $4/T$ の正規分布となる．この分布を検定に用いることもできるが，それは結局，直接 $r(1)$ を用いて検定を行なうことと同等である．

### 風呂敷検定統計量

標本自己相関関数の漸近分布の結果より，独立性の検定方法として始めの $P$ 個の自己相関に基づく統計量として，例えば，

$$Q = T\sum_{\tau=1}^{P} r^2(\tau) \tag{2.3}$$

などが考えられよう．ホワイト・ノイズ過程からの観測量に対しては大標本において $r(\tau)$ は互いに独立に分布するので $Q$ は漸近的に $\chi_P^2$ に従う．ここで標本自己相関係数の数 $P$ がゼロに近ければ $Q$ は漸近分布よりも大きくなる傾向があることに注意しよう．

ここで，$P$ の選択は任意であるが，$P$ を大きくすると比較的高次で有意な $r(\tau)$ をとらえうることになる．他方，$P$ を大きくとると検出力が低下する．例えば真のモデルが MA (1) 過程であるとしよう．このときには $r(1)$ に基づく検定やフォン・ノイマン検定は，この対立仮説に対し比較的検出力が高いのに対し，風呂敷 (portmanteau) 検定はそれほど有効ではないであろう．$r(1)$ がややゼロから離れていたとしても，その $Q$ に対する影響は他の $P-1$ 個の標本自己相関によって薄められる．したがってこの風呂敷検定で有意でない結果は自己相関での顕著な特色を見逃しているかもしれないので，これをコレログラムの代りとすることはできないのである．

この風呂敷検定は漸近的に正当化されるが，小標本においては帰無仮説の下で $\chi_P^2$ 分布は $Q$ の分布の近似としてそれほど優れたものではない．この点についてはより満足できる修正された統計量として，

$$Q^* = T(T+2)\sum_{\tau=1}^{P}(T-\tau)^{-1}r^2(\tau) \tag{2.4}$$



をあげることができる．$(T-\tau)/T(T+\tau)$ が $1/T$ よりも $r(\tau)$ の分散の近似として，より良いとの合理づけによって Ljung=Box (1978) は (2.4) を提案している．

風呂敷検定を用いることの合理性は実用的な点にある．独立性からいかなる形であれ乖離すれば定義より標本自己相関に影響が及び，風呂敷検定統計量はそれをひろいあげるように作られている．さらにその漸近分布は標本自己相関の分布よりほとんど自明である．ところでこの風呂敷検定についてもう少し別の正当化をすることができる．すなわち，ホワイト・ノイズの帰無仮説に対し AR($P$) 過程，あるいは MA($P$) 過程を対立仮説とするラグランジュ乗数検定は (2.3) に基づく風呂敷検定に一致する．高次の AR 過程や MA 過程を任意の線型確率への近似とみなせば風呂敷検定は非独立性に関する一般的な検定方法を与えているのである．

ここで対立仮説としてある AR($P$) 過程を考え，さらに簡単化のためにその確率過程の平均をゼロとしよう．任意の AR($P$) 過程の漸近情報行列は $\phi$ と $\sigma^2$ に関してブロック対角であるからラグランジュ乗数統計量は 1 章 (3.5) より $\psi=\phi$ として直ちに求めることができる．さて 5 章 (4.2) より，

$$IA(\phi) = V_p \tag{2.5}$$

と表現できるが，ここで $\sigma^2 V_p$ は $p$ 個の連続した時刻の観測量の共分散行列である．帰無仮説 $H_0: \phi_1=\phi_2=\cdots=\phi_p=0$ の下で $V_p$ は単位行列となるのでラグランジュ乗数統計量を $\phi=0$ で評価すると，

$$\mathrm{LM} = T^{-1}\left(\frac{\partial \log L}{\partial \phi}\right)'\left(\frac{\partial \log L}{\partial \phi}\right) = T^{-1}\sum_{i=1}^{P}\left(\frac{\partial \log L}{\partial \phi_i}\right)^2 \tag{2.6}$$

となる．条件付対数尤度関数は，

$$\log L = -(1/2)\,T\log 2\pi - (1/2)\,T\log\sigma^2 - (1/2)\,\sigma^{-2}S(\phi)$$

であり，$S(\phi)$ は 5 章 (2.3) で与えられている．ここで $\phi$ の要素について微分すると，

$$\frac{\partial \log L}{\partial \phi_i} = \sigma^{-2}\sum_{t=P+1}^{T} y_{t-i}\,(y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-P}), \qquad i=1,\cdots,P \tag{2.7}$$

となる．$H_0$ の下では $\tilde{\sigma}_0^2 = T^{-1}\sum_t y_t^2$ であるから (2.7) は

$$\left.\frac{\partial \log L}{\partial \phi_i}\right|_{\phi=0} = \tilde{\sigma}_0^{-2}\sum_{t=P+1}^{T} y_{t-i}y_t = Tr(i), \qquad i=1,\cdots,P \tag{2.8}$$

となる．これを(2.6)に代入するとラグランジュ乗数統計量は，

$$\mathrm{LM} = T\sum_{\tau=1}^{P} r^2(\tau)$$

となり，自己共分散関数に表われるべき標本平均 $\bar{y}$ を除き(2.31)で与えられる風呂敷統計量に一致する．最後の変形は単に確率過程の平均をゼロと仮定したためである．

さらに風呂敷検定は MA$(p)$過程を対立仮説とした検定方法としても導かれる．AR$(p)$過程の場合と同様にして仮説の下での情報行列は $P\times P$ の単位行列の $T$ 倍に等しい．ラグランジュ乗数統計量は(2.6)に類似の形となり，$\partial \log L/\partial\phi_i$ を

$$\frac{\partial \log L}{\partial \theta_j} = \sigma^{-2}\sum_{t=1}^{T}\left(\epsilon_{t-j} + \sum_{i=1}^{q}\theta_i\frac{\partial \epsilon_{t-i}}{\partial \theta_j}\right)\epsilon_t, \qquad j=1,\cdots,P$$

に置き換えればよい．残差 $\epsilon_t$ とその導関数はともに5章(3.3)と(3.4)の漸化式に従うが，帰無仮説の下では $\theta=0$ であって $\epsilon_t = y_t$ となる．したがって，

$$\left.\frac{\partial \log L}{\partial \theta_j}\right|_{\theta=0} = Tr(j), \qquad j=1,\cdots,P \tag{2.9}$$

となって前の形と同一になる．

ところで対立仮説が AR 過程と MA 過程の双方についてのもう一つのアプローチは，条件付残差二乗和の最小化を尤度関数の最大化とみなして扱う方法がある．この場合には $R^2$ を残差 $\epsilon_t$ をその導関数ベクトル $z_t$ へ回帰することによって得られる重相関係数とすれば，検定は統計量 $T\cdot R^2$ によって行なわれる．どちらの確率過程の場合にもこのことは単に $y_t$ を $y_{t-1}, y_{t-2}, \cdots, y_{t-P}$ に回帰することに帰着されることになり，$TR^2$ 統計量は $H_0$ の下で自由度 $P$ の $\chi^2$ 変数とみなして検定が行なわれることになる．

### 累積ピリオドグラム

周波数領域において風呂敷検定に対応するのが累積ピリオドグラム検定である．この検定は **3.2** 節で定義したピリオドグラム座標 $p_j^2$ に基づき，

$$s_i = \sum_{j=1}^{i} p_j^2 \Big/ \sum_{j=1}^{n} p_j^2, \qquad i=1,\cdots,n \tag{2.10}$$

なる一連の統計量が用いられる．検定方法は累積ピリオドグラムとして知られている $i$ に対する $s_i$ のプロットによって行なわれる．$p_j$ 自体はきわめて不安定な値をとるが，それを累積した累積ピリオドグラムは安定的である．すなわちピリオドグラム自体はあまり価値がないのに対し，累積ピリオドグラムの方はモデルの診断法の一つとなりうる．



ホワイト・ノイズ過程においては $i$ に対する $s_i$ のグラフ上で $s_i$ は 45° 近くにプロットされねばならない．他方，3章図3.2のように低周波数が強い確率過程に対しては累積ピリオドグラムは 45° 線より上方に現われる．これとは対照的に高周波数が強い確率過程に対しては，累積ピリオドグラムは 45° 線より下方に現われることになる．

独立性の検定は 45° 線 $s=i/n$ に平行な二直線によって行なわれる．二本の直線を

$$s = \pm c_0 + i/n \tag{2.11}$$

によって決めよう．ただし $c_0$ は $n$ に依存する有意水準を表わし，Durbin (1969)にその数値が与えられている．その数値は Harvey (1981 a, pp. 364-365)に表Cとして与えられているが，そこでは有意水準は $n-1$ に対応づけられている．水準 $\alpha$ の両側検定に対しては $\alpha/2$ に対応した $c_0$ を用いる．もし標本より計算される軌跡 $s_1, \cdots, s_n$ が(2.11)のどちらかの線を越えれば帰無仮説は棄却されることになる．

ところで，場合によっては片側検定が適当なこともある．例えば，もし低周波数が過大とする対立仮説が適当なときには，上側の直線のみが意味を持っている．低周波数が過大であることは正の系列相関に対応しているので，このような対立仮説が妥当することがしばしばある．そのときには有意水準は片側検定として水準 $\alpha$ に対応した欄を見ればよい．

$T$ が奇数であるときには片側検定の水準は厳密である．しかし両側検定や $T$ が偶数の場合の片側検定では，水準は近似値であるがその近似は無視できるほど正確である．またこの検定では必ずしも実際に $s_i$ のグラフを書く必要はなく，例えば両側検定の場合には，もし

$$\max_i |s_i - i/n| > c_0$$

ならば仮説 $H_0$ を棄却すればよいこととなる．

## 6.3　モデル選択

前章においては ARMA$(p,q)$ モデルの推定問題が $p$ と $q$ があらかじめ与えら

れているものとして議論された．ところが定常時系列データのモデル分析では，$p$ や $q$ が既知でないのが一般的である．そこで推定に先だって，一つのモデルを選択しなければならないが，それは最小限の未知母数であてはまりのよい $p$ と $q$ を選ぶことによって実行される．ボックス=ジェンキンズはこの方式を「ケチの原理」(principle of parsimony) と呼んでいる．

さて Harvey (1981 a) は 5.8 節で，モデル選択の一般原理を提示した．そこではまず最も一般的モデルから出発して，次々に制約条件を検定してゆくのが最適であると主張している．その方法はもし純粋 AR (あるいは MA) 過程が妥当と考えられる場合には適用することができよう．AR ($p$) モデルが与えられるとき検定すべき最初の仮説は自己回帰過程の次数が $p$ であることである．次の仮説は次数が $p-1$ となる等々として，モデルの次数はある特定の仮説が受け入れられたときに決定することになる．この方法については Anderson (1971, pp. 270-276) が詳しく説明している．ところでこの方法に対しては二つの反論がありうる．第一にはそれほど重要ではないが，この方式によって決定される次数は必ずしも適当でないという反論である．最適なモデルは非負係数の数よりも高い次数，たとえば $\phi_1 = 0$ である AR (2) 過程となる可能性がある．ただし季節データを除くとこうしたモデルの定式化はそれほど一般的でないし，次数 $p$ が決まればゼロとなる係数をテストすることは困難ではない．より基本的問題としては，任意の線型時系列過程を近似するには比較的多くの自己回帰係数を用いなければならないことであろう．このことは次の二点を意味する．第一に始めのモデルを十分高次モデルとする必要がある．さらに逐次検定方式によって，選択されるモデルが非常に多くの母数を含むことになりかねないので，先ほどの「ケチの原理」に反することになりやすい．

混合 ARMA モデルの一つの特徴は多くの場合比較的少数の母数で任意の線型確率過程を近似できることにある．ただし混合 ARMA モデルの場合には，仮説の自然な順序づけが困難となる別の問題が生じる．さらにモデルが過剰母数を含むときには一つの技術的問題も生じる．例えば，簡単な例としてはホワイト・ノイズ過程からの観測量に ARMA (1, 1) モデル，

$$y_t = \phi y_{t-1} + \epsilon_t + \theta \epsilon_{t-1} \tag{3.1}$$

をあてはめることを考えよう．モデル (3.1) は $\phi = -\theta$ のときホワイト・ノイズ過程となるが，情報行列 (5 章 (4.7) 参照) は特異になる．この場合多くの最尤推定の計算法では，情報行列と漸近的に同等な行列を用いるために計算上困



難な問題が生じる．計算上 $\tilde{\phi} \fallingdotseq -\tilde{\theta}$ の値に収束したとしても，情報行列の推定値が特異に近くなると標準誤差推定値が大きな値をとり，このことが過剰母数であることを示唆することとなる．

以上のような問題をふまえてボックス=ジェンキンズは，三段階の手続きからなるモデル選択の戦略を提唱している．第一段階は識別 (identification) と呼ばれ，コレログラムや他の関連した統計量をもとにして一つのモデルを選択する．次に $p$ と $q$ が与えられたとして，前章で述べた方法の一つによって母数を推定する．この推定 (estimation) の段階より残差が計算され，それに基づいてモデルの診断 (diagnostic checking) が行なわれる．独立性からの乖離が有意であれば，何らかの意味でモデルが不適切となり識別段階へ再び戻り，適切なモデルが見い出されるまでこの手続きが繰り返されることになる．

### 識　　別

適切な ARMA ($p, q$) 過程を識別する最も重要な手がかりはコレログラムである．標本数が十分大きければ，コレログラムは確率過程の自己相関の理論値を反映する．2 章においてすでに自己相関関数の挙動を調べ，いくつかの確率過程についてその特色に注意を払った．すなわち MA ($q$) 過程では $q$ 以上の時差の自己相関は切断され，他方 AR 過程や混合 ARMA 過程の場合には徐々にゼロへ向って減衰してゆく．例えば図 6.1 で与えられるコレログラムからは多分 MA (2) モデルがまず選ばれるであろう．$\tau = 2$ 以上の自己相関はゼロに一致しないが，$r(1)$ と $r(2)$ に比較すると十分小さく主に標本誤差によるのではないかと推察される．

もし $q$ 以上の次数の理論上の自己相関がゼロであるならば，標本自己相関は近似的に平均ゼロ，分散

$$\text{Avar}[r(\tau)] = \left[1 + 2\sum_{j=1}^{q} \rho(j)^2\right] \Big/ T, \qquad \tau > q \tag{3.2}$$

となることが知られている．Avar $[r(\tau)]$ の一つの推定値は (3.2) における自己相関の理論値を標本自己相関に置き換えることによって得ることができよう．図 6.1 の破線は MA (2) 過程の仮定の下で $\tau > 2$ となる $r(\tau)$ の近似的 95% 信頼区間を示している．

純粋 AR ($p$) 過程については，$p = 1$ の場合を除いてコレログラムより次数を決定するのは困難である．それを補完する方法としては標本偏自己相関関数

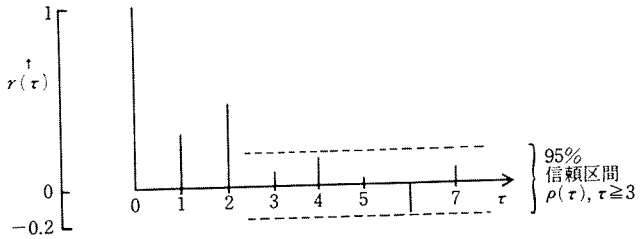



**図 6.1** ある MA(2)過程から 200 個の観測値に対する標本自己相関関数.

(sample partial autocorrelation function)がある．AR($\tau$)モデルにおいて $\tilde{\phi}(\tau)$ を $y_{t-\tau}$ の係数推定値としよう．このとき標本偏自己相関関数は，$\tau$ に対する $\tilde{\phi}(\tau)$ のグラフとして定義される．純粋 AR($p$)過程では $p$ 以上の偏自己相関の理論値はゼロとなるのに対し，MA($q$)過程では徐々に減衰してゆく．もし観測値が AR($p$)過程から生成されたものであれば，$p$ 次以上の標本偏自己相関は漸近的に平均ゼロで，分散

$$\text{Avar}\,[\tilde{\phi}(\tau)] = 1/T \tag{3.3}$$

の正規分布に従っている．

以上からもしデータが純粋 AR モデルや純粋 MA モデルで近似されれば，その自己相関と偏自己相関の推定値を調べることによって適当な次数 $p$ と $q$ を選ぶことができる．混合 ARMA モデルの場合には，次数 $p$ と $q$ の識別はより困難となる．というのは，この場合自己相関関数と偏自己相関関数はともに明確な切断点を持たず，そのパターンを解釈するには相当の判断力を要するからである．したがって，適切なモデルが選択されるまで識別－推定－診断のプロセスを何度も繰り返すことが必要となろう．

ところで標本誤差は累積することによって一層やっかいな問題が生じる．例えば，もとのモデルが AR(1)過程のとき $r(\tau)$ と $r(\tau+1)$ の相関はほぼ

$$\left(\frac{1}{T}\right)\frac{\phi^k[(k+1)-(k-1)\phi^2]}{1-\phi^2} \tag{3.4}$$

となる．$\phi$ がそれほど大きくなく，例えば 0.5 とすると $r(\tau)$ と $r(\tau+1)$ の相関は 0.80 となる．すなわち，標本誤差は累積的に影響を及ぼしコレログラムは真の自己相関関数を歪めて示す可能性がある．さらにこれに関連する問題としては，コレログラムは対応する自己相関関数ほどなめらかに減衰しない．



Kendall(1973, pp. 84, 90)の図はこの現象をうまく例示している．

## 診　　断

残差の独立性からの乖離を調べることによってモデルの診断を行なうことができる．残差のコレログラムや累積ピリオドグラムに基づいて検定方法を考えることができるが，それにもまして残差のプロットが重要である．Box=Jenkins(1976, p. 289)の言葉を引用するならば，「残差プロットを視覚的に調べることがモデルの診断に際しての第一歩であることをいくら強調してもしすぎることはない」．

むろん残差のプロットと残差からの統計量を調べることはきわめて有益であるが，結果の解釈に際しては十分に注意する必要がある．しばしば残差に関連した統計量は，モデルが正しく定式化されていれば残差は撹乱項自体と同一の性質を有する，との仮定の下で構成される．しかし残念ながら ARMA モデルの残差については，この命題は大標本でも成立しない．例えば AR(1)モデルをデータにあてはめると残差からの一次の自己相関係数の漸近分散は $\phi^2/T$ となって，ホワイト・ノイズ過程からの $r(1)$ の分散である $1/T$ よりも小さくなりうるが，このことは任意の ARMA($p, q$)過程からの残差の自己相関の挙動についてある程度示唆を与えている．一般に低次数においては分散は小さくなる傾向があり，しかも $r(\tau)$ は互いに高く相関している．通常は高次においてこのような影響は消滅するが，しばしば風呂敷統計量や累積ピリオドグラムに対し重大なバイヤスをもたらすことがあり，そしてそれは検定がホワイト・ノイズからの乖離の有意性を過小評価する方向に影響を及ぼしがちとなる．

このような状況では，グラフによる検定方法は一般的に正しいとはいいきれないが，有益なガイド・ラインとなっている．もちろん適当な検定統計量を用いることが望ましいことも明らかであろう．Box=Pierce(1970)は $P$ が十分大きければ ARMA モデルの残差から作られた風呂敷統計量の分布が帰無仮説の下で $\chi^2_{P-p-q}$ とみなしうることを示した．母数を推定したことで自由度が減り，自由度 $P$ の $\chi^2$ 分布に基づく風呂敷検定と比較すると，帰無仮説を棄却する確率をより大きくする傾向を示している．さらにこの統計量は小標本では **6.2** 節に述べた欠点を持つので，修正統計量(2.4)を用いることが望ましい．ただし，ここで有意水準は $\chi^2_{P-p-q}$ より求めるべきである．(2.3)と(2.4)を用いる方法のことはしばしばボックス=ピアース(Box=Pierce)検定及び修正，ボックス=ピ

アース(Box=Pierce)またはボックス=リュング(Box=Ljung)検定として言及される.

以上の方法とは別にあるARMA$(p,q)$モデルの定式化を検定するために,そのモデルを含むARMA$(p+P,q)$過程を考える方法がある.モデルにおいて仮説$H_0:\phi_{p+1}=\phi_{p+2}=\cdots=\phi_{p+P}=0$にラグランジュ乗数検定を用いることを考えよう.残差$\epsilon_t$を仮説$H_0$の下で$p+P+q$個の微係数に回帰して計算された統計量$TR^2$を$\chi^2_P$に従う変数とみなすことによって検定が行なわれる.さらにこのラグランジュ乗数検定の修正としては$\phi_{p+1}$から$\phi_{p+P}$までの微係数の有意性に対し近似的に$F$検定を用いることが考えられる.

**例1** ARMA$(1,1)$モデルが推定されたとしよう.このモデルをARMA$(1+P,1)$モデルに対して検定を行なうと,残差

$$\epsilon_t=y_t-\phi_1 y_{t-1}-\cdots-\phi_{1+P}y_{t-1-P}-\theta\epsilon_{t-1},\qquad t=2+P,\cdots,T \tag{3.5}$$

の微係数を$\epsilon_1=\epsilon_2=\cdots=\epsilon_{1+P}=0$として評価すると$i=1,\cdots,1+P$に対して,

$$\frac{\partial\epsilon_t}{\partial\phi_i}=-y_{t-i}-\theta\frac{\partial\epsilon_{t-1}}{\partial\phi_i}=-\sum_{j=0}^{t-P-2}(-\theta)^j y_{t-i-j},\qquad t=2+P,\cdots,T \tag{3.6 a}$$

および

$$\frac{\partial\epsilon_t}{\partial\theta}=-\epsilon_{t-1}-\theta\frac{\partial\epsilon_{t-1}}{\partial\theta}=-\sum_{j=0}^{t-P-2}(-\theta)^j\epsilon_{t-1-j},\qquad t=2+P,\cdots,T \tag{3.6 b}$$

となる.ここで$\epsilon_{1+P}$は固定されていると仮定しているので,時刻$t=1+P$におけるすべての微係数はゼロとなる.仮説の下でこの表現を評価すると$\phi_1,\theta,\epsilon_t$をその推定値で置き代えて,$y_t-\phi_1 y_{t-1}-\theta\epsilon_{t-1}$を(3.6)に回帰することとなり,その結果計算される統計量$TR^2$に$\chi^2_P$分布を用いて検定を行なうことができよう.

さてMA要素を大きめにとりARMA$(p,q+P)$として別のラグランジュ乗数検定を作ることもできる(5章の例3はその特殊な例である).一見するとこの方法では前に述べた検定とは異なる統計量が導出されるようであるが,実は二つの統計量$TR^2$は一致することがPoskitt=Tremayne(1980)によって示されている.

最後にラグランジュ乗数検定について注意すべき点としては,それがいかなる$P$についても正当化されることである.このことはボックス=ピアース検定が漸近的に正当化されるためには十分大きい$P$,例えば$P\log T$を取る必要があるのに比較して,利点としてあげることができよう.



### 適合度

一つの応用例に際しボックス=ジェンキンズのモデル選択の手続きが複数の定式化を導き,その各々がモデル診断基準を満たすこともありうる.その場合には,異なるモデルを区別するある種の適合度の尺度が必要となろう.**1.3**節で述べた赤池の情報量基準(Akaike Information Criterion : AIC)はその一つの可能性であろう.AIC統計量は,

$$\mathrm{AIC}^{\dagger}=\tilde{\sigma}^2\exp\left[\log|V|+2(p+q)/T\right] \tag{3.7}$$

と簡潔に表現される.ただし$\sigma^2V$は$y$の共分散行列である.もしモデルが完全最尤法でなく制約条件最小二乗法で推定される場合には$\log|V|$項は消える.そこで$\mathrm{AIC}^{\dagger}$が最小となるモデルが選択される.$2(p+q)/T$項は「ケチの原理」に反することへのペナルティ(罰金)を意味している.

ボックス=ジェンキンズ法の欠点の一つは,とくに識別段階で相当の時間と技術を必要とすることであろう.それに対しもう一つのアプローチとしてAICを全般的に用いることが考えられよう.可能なモデルの範囲を事前情報で定め,大ざっぱな識別手順によって(3.7)を最小化するようにモデルを選択する.この方法では計算機による計算がより重要な役割を演じる.しかしながら,手続きが半ば自動的に適用できるからといってこうして選択されたモデルに対して,全く診断手続きをせずに済ませることはあまり賢明ではないであろう.

## 6.4 予　　測

最小平均二乗推定量(MMSE)の概念は4章で導入されたが,この推定量は予測誤差分散を最小化している.ここで$\hat{y}_{T+l|T}$を$y_{T+l}$の最小平均二乗誤差推定量とすると,少なくとも$E[(y_{T+l}-\hat{y}_{T+l|T})^2]$はあらゆる他の予測量の平均二乗誤差より大きくならない.

ここでまず次の三つの仮定をもうけることにしよう.第一に撹乱項$\epsilon_t$は正規分布に従う.正規性の下では最小平均二乗推定量は最小平均二乗線型推定量となるので,この仮定によって以下の議論は線型推定量に限定することになる.第二にARMAモデルの母数が既知であると仮定する.このことは実際には正しくないが,標本数が小さくなければ母数の推定誤差は予測にあまり影響を与えない.この問題については本節の最後で多少詳しく言及するであろう.最後

に MA 過程と混合過程については，現在と過去のすべての撹乱項 $\epsilon_T, \epsilon_{T-1}, \cdots$ は既知であると仮定しよう．このことは事実上観測量の無限の実現 (realization) を仮定することを意味する．この仮定は $T$ が小さかったり MA 過程の随伴多項式の根の中に単位円周上あるいはその近くに来るものがない限りそれほど制約的ではない．最後の点が重要となる場合には有限標本予測 (finite sample prediction) が適切となるが，それについては後にふれることにする．

時刻 $T+l$ における ARMA $(p, q)$ 過程は

$$(4.1)\quad y_{T+l} = \phi_1 y_{T+l-1} + \cdots + \phi_p y_{T+l-p} + \epsilon_{T+l} + \cdots + \theta_q \epsilon_{T+l-q}$$

で与えられる．$\epsilon_t$ の将来値は未知であり予測することはできない．そこで $y_{T+l}$ の予測に (4.1) を用いるときには，$t > T$ に対する $\epsilon_t$ をその期待値に等しくおこう．$y_t$ の将来値もやはり未知であるが，予測は漸化式

$$(4.2)\quad \tilde{y}_{T+l/T} = \phi_1 \tilde{y}_{T+l-1/T} + \cdots + \phi_p \tilde{y}_{T+l-p/T} + \tilde{\epsilon}_{T+l/T} + \cdots + \theta_q \tilde{\epsilon}_{T+l-q/T}, \qquad l = 1, 2, \cdots$$

によって行なうことができる．ここで $j \leq 0$ に対し $\tilde{y}_{T+l/T} = y_{T+j}$ かつ

$$\tilde{\epsilon}_{T+j/T} = \begin{cases} 0 & \text{for } j > 0 \\ \epsilon_{T+j} & \text{for } j \leq 0 \end{cases}$$

である．

**例 1**　AR (1) 過程に対して (4.2) は，

$$(4.3)\quad \tilde{y}_{T+l/T} = \phi \tilde{y}_{T+l-1/T}, \qquad l = 1, 2, \cdots$$

となる差分方程式を導く．出発値は $\tilde{y}_{T/T} = y_T$ で与えられるので，(4.3) を解くと

$$(4.4)\quad \tilde{y}_{T+l/T} = \phi^l y_T$$

となる．したがって予測値は指数的にゼロに近づき，予測関数は自己共分散関数と全く同一の形になる．

**例 2**　時刻 $T+1$ において MA (1) 過程は

$$(4.5)\quad y_{T+1} = \epsilon_{T+1} + \theta \epsilon_T$$

と表現される．$\epsilon_{T+1}$ は未知であるから，期待値ゼロで置き換えると予測方程式は

$$(4.6)\quad \tilde{y}_{T+1/T} = \theta \epsilon_T$$

となる．$l > 1$ に対しては $\tilde{y}_{T+l/T} = 0$ となるので，この過程から生成されたデータに関する知識は 1 期先以上の予測には役に立たないことになる．

**例 3**　ARMA (2, 2) 過程

$$(4.7)\quad y_t = 0.6 y_{t-1} + 0.2 y_{t-2} + \epsilon_t + 0.3 \epsilon_{t-1} - 0.4 \epsilon_{t-2}$$

を考え，さらに $y_T = 4.0, y_{T-1} = 5.0, \epsilon_T = 1.0, \epsilon_{T-1} = 0.5$ としよう．このときには



$$\tilde{y}_{T+1/T} = 0.6 y_T + 0.2 y_{T-1} + 0.3 \epsilon_T - 0.4 \epsilon_{T-1} = 3.5$$

および

$$\tilde{y}_{T+2/T} = 0.6 \tilde{y}_{T+1/T} + 0.2 y_T - 0.4 \epsilon_T = 2.5$$

となる．さらに $l = 3, 4, \cdots$ に対する予測量は差分方程式，

$$\tilde{y}_{T+l/T} = 0.6 \tilde{y}_{T+l-1/T} + 0.2 \tilde{y}_{T+l-2/T}, \qquad l = 3, 4, \cdots$$

で与えられる．

### 最適予測とその平均二乗誤差

例 2 の MA (1) モデルでは $\tilde{y}_{T+1/T}$ の平均二乗誤差は直ちに求めることができる．(4.5) と (4.6) より，

$$\mathrm{MSE}\,(\tilde{y}_{T+1/T}) = E[(y_{T+1} - \tilde{y}_{T+1/T})^2] = E(\epsilon_{T+1}^2) = \sigma^2$$

となる．$l > 1$ に対しては $\tilde{y}_{T+l/T} = 0$ となるので，$\mathrm{MSE}\,(\tilde{y}_{T+l/T}) = \mathrm{Var}\,(y_{T+l}) = (1+\theta^2)\sigma^2$ であることは自明であろう．任意の MA $(q)$ 過程についても予測の平均二乗誤差は直ちに計算できる．これに対して，AR 過程や混合過程における予測の平均二乗誤差はそれほど自明ではない．任意の ARMA 過程は無限次 MA 過程

$$(4.8)\quad y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

と表現される．ただしここで係数 $\psi_j$ は 2 章 (4.6) によって定められる．さて観測量に関する任意の線型予測量は撹乱項についても線型でなければならない．撹乱項 $\epsilon_t$ の将来値は未知であるから係数 $\psi_l^*, \psi_{l+1}^*, \cdots$ を用いて，

$$(4.9)\quad \hat{y}_{T+l/T} = \psi_l^* \epsilon_T + \psi_{l+1}^* \epsilon_{T-1} + \cdots, \qquad l = 1, 2, \cdots$$

と表わすことができる．ところで

$$y_{T+l} - \hat{y}_{T+l/T} = \epsilon_{T+l} + \psi_1 \epsilon_{T+l-1} + \cdots + \psi_{l-1} \epsilon_{T+1} + (\psi_l - \psi_l^*) \epsilon_T + (\psi_{l+1} - \psi_{l+1}^*) \epsilon_{T-1} + \cdots$$

の期待値はゼロであるから，$\hat{y}_{T+l/T}$ は $u$-不偏であってその平均二乗誤差は

$$(4.10)\quad \mathrm{MSE}\,(\hat{y}_{T+l/T}) = \sigma^2 (1 + \psi_1^2 + \cdots + \psi_{l-1}^2) + \sigma^2 \sum_{j=0}^{\infty} (\psi_{l+j} - \psi_{l+j}^*)^2$$

で与えられる．(4.10) は $\psi_{l+j}^* = \psi_{l+j}$ とすると最小になる．したがって $y_{T+l}$ の最小平均二乗推定量は，

$$(4.11)\quad \hat{y}_{T+l/T} = \sum_{j=0}^{\infty} \psi_{l+j} \epsilon_{T-j}$$

となり，その平均二乗誤差は

(4.12) $$\mathrm{MSE}\,(\tilde{y}_{T+l/T}) = (1+\psi_1^2+\cdots+\psi_{l-1}^2)\,\sigma^2$$

となる．

**例 4** AR(1)モデルについては，$\psi_{l+j}=\phi^{l+j}$ となるので，

$$\tilde{y}_{T+l/T} = \sum_{j=0}^{\infty}\phi^{l+j}\epsilon_{T-j} = \phi^l\sum_{j=0}^{\infty}\phi^j\epsilon_{T-j} = \phi^l y_T$$

である．((4.4)参照)．AR(1)モデルにおける $y_{T+l/T}$ の平均二乗誤差は

(4.13) $$\mathrm{MSE}\,(\tilde{y}_{T+l/T}) = [1+\phi^2+\cdots+\phi^{2(l-1)}]\sigma^2$$
$$= \frac{1-\phi^{2l}}{1-\phi^2}\cdot\sigma^2$$

で与えられる．ここで $l\to+\infty$ とすると $y$ の分散 $\sigma^2/(1-\phi^2)$ に近づく．

さて $y_{T+l}$ の最小平均二乗推定量は攪乱項の線型関数として表現されるので，攪乱項が正規分布に従うとき同じく正規分布に従う．これより $y_{T+l}$ の95%予測区間は，

(4.14) $$y_{T+l} = \tilde{y}_{T+l/T} \pm 1.96\left(1+\sum_{j=1}^{l-1}\psi_j^2\right)^{1/2}\sigma$$

で与えられる．

### 有限標本予測と反転可能性*

MA(1)モデルにおける $y_{T+1}$ の最小平均二乗推定量は $\epsilon_T$ に依存していた．ところで仮に $\theta$ が既知としても $\epsilon_T$ は有限個の標本 $y_1,\cdots,y_T$ からは正確に求めることができない．実際の予測に際しては $\epsilon_T$ を5章(3.3)の漸化式より初期値 $\epsilon_0(\theta)=0$ から出発して求めた残差 $e_T=\epsilon_T(\theta)$ が用いられる．

ここで $\epsilon_T$ を $e_T$ で置き換えることによって予測量の平均二乗誤差は増加する．$\epsilon_t = y_t-\theta\epsilon_{t-1}$ の代入を繰り返すことによって，

$$\epsilon_t = y_t+(-\theta)y_{t-1}+\theta^2 y_{t-2}+\cdots+(-\theta)^{t-1}y_1+(-\theta)^t\epsilon_0$$

となる．この表現は $\epsilon_0$ をゼロで置き換えたことを除き最小二乗の漸化式に対応している．したがって平均二乗誤差公式は，

(4.15) $$\mathrm{MSE}\,(\tilde{y}_{T+1/T}) = E\{[\epsilon_{T+1}+\theta(\epsilon_T-e_T)]^2\} = \sigma^2[1+\theta^{2(T+1)}]$$

となる．ここでもし $T$ が十分大きければ $|\theta|<1$ の仮定の下で $\epsilon_T$ と $e_T$ の差は無視されよう．すなわち $T\to\infty$ のとき(4.15)は $\sigma^2$ に近づく．他方 $T$ が比較的小さく $|\theta|$ が1に近ければ，$e_T$ を用いる予測量は(4.6)とは相当違っているであろう．$\theta$ が単位円周上にあるときはその最も極端な場合であろう．このときMA過程は反転可能ではなくなり，予測誤差は $y_{T+1}-\theta e_T = \epsilon_{T+1}-(-\theta)^{T+1}\epsilon_0 = \epsilon_{T+1}\pm\epsilon_0$ となり標本数の大小にかかわらず $\epsilon_0$ に依存し，予測の平均二乗誤差は $2\sigma^2$ となる．



MA過程や混合過程において将来観測値の精密な最小平均二乗推定量を求める問題はしばしば有限標本予測(finite sample prediction)問題と呼ばれる．その一つの解決法としてはカルマン・フィルターによる方法があり，この方法については **5.3** 節ですでに述べた．与えられた $\phi$ と $\theta$ に対し $y_{T+l}$ の最小平均二乗推定量は **4.5** 節で述べた多段予測の結果を用いて求めることができよう．例えば，MA(1)の場合には $f_{T+1}$ が4章(4.9)で与えられるとすると，$\mathrm{MSE}\,(\tilde{y}_{T+1/T})=\sigma^2 f_{T+1}$ となる．これを(4.15)と比較すると $\theta=0$ となる場合を除いて上述の疑似最小平均二乗推定量の平均二乗誤差よりも小さくなる．$\theta=\pm1$ となる場合にも4章(5.15)より $T\to\infty$ のときに $\mathrm{MSE}(\tilde{y}_{T+1/T})\to\sigma^2$ となる．

最後に一般の場合について，MA過程がたとえ反転可能でないときにも有限標本予測を行なうことができることに注意しよう．例えばMA(1)モデルにおいて $|\theta|>1$ のような場合である．ただしこのようなモデルでの予測はあまり魅力がないが，それは予測が有効でなくなるからである．すなわち，5章(3.9)を用いると，反転可能なモデルでは $T\to\infty$ のとき $\mathrm{MSE}(\tilde{y}_{T+1})\to\sigma^2$ となるのに対し，$|\theta|>1$ のときには $\mathrm{MSE}(\tilde{y}_{T+1})\to\sigma^2\theta^2$ となることがわかる．

### 母数推定値を用いる予測*

実際の予測においてはほとんど常に未知母数 $\phi$ と $\theta$ にその推定値が代入される．このことは予測量に推定量の変動要因が加えられることを意味するので，それを含めて予測の平均二乗誤差を評価することが望ましい．

さて，ここでAR(1)過程を例にとってみよう．母数 $\phi$ が既知ならば $l$ 期先の最小平均二乗推定量は(4.4)で与えられる．次に母数 $\phi$ が未知のときには $\phi$ をその最尤推定値，あるいはそれを漸近的に同等な推定値で置き換える．すなわち予測量は

(4.16) $$\tilde{y}^*_{T+l/T} = \tilde{\phi}^l y_T$$

となる．一般のモデルでも同様にして(4.2)と同じ差分方程式によって $\tilde{y}^*_{T+l/T}$ が計算されよう．ここで(4.16)の予測誤差は二つの項に分解され，

(4.17) $y_{T+l}-\tilde{y}^{*}_{T+l/T}=(y_{T+l}-\tilde{y}_{T+l/T})+(\tilde{y}_{T+l/T}-\tilde{y}^{*}_{T+l/T})$

と書ける．(4.17)の第一項は$\phi$が既知のときの予測誤差を示し，第二項は$\phi$の推定から生じる誤差を示している．この種の分解はあらゆるARMAモデルについて妥当する．AR(1)モデルの場合には(4.16)を用いることによって簡単化されて，

(4.18) $y_{T+l}-\tilde{y}^{*}_{T+l/T}=(y_{T+l}-\tilde{y}_{T+l/T})+(\phi^{l}-\tilde{\phi}^{l})\,y_T$

となる．

ここで一期先予測量を考えるとその平均二乗誤差は，

(4.19) $\mathrm{MSE}\,(\tilde{y}^{*}_{T+1/T})=\mathrm{MSE}\,(\tilde{y}_{T+1/T})+y_T^2E[(\tilde{\phi}-\phi)^2]$

と書ける[訳注]．(4.19)において推定誤差の大きさを評価する際には$y_T$を固定し，$\tilde{\phi}$を確率変数とみなそう．このことは$\tilde{\phi}$を構成する際に$y_T$を用いるので一見矛盾しているように見える．しかしいかなる予測も標本観測値は与えられたものとして行なわれるので，平均二乗誤差の一つの定義としては意味がないとはいえないであろう．そして$E[(\tilde{\phi}-\phi)^2]$をその漸近分散で置き換えると平均二乗誤差への一つの近似として，

(4.20) $\mathrm{MSE}\,(y^{*}_{T+l/T})\simeq\sigma^2+y_T^2(1-\phi^2)/T.$

が求まる．AR(1)モデルはしばしば最小二乗法によって推定される．通常の回帰分析の公式を$\tilde{y}^{*}_{T+1/T}$の平均二乗誤差の推定に用いると

(4.21) $\mathrm{MSE}\,(y^{*}_{T+1/T})=s^2\left(1+y_T^2\Big/\sum_{t=2}^{T}y_{t-1}^2\right)$

となる．本標本では

$$\sum_{t=2}^{T}y_{t-1}^2\simeq T\sigma^2/(1-\phi^2)$$

であるから，(4.20)とほぼ同等になる．Fuller=Hasza(1978)は最近一般のARモデルにおいても通常の回帰分析の公式が多段予測の平均二乗誤差のよい推定値であることを示している．

ところで$l>1$のときには(4.19)の最後の項は$y_T^2E[(\tilde{\phi}^{l}-\phi^{l})^2]$となる．そこで

訳注　ここではモデルが正しい場合の予測量の平均二乗誤差を問題としている．モデルの定式化が誤まっていれば，予測量の平均二乗誤差も修正されねばならない．詳しくは次の論文を参照されたい．

Kunitomo, N. and T. Yamamoto (1985), "Properties of Predictors in Misspecified Autoregressive Modeles", *Journal of the American Statistical Association*, forthcoming.



$$\phi^{l}-\tilde{\phi}^{l}=\phi^{l}-\phi^{l}\left[1-\frac{(\phi-\tilde{\phi})}{\phi}\right]^{l}$$

と表現し，カッコ内を展開し高次の項を無視すると，

$$\phi^{l}-\tilde{\phi}^{l}\simeq l\,\phi^{l-1}(\phi-\tilde{\phi})$$

となるので，

$$E[(\tilde{\phi}^{l}-\phi^{l})^2]\simeq l^2\phi^{2(l-1)}E[(\tilde{\phi}-\phi)^2]$$

である．これを(4.13)と結びつけると，

(4.22) $\mathrm{MSE}\,(y^{*}_{T+l/T})\simeq\sigma^2\dfrac{1-\phi^{2l}}{1-\phi^2}+y_T^2\dfrac{(1-\phi^2)l^2\phi^{2(l-1)}}{T}$

を得る．この(4.22)によって標本が与えられたときの多段予測量の平均二乗誤差の近似が得られる．予測量の平均的な平均二乗誤差の動きを知るには$y_T^2$をその期待値$\sigma^2/(1-\phi^2)$で置き換えればよい．すなわち

(4.23) $\mathrm{AMSE}\,(y^{*}_{T+l/T})\simeq\sigma^2\left[\dfrac{1-\phi^{2l}}{1-\phi^2}+\dfrac{l^2\phi^{2(l-1)}}{T}\right],\qquad l=1,2,\cdots$

となり，とくに$l=1$ならば

(4.24) $\mathrm{AMSE}\,(y^{*}_{T+1/T})=\sigma^2(1+T^{-1}).$

となる．

ところで，(4.23)と(4.24)は，ともに$\phi$を推定することの誤差への影響は$0(T^{-1})$のオーダーであることを示している．すなわち，ここでは$\phi$を既知としたときの最適予測量の平均二乗誤差が支配的になっている．むろん$\phi$を推定する影響を無視することによって予測誤差を過小評価することになるが，$T$がきわめて小さくない限りそのためのバイヤスはそれほど大きくないであろう．このことはより一般のモデルについても成立する．より詳しくはBox=Jenkins(1976, pp. 267-269)及びYamamoto(1976)を参照されたい．

## 6.5 定　常　性

これまで議論した時系列モデルは定常性の仮定に本質的に依存している．しかしながら多くの実際問題ではこの仮定はあまりに制約的である．時系列データをプロットしてみると平均や分散に何らかのトレンドがしばしば見受けられる．

さて，ARMA過程をもとにして非定常性を扱う手段は基本的に二つの方法

がある．第一にはデータの階差をとることによって定常系列を作り出し，階差系列に ARMA 過程をあてはめる方法が考えられる．通常 1 回か 2 回の階差をとることが必要であるが，全体のモデルのことを和分(integrated)過程と呼ぶ．第二にはトレンド項と誤差項の加法モデル(trends plus error model)

(5.1) $$y_t = f(t) + u_t$$

とよばれるものがある．ここで $u_t \sim \text{ARMA}(p, q)$ であって $f(t)$ はある非確率的な時間の関数である．ところで(5.1)における非確率的トレンド項の形について，なんの事前情報もない場合には多項式を用いることが考えられよう．すなわち $h$ を多項式の最高次数として，

(5.2) $$f(t) = \alpha + \beta_1 t + \beta_2 t^2 + \cdots + \beta_h t^h$$

を想定する方法がある．このような(5.1)タイプのアプローチの欠点としては全観測期間を通じて一つのトレンド項を仮定していることをあげることができよう．したがって，データの時間的経過に伴なうトレンドの変化に対応できずに，モデルの予測力があまりよくない可能性がある．この点に関しては階差の方法は(5.2)の大域的(global)トレンドの代りに局所的(local)トレンドを用いているのでより柔軟性がある．その場合予測はやはり多項式の外挿によって行なわれるが，予測の出発点を決めるに当って予測期間に近い観測値のウェイトが高くなっている．(**1.1** 節参照)．

以下では，まず和分過程の性質を調べ次に予測の手順を解説する．そして予測の平均二乗誤差の公式を与える．予測の平均二乗誤差は任意の将来時点における信頼度を明示するので，いかなる科学的予測法においても重要である．次に，局所的トレンドと大域的トレンドを(5.1)における非確率的トレンドと確率的トレンドに置き換えることによって関係づける．

ところで非定常な観測値を生成する方法は無限にあり得る．平均項にトレンド項を持つモデルはほんの一例にしかすぎない．ここでは共分散の変化に非定常性が反映するモデルについては説明しない．ただし分散のみに表われるトレンドについては，しばしば対数変換などの簡単な操作によって取り除くことができる．より一般的にはいわゆるボックス=コックス(Box=Cox)変換

(5.3) $$y_t^{(\lambda)} = \begin{cases} (y_t^{\lambda} - 1)/\lambda, & 0 < \lambda \leq 1 \\ \log y_t, & \lambda = 0 \end{cases}$$

が適用されよう．ところでこのボックス=コックス変換は一般性を持つが，多くの分析目的にとっては対数変換で十分なことが多い．このことはとくに経済時系列の場合によく妥当するが，それは対数の 1 回の階差が成長率の近似になっていることによるものであろう．



### 自己回帰和分移動平均過程

和分過程の最も単純な例は酔歩(random walk)過程

(5.4) $$y_t = y_{t-1} + \epsilon_t$$

であろう．このモデルは 2 章(2.2)の 1 次自己回帰モデルと同一の形をとっているが，未知母数 $\phi$ が単位円周内に含まれるという仮定を満たしていないので(5.4)は定常過程ではない．$y_t$ の過去の値を繰り返し代入すると，

(5.5) $$y_t = \sum_{j=0}^{t-1} \epsilon_{t-j} + y_0, \qquad t = 1, 2, \cdots, T$$

となる．すなわち $y_t$ の平均は初期値 $y_0$ の関数であって $E(y_t) = E(y_0)$ となる．ここでもし $y_0$ が固定されていれば定常性の第一条件である平均の時間についての一定性を満たしている．しかしながら，

(5.6) $$\text{Var}(y_t) = t\,\sigma^2$$

(5.7) $$\text{Cov}(y_t, y_{t-\tau}) = |t - \tau|\,\sigma^2$$

となることから確率過程は非定常となる．酔歩過程では初期値から出発してさまよい動くが特定のトレンドは示さない．ところでこの過程よりきわめて容易に一つの定常過程を求めることができる．(5.4)を整理すると，

(5.8) $$\Delta y_t = y_t - y_{t-1} = \epsilon_t$$

となるので一次階差はホワイト・ノイズ過程に従うことになる．

次に(5.8)を拡張して $\epsilon_t$ を一般の定常 ARMA$(p, q)$ 過程としてみよう．さらに $y_t$ の階差を何度もとることにする．ここでその階差の回数を $d$ とすると，その結果モデルは

(5.9) $$\phi(L)\,\Delta^d y_t = \theta(L)\,\epsilon_t$$

となり，ARIMA$(p, d, q)$ 過程として知られている．酔歩過程では $y_t \sim$ ARIMA$(0, 1, 0)$ となり，定常 ARMA 過程は $d = 0$ とした場合となる．

モデル(5.9)に母数 $\beta$ で表現された定数項を加え，

(5.10) $$\Delta^d y_t = \beta + \phi^{-1}(L)\,\theta(L)\,\epsilon_t$$

とすることができよう．定数項を定常過程に加えることによって確率過程は平均を持つことになるが，その挙動は変わらない．これに対して和分過程では平均の影響は劇的なまでに大きい．(5.4)に定数項を加えると，

(5.11) $$\Delta y_t = \beta + \epsilon_t$$

となるが，このモデルはトレンド付酔歩(random walk with drift)として知られている．$y_0$ を固定し $y_t$ のラグ値を繰り返し代入して期待値をとると，

(5.12) $$E(y_t) = y_0 + \beta t$$

となる．すなわち，$y_t$ の水準は線型トレンドに支配されている．トレンドは(5.1)と(5.2)のモデルにおいても導入することができるが，(5.11)の特徴はそれが線型トレンド項と定常撹乱項の和に分割されえないところにあろう．

## 予　　測

定常モデルにおける予測の基本的性質は，予測期間が増加するにつれて予測値が時系列の平均値に近づくことであった．もし $l$ が十分大きければモデルの構造は予測上それほど問題にならなくなる．ところが和分過程ではこれとは異なり，予測関数が階差の次数及び階差データにあてはめる ARIMA モデルの依存する非確率的要素を含むことになる．

ARIMA モデルにおける予測量の構成法は ARMA モデルと全く同じである．まず(5.9)の $\phi(L)\Delta^d$ を展開し $p+d$ の AR 多項式

(5.13) $$\phi(L)\Delta^d = \varphi(L) = 1 - \varphi_1 L - \cdots - \varphi_{p+d} L^{p+d}$$

を求める．次に予測は差分方程式

(5.14) $$\tilde{y}_{T+l/T} = \varphi_1 \tilde{y}_{T+l-1/T} + \cdots + \varphi_{p+d} \tilde{y}_{T+l-p-d/T} + \tilde{\epsilon}_{T+l/T} + \cdots + \theta_q \tilde{\epsilon}_{T+l-q/T}, \qquad l = 1, 2, \cdots$$

から求めることができる．ここで $\tilde{y}_{T+j/T}$ と $\tilde{\epsilon}_{T+j}$ は(4.2)で定義されたものである．

**例 1**　ARIMA(0, 1, 1)モデル

(5.15) $$\Delta y_t = \epsilon_t + \theta \epsilon_{t-1},$$

において予測量を差分方程式

(5.16) $$\tilde{y}_{T+l/T} = \tilde{y}_{T+l-1/T} + \tilde{\epsilon}_{T+l/T} + \theta \tilde{\epsilon}_{T+l-1/T}, \qquad l = 1, 2, \cdots$$

によって構成する．将来の $\epsilon_t$ の期待値はゼロであるから

(5.17 a) $$\tilde{y}_{T+1/T} = y_T + \theta \epsilon_T$$

(5.17 b) $$\tilde{y}_{T+l/T} = \tilde{y}_{T+l-1/T}, \qquad l = 2, 3, \cdots$$

となる．したがって時刻 $T$ においてすべての予測期間の予測値は水準線となる．撹乱項 $\epsilon_t$ は $\epsilon_1 = 0$ として差分方程式

(5.18) $$\epsilon_t = y_t - y_{t-1} - \theta \epsilon_{t-1}, \qquad t = 2, \cdots, T$$



より作られる(5章(3.3)で与えられる制約付最小二乗法による漸化式を参照)．$\epsilon_T$ に関する方程式で $\epsilon_t$ のラグ付き変数を繰り返し代入すると

(5.19) $$\epsilon_T = y_T - (1+\theta) \sum_{j=1}^{T-1} (-\theta)^{j-1} y_{T-j}$$

となる．したがって，一期先予測量(5.17 a)は

(5.20) $$\tilde{y}_{T+1/T} = (1+\theta) \sum_{j=0}^{T-1} (-\theta)^j y_{T-j}$$

と表わすことができる．すなわち $\tilde{y}_{T+1/T}$ は現在と過去の観測値の指数型加重移動平均(exponentially weighted moving average : EWMA)となっている．さらにより簡単に予測を行なう方法は(5.20)を $\lambda = 1+\theta$ として書き直し，

(5.21) $$\tilde{y}_{T+1/T} = \lambda y_T + (1-\lambda) \tilde{y}_{T/T-1}$$

とすると，現在時点における予測量は現在の観測値と一期前の予測との加重和となっていることがわかる．

この指数型加重移動平均(EWMA)予測はその簡便さの故に，多量の時系列の予測や複雑なモデルを作成する余裕がないときには非常に魅力的な方法となる．とくに在庫問題に関連した OR 分野においてはしばしば用いられる．平滑化定数(smoothing constant)$\lambda$ は 0 と 1 の間の値を与えられるが，0 に近ければ予測は過去の観測値に強く依存し，$\lambda = 1$ ならば過去のデータの情報は完全に無視され，一期先予測は現在の観測値に一致する．

**例 2**　ARIMA(0, 2, 2)モデル

(5.22) $$\Delta^2 y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$$

における随伴予測関数(eventual forecasting function)は $(1-L)^2 \tilde{y}_{T+l/T} = 0$ の解となる．この解は二期間の予測値 $\tilde{y}_{T+1/T}$ と $\tilde{y}_{T+2/T}$ を通る直線となる．このとき **1.1** 節で述べたように更新は容易にでき，1章(1.8)の更新方程式は(5.21)の一般化になっている．

**例 3**　ARIMA(1, 1, 0)モデルにおいて

$$\varphi(L) = (1-\phi L)(1-L) = 1 - (1+\phi)L + \phi L^2$$

とすると，予測量は差分方程式

(5.23) $$\tilde{y}_{T+l/T} = (1+\phi)\tilde{y}_{T+l-1/T} - \phi \tilde{y}_{T+l-2/T}, \qquad l = 1, 2, \cdots$$

より構成される．それに対応した予測関数は

(5.24) $$\tilde{y}_{T+l/T} = y_T + (y_T - y_{T-1}) \frac{\phi(1-\phi^l)}{1-\phi}, \qquad l = 1, 2, \cdots$$

となる．ここで $l \to \infty$ となるにつれて予測関数は水平線，

(5.25) $$y_T + (y_T - y_{T-1})\phi/(1-\phi)$$

に近づく．

ARIMA (0, 1, 1) モデルにおいては予測線は水平線になりその水準はすべての過去の値に依存している．(5.20)参照．他方，(5.24)は観測値の中でも最後の二点のみに依存している．すなわち外挿トレンドはきわめて局所的となる．このことは MA 要素を含まないすべての ARIMA モデルの特色といってよい．

以上の例から導かれる一般的命題は任意の ARIMA $(p, d, q)$ モデルに対する随伴予測関数は $d-1$ 次多項式となることである．モデルがゼロでない平均を持つときには随伴予測関数は $d$ 次の多項式となる．

**例 4**　モデル(5.11)において $\alpha = y_T$ とすると予測関数は，

$$\tilde{y}_{T+l/T} = \alpha + \beta l, \qquad l = 1, 2, \cdots \tag{5.26}$$

となる．例 2 において指摘したように定数項のない ARIMA (0, 2, 2) 過程に対する随伴予測関数はやはり線型トレンドとなる．上式(5.26)の特長としては傾きの母数 $\beta$ が全観測期間における $y_t$ の平均増加幅と解釈されることであろう．これに対し ARIMA (0, 2, 2) モデルでは予測関数の係数を決定する際に後半の観測値がより重視されることになっている．

さて和分過程に定数項を加えると時系列に非確率的要素を導入することになる．Box=Jenkins (1976, pp. 92-93) は多くの場合この方法が非現実的であることを主張している．すなわち $\beta = 0$ の仮定を「データが示す事実に反しない限り」置くことを好ましいとしている．

任意の ARIMA モデルにおける予測量の平均二乗誤差 (MSE) は，

$$\varphi(L)\,\psi(L) = \theta(L) \tag{5.27}$$

より求めることができる．(5.27)より

$$\psi_j = \sum_{i=1}^{\min(j,\,p+d)} \varphi_i\,\psi_{j-i} + \theta_j, \qquad 1 \leqslant j \leqslant q \tag{5.28 a}$$

$$\psi_j = \sum_{i=1}^{\min(j,\,p+d)} \varphi_i\,\psi_{j-i}, \qquad j > q \tag{5.28 b}$$

となる．ここで $\psi_0 = 1$ に注意しよう．

**例 5**　ARIMA (0, 2, 1) モデル

$$\Delta^2 y_t = \epsilon_t + 0.5\,\epsilon_{t-1} \tag{5.29}$$

において $\psi_j$ を計算すると，

$$\psi_1 = \varphi_1 + \theta_1 = 2.0 + 0.5 = 2.5$$



$$\psi_2 = \varphi_1\,\psi_1 + \phi_2 = 2(2.5) - 1.0 = 4.0$$

$$\psi_3 = \varphi_1\,\psi_2 + \varphi_2\,\psi_1 = 8.0 - 2.5 = 5.5$$

となる．したがって $l = 3$ に対しては

$$\mathrm{MSE}\,(\tilde{y}_{T+3/T}) = \sigma^2\,(1.0^2 + 2.5^2 + 4.0^2) = 23.25\,\sigma^2$$

であって $y_{T+3}$ に対する 95%予測区間は

$$\tilde{y}_{T+3/T} \pm 9.5\,\sigma$$

となる．

ARIMA モデルにおける予測量の平均二乗誤差は予測期間が長くなるにつれて急速に増加する．したがって ARIMA モデルの主な価値は短期予測にあることになる．

### 確率的トレンド項と誤差項の加法モデル

ARIMA モデルの構造をさらにとらえるために，モデル(5.1)における非確率的トレンドを確率的トレンド $m_t$ で置き換えてみよう．ここでトレンド項として意味をもたせるためには $m_t$ をある非定常過程とする必要があろう．

**例 6**　$m_t$ を酔歩過程

$$m_t = m_{t-1} + \eta_t \tag{5.30}$$

に従うとしよう．ここで $\eta_t \sim \mathrm{WN}(0, \sigma_\eta^2)$ であり，誤差項 $\epsilon_t$ は分散 $\sigma^2$ ホワイト・ノイズ過程とする．そこでこのモデルは

$$y_t = m_t + \epsilon_t, \qquad t = 1, \cdots, T \tag{5.31}$$

となり一次の階差をとると，

$$\Delta y_t = \eta_t + \epsilon_t - \epsilon_{t-1}, \qquad t = 2, \cdots, T \tag{5.32}$$

となる．これは ARIMA (0, 1, 1) 過程

$$\Delta y_t = \xi_t + \theta\,\xi_{t-1}, \qquad t = 2, \cdots, T \tag{5.33}$$

であって，$\xi_t$ はホワイト・ノイズとなる．**2.5** 節の例 2 参照．さらに(5.32)と(5.33)のモデルの一次の自己相関が同一であることより $\kappa$ を分散比 $\sigma_\eta^2/\sigma^2$ とすると，$\kappa = -(1+\theta^2)/\theta$ となる．これが 2 次関数であることから反転可能である根をとると，

$$\theta = [(\kappa^2 + 4\kappa)^{1/2} - 2 - \kappa]/2 \tag{5.34}$$

となる．すでに例 1 において ARIMA (0, 1, 1) モデルにおける最適予測は過去の観測値の指数型加重和移動平均 (EWMA) であることが示されている．この例のモデルは平均が徐々に変化してゆく時系列データを予測する方法としてこの EWMA 法が適当である一つの根拠を与えている．

**例 7** (5.31)の確率的トレンドの $m_t$ をトレンドをもつ酔歩過程へ一般化し，さらに酔歩トレンド項自体も確率変数にすると，

(5.35 a) $$m_t = m_{t-1} + \beta_t + \eta_t$$

(5.35 b) $$\beta_t = \beta_{t-1} + \zeta_t$$

であって $\zeta_t \sim \mathrm{WN}(0, \sigma_\zeta^2)$ と表現される．2回の階差をとれば，

(5.36) $$\Delta^2 y_t = (\eta_t + \zeta_t + \epsilon_t) - (\eta_{t-1} + 2\epsilon_{t-1}) + \epsilon_{t-2}.$$

となるが，これは ARIMA (0, 2, 2) モデルと同一であって，その下での予測量は1章(1.8)の漸化式で構成される．ここで平滑化の定数は誤差項同士の分散比に依存することになる．したがって(5.31)と(5.35)で与えられるモデルはホルト=ウィンターズの予測法への一つの合理づけを与えていると解釈することができよう．

## 6.6 季 節 性

観測値が月次あるいは四半期別に得られるときには季節要素について注意が払われねばならない．このことを回避するアプローチとして季節調整済みデータを用いる方法があるが，本節の最後で述べる理由によって一般的には望ましいこととはいえない．そこで本節では季節性を時系列モデルに組み込む方法について考えることにしよう．

季節性には大きく分けて二つの側面があろう．第一に毎年多かれ少なかれ規則的に繰り返されて観測値に現われる季節パターンがある．この種の季節要素は永続的ともみなすことができ，例えば天候のような要因から生じる要素といえよう．この種の季節要素をモデル化するにはトレンドをモデル化する方法と同様に考えればよいであろう．もし季節パターンが毎年一定と思われれば非確率的季節要素が適当であるし，季節パターンが徐々に変化しているようであれば，確率的季節要素を導入すればよいであろう．これら両方のモデルはともに規則的季節パターンを将来とも継続させるが，確率的トレンドの方がより最近の観測値に重点を置くこととなる．このことは非季節データ処理での局所的トレンドの際の議論と同一である．

以上で述べた季節性に欠けている要素としてもう一つの季節性の側面が考えられる．例えば，ある年3月の港湾ストライキはもし再びストライキが起こることを高い確率で予測するならば，次年3月の生産ターゲットに影響を及ぼすかもしれない．しかしながら3月にストライキが規則的に発生しない限り前年のストライキの影響は徐々に消滅するであろう．すなわち，異なる年の同じ季節の観測値間には相関があるが，年が経過するにつれてその相関が小さくなることが期待されよう．こう考えると季節パターンが予測期間を長くとるにつれ消滅する定常モデルが示唆されることとなろう．以下では，このようなモデルを論じ，次に非確率的季節要素と徐々に変化する季節要素を含むモデルを説明することにしたい．



### 季節自己回帰移動平均過程

四半期データの定常時系列を例にとってみよう．季節要素をとらえる一つの非常に簡単なモデルとして四次の季節自己回帰過程

(6.1) $$y_t = \phi_4 y_{t-4} + \epsilon_t, \qquad |\phi_4| < 1$$

が考えられよう．このモデルは AR(4) 過程において，$\phi_1 = \phi_2 = \phi_3 = 0$ の制約条件を持つ特殊なモデルである．(6.1)は定常 AR(1) 過程とその性質が類似しているので，**2.2** の方法を用いることによって自己相関関数が

(6.2) $$\rho(\tau) = \begin{cases} \phi_4^{\tau/4}, & \tau = 0, 4, 8, \cdots \\ 0, & \tau \neq 0, 4, 8 \end{cases}$$

となることがわかる．$|\phi_4|$ が1に近いほど季節パターンが顕著となる．ただ $\phi_4$ が単位円周上内にある限り季節要素は確率的であって，予測期間が長くなればなる程季節パターンも徐々に消滅してゆく．

次にモデル(6.1)を季節要素とともに AR 項と MA 項の双方を持つように拡張することができよう．もし $s$ を1年間における季節数とすると，

(6.3) $$\phi^\dagger(L^s) y_t = \theta^\dagger(L^s) \zeta_t,$$

ただし

(6.4 a) $$\phi^\dagger(L^s) = 1 - \phi_1^\dagger L^s - \cdots - \phi_P^\dagger L^{Ps}$$

(6.4 b) $$\theta^\dagger(L^s) = 1 + \theta_1^\dagger L^s + \cdots + \theta_Q^\dagger L^{Qs}$$

となるモデルを一般的に定式化とすることができよう．ここで $\zeta_t$ は独立な攪乱項列とする．経済における応用では $s$ の値は通常4か12であり，それぞれ四半期と月次の観測値に対応している．

表現(6.3)は $(P, Q)_s$ 次の純粋な季節 ARMA 過程を定義している．この場合には(6.1)と同様に自己相関関数は非季節時差ではゼロとなる．ところで季節的変動のみが時系列の予測可能な要素であると考えられない場合を除いてこのモデルは適切でない．例えば月次データでは本年3月の観測値は前年3月の観測値とともに本年2月の観測値とも関係していると想定する方が自然であろう．

季節的変動と非季節的変動の両者を含むモデルを構成するには，基本的に二つの方法が考えられる．第一の方法は単に季節モデルにおいて生じたすき間を埋めるやり方である．すなわち，例えば(6.1)に一次の項を導入すると

(6.5) $$y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \epsilon_t$$

となる．

第二の方法としてはモデル(6.3)の独立な攪乱項を非季節的 ARMA $(p, q)$ 過程で置き換えて，

(6.6) $$\phi(L)\, u_t = \theta(L)\, \epsilon_t$$

とすることが考えられよう．ここで $\phi(L)$ と $\theta(L)$ は 2 章(1.26)で定義された随伴多項式である．ここで(6.3)と(6.6)を組み合わせると，

(6.7) $$\phi^{\dagger}(L^s)\, \phi(L)\, y_t = \theta^{\dagger}(L^s)\, \theta(L)\, \epsilon_t$$

となるが，これを次数 $(p, q) \times (P, Q)_s$ の乗法的(multiplicative)季節 ARMA 過程と呼ぶ．簡単な例として AR(1)過程を(6.1)に結びつけるとモデルは，

(6.8) $$(1-\phi_1 L)(1-\phi_4 L^4) = \epsilon_t$$

となる．これを書き直すと

(6.9) $$y_t = \phi_1 y_{t-1} + \phi_4 y_{t-4} + \phi_5 y_{t-5} + \epsilon_t$$

であって $\phi_5 = -\phi_1 \phi_4$ となる．

このような乗法型モデルは時系列分析においては広く用いられている．しかしながら，これまでの議論では(6.9)が(6.5)よりも好まれる理由は明らかではない．後に階差を取ったときに乗法モデルが最も自然に生み出されることについて述べることにする．

以上で述べた二つの方法は定常季節モデルの定式化の最も一般的なものであるが，三番目の方法として次のモデルも考えられる．すなわち加法(additive)モデル

(6.10) $$y_t = s_t + u_t$$

において，$u_t$ を(6.6)の ARMA $(p, q)$ 過程，$s_t$ を次数 $(P, Q)_s$ の季節 ARMA 過程とする方法である．ここで二つの確率過程の攪乱項は互いに独立と仮定される．例えば

(6.11) $$y_t = \frac{\zeta_t}{1-\phi_4 L^4} + \frac{\epsilon_t}{1-\phi_1 L}$$

を考え，変形すると

(6.12) $$(1-\phi_1 L)(1-\phi_4 L^4)\, y_t = \zeta_t + \epsilon_t - \phi_1 \zeta_{t-1} - \phi_4 \epsilon_{t-4}$$



となる．(6.12)の右辺は MA(4)過程であるがその母数には $\phi_1$ と $\phi_4$ からの制約条件がある．モデルを推定する際にこのような制約条件を考慮することは(6.11)を状態空間形に表現し，カルマン・フィルターを用いて尤度関数を計算することによって可能である．(4 章 **4.3** 節参照)．他方，Nerlove *et al.* (1979)はこの問題への別のアプローチについて述べている．

### 非確率的季節要素モデル

さて $u_t$ を平均ゼロの定常確率過程で，これまでに述べたような季節要素を含み得るとしよう．このモデルへ $s$ 個のダミー変数を導入することによって非確率的季節要素を埋め込むことができよう．ここでもし $z_{jt}$ を季節 $j$ に 1 をとり，さもなければゼロをとるダミー変数とすると，モデルは

(6.13) $$y_t = \sum_{j=1}^{s} \gamma_j z_{jt} + u_t, \qquad t = 1, \cdots, T$$

となる．ここで $j$ 番目の季節母数 $\gamma_j$ は $j$ 番目の季節における $y_t$ の期待値と解釈することができる．

非確率的季節要素のモデル化のもう一つの方法として三角関数を用いることも考えられる．このときには方程式(6.13)の代りに，

(6.14) $$y_t = \alpha_0 + \sum_j (\alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t) + u_t$$

を考える．ここで和記号は $s$ が偶数なら $j = 1$ から $s/2$ まで，$s$ が奇数ならば $j = 1$ から $(s-1)/2$ をとり，さらに $\lambda_j = 2\pi j/s$ とする．

方程式(6.14)は $s$ 個の季節母数を含んでいる．したがって(6.13)における $s$ 個のダミー変数と全く同じように季節要素を表わすことになる．三角関数表現を用いる利点としては，しばしば $s$ 個以下の係数で季節要素を表現できることにある．このことはとくに月次データについては重要な意味を持っている．**3.2** 節で引用したバター市場の例で考えると，12 ヵ月と 6 ヵ月に対応する二つの周波数要素が季節変動の 90%以上を占めていた．そこで季節モデルとしてはこの二つの周波数のみを考慮すれば十分であり，季節母数数は 12 個から 5 個へと減少するになる．

### 可変的季節パターン

モデル(6.13)を書き直すと

(6.15) $$y_t = s_t + u_t, \qquad t = 1, \cdots, T$$

となる．毎年 $j$ 番目の季節中では非確率的要素 $s_t$ は $\gamma_j$ に等しく，すべての $t$ について

(6.16) $$s_t = s_{t-s}$$

となっている．このモデルの季節性へ一定の柔軟性を認めて，(6.16)へ $u_t$ とは独立な攪乱項 $\omega_t$ を加えることが考えられよう．すなわち

(6.17) $$s_t = s_{t-s} + \omega_t$$

とする．このことは $\gamma_j$ を確率変数とすることに対応している．

ここで $s_t$ を(6.17)で与えられる酔歩過程に従っているとすると，季節パターンは時間とともに徐々に変化する．さらに $s_t$ は非定常的であっても季節階差(seasonal differencing)によって定常過程を得ることができる．すなわち，

(6.18) $$\Delta_s y_t = \omega_t + u_t - u_{t-s}$$

を得る．ここで季節階差作用素は

(6.19) $$\Delta_s = 1 - L^s$$

である．$\omega_t$ の分散が厳密に正であるとすると(6.18)の右辺は反転可能となる．

**例 1** $u_t$ と $\omega_t$ をホワイト・ノイズ過程とすると **6.5** 節の例 6 と同様の議論によって $\Delta_s y_t$ が季節 MA(1)過程となる．月次データについては，

(6.20) $$\Delta_{12} y_t = \xi_t + \theta_{12}\xi_{t-12}$$

となる．$\xi_t$ はホワイト・ノイズである．

ところで(6.17)によってモデル化された季節要素は確率的であるにもかかわらず将来へはある固定的季節パターンが予測される．というのは $l \geqslant 13$ に対して(6.20)の予測関数は，

(6.21) $$\tilde{y}_{T+l/T} = \tilde{y}_{T+l-12/T}$$

となるので，最適予測値 $\tilde{y}_{T+1/T}$ から $\tilde{y}_{T+12/T}$ までは後半の観測値に大きく依存しているからである．極端な場合にはすべての $u_t \equiv 0$，$\theta_{12} = 0$ となり得るので，最後の 12 個の観測値の変動パターンが将来再現されることになる．

もし $\omega_t$ が正の系列相関を示していれば多少滑らかな季節パターンが得られるであろう．もちろん非季節的 ARMA 過程の場合にも同様のパターンを得ることはできる．

**例 2** 恒常的季節要素と一時的季節要素を組み合わせることが可能である．ここで観測値は四半期ごとに得られ，$y_t$ が恒常的要素(6.17)と(6.1)の定常季節 AR(4)過程の和として得られているとしよう．このとき

(6.22) $$(1-\phi_4 L^4)(1-L^4) y_t = \omega_t - \phi_4 \omega_{t-4} + \epsilon_t - \epsilon_{t-4}$$



となって，$\Delta_4 y_t$ は次数 $(1,1)_4$ の季節 ARMA 過程に従うことになる．

**季節調整**

官庁統計はしばしば季節調整済み系列として公表される．そこで季節調整済みデータを用いれば，季節要素を明示的に考慮する必要はないとの考え方が流布している．しかしながら **3.5** 節で示されたように，官庁統計が採用している方法は季節調整を過剰に行ない，季節周波数におけるスペクトラムを過剰に取り除く傾向がある．このことは時間領域に反映され季調済み系列は季節ラグにおいて負の相関を示す傾向がある．1970 年代初期に英国の官庁で用いられていた季節調整は米国センサス局で作られた X-11 法の一つの修正法であった．Wallis (1974) はホワイト・ノイズを含めた多くの系列を用いてこの方法の四半期データへの影響を調べている．季節調整を行なったあるホワイト・ノイズ系列の自己相関関数が図 6.2 に示されている．理論的には季調済み系列もホワイト・ノイズ過程となることが望ましいが，ラグが 1 から 3 及び 5 から 7, …に正の相関が多少見受けられ，ラグ 4, 8, …に負の自己相関が見受けられるので理想状態からはズレている．実際，この季調済みデータに対し AR(4) モデルをあてはめることが考えられるかもしれないが，それはモデルが複雑な移動平均に対する適当な近似となることによる．さらに彼の研究では(6.1)の形の季節 AR 過程に対する季節調整の影響を考えている．それによると季節ラグでは自己相関を減少させるが，他のラグでは自己相関を導くことになる．((6.2)参照)．そして結論的には季調済み系列から AR(1) モデルが採用されることもあると述べている．

Wallis (1974) やその他の研究からの結論として一般に季節調整をしない系列を扱う方が望ましいといえるであろう．というのは，季節調整は時系列にかなりの歪みを導入するとともに，季調済み系列から季節要素が除れている保障はどこにもないからである．

## 6.7　時系列モデルの一般的クラス

これまで二節にわたって階差に基づいたモデル作成の手続への一般的アプローチを説明してきた．実際，Box=Jenkins (1976) が採用したのは以上の方法であった．通常の階差と季節階差を時系列が定常になるまで適用し，得られた

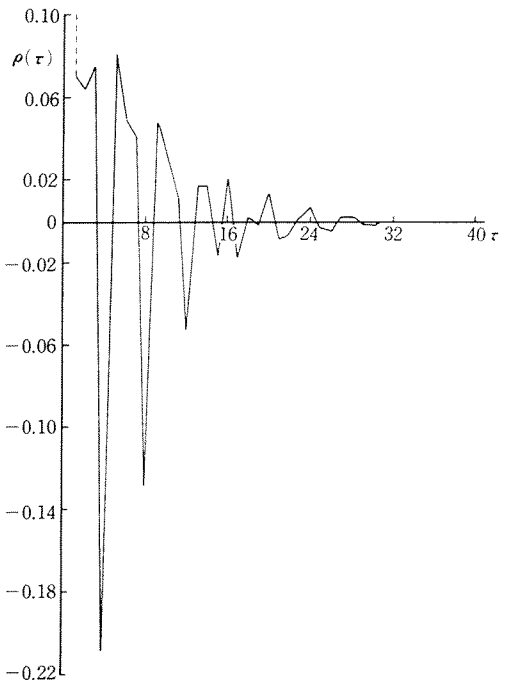



**図 6.2** 季調済みホワイト・ノイズ系列の自己相関関数.
出典：Wallis (1974).

系列に (6.7) で与えられる乗法的季節 ARMA 過程を用いてモデル分析を行なうのが彼らのアプローチである．このモデルは $D$ と $d$ をそれぞれ季節階差と 1 次階差を適用する回数とすると，

$$\phi^{\dagger}(L^s)\,\phi(L)\,\Delta^d\,\Delta_s^D y_t = \theta^{\dagger}(L^s)\,\theta(L)\,\epsilon_t \tag{7.1}$$

と表現され，次数 $(p, d, q)\times(P, D, Q)_s$ の乗法的季節 (multiplicative seasonal) ARIMA 過程として知られている．ここでモデルの定常性部分を乗法過程とする仮定をゆるめれば，さらに一般的モデルの定式化が得られ，

$$\phi(L)\,\Delta^d\,\Delta_s^D y_t = \theta(L)\,\epsilon_t \tag{7.2}$$

となろう．ここで多項式 $\phi(L)$ と $\theta(L)$ は通常は季節母数と非季節母数の両方を含み，さらに (6.5) のようなゼロ係数も可能とする．乗法モデルは (7.2) において母数が (7.1) を満たす制約条件を課されている意味で (7.2) の特殊な場合となっている．



これら二つのモデルのいずれにおいてもその推定や予測に関しては新たな問題は生じない．むしろ困難な問題はモデルの選択に関わることであろう．ここでただ一つモデルの推定について付け加えると，Ansley and Newbold (1980) の結果では，非季節モデルよりも季節モデルの方が厳密な最尤法の有効性が一層顕著になることが示唆されている．

**例 1**　Box=Jenkins (1976, p. 531) で分析されている航空乗客量データは時系列分析の古典的例となっている．データは 1949 年 1 月から 1960 年 12 月間の毎月の乗客者数である．図 6.3 ではその数字 (千単位) の対数値とともにモデル

$$\Delta\Delta_{12} y_t = (1+\theta_1 L)\,(1+\theta_{12} L^{12})\,\epsilon_t \tag{7.3}$$

を用いて 1957 年からの予測値がプロットされている．このモデルは $y_t$ の対数値に関する $(0, 1, 1)\times(0, 1, 1)_{12}$ の乗法季節 ARIMA 過程である．すべての観測値を用いた時の厳密な最尤推定値は，

$$\tilde{\theta}_1 = -0.40 \qquad \tilde{\theta}_{12} = -0.55$$
$$(0.09) \qquad\qquad (0.07)$$

となる．ここでカッコ内の数字は漸近標準誤差である．最初の 48 個の残差自己相関に基づく修正ボックス=ピアース統計量は $Q^*(46) = 44.12$ となって定式化が妥当なことを示唆している．

ところで (7.3) に対応した (7.2) のモデルではラグ 1, 12, 13 の MA 母数を持つ．すなわち (7.2) のモデルでは $\theta_{13}$ は非線型制約条件 $\theta_{13} = \theta_1\,\theta_{12}$ を満足せねばならない．ここで (7.3) の右辺を展開して推定値を用いて整理すると，

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \epsilon_t - 0.40\,\epsilon_{t-1} - 0.55\,\epsilon_{t-12} + 0.22\,\epsilon_{t-13} \tag{7.4}$$

となり，さらに予測値もいつものようにこの方程式から帰納的に計算される．随伴予測関数は線型トレンドとその上に規則的な季節パターンを組込む形になっている．

### トレンド要素，季節要素および誤差要素

季節 ARIMA モデルに対する一つの合理づけをトレンド要素，季節要素および誤差要素から成る加法モデルを考えることによって行なうことができる．加法モデルは

$$y_t = m_t + s_t + u_t, \qquad t = 1, \cdots, T \tag{7.5}$$

と表わされるが，これは (5.31) と (6.15) の一つの一般化になっている．もし $m_t$ がトレンド付酔歩過程

$$m_t = \beta + m_{t-1} + \eta_t \tag{7.6}$$

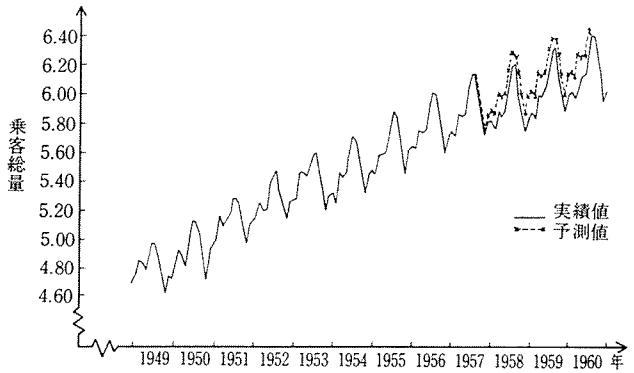



**図 6.3** 月次航空総乗客数(1000 人)の対数値
データの出典：Box=Jenkins(1976, p. 531).

に従い，さらに $s_t$ を(6.17)の確率的季節要素とすると一次階差をほどこし，次に季節階差をとると一つの定常過程

(7.7) $$\Delta\Delta_s y_t = (1-L^s)\eta_t + (1-L)\omega_t + (1-L)(1-L^s)u_t, \qquad t = s+1, \cdots, T$$

を得る．$u_t$ がホワイト・ノイズ撹乱項のときにはこれを $\epsilon_t$ で表わし，(7.7)を書き直すと

(7.8) $$\Delta\Delta_s y_t = (\eta_t + \omega_t + \epsilon_t) - (\omega_{t-1} + \epsilon_{t-1}) - (\eta_{t-s} + \epsilon_{t-s}) + \epsilon_{t-s-1}$$

となる．(7.8)の右辺は(6.7)の形の乗法的季節 ARMA モデルである．したがってこのモデル全体として次数 $(0,1,1)\times(0,1,1)_s$ の季節 ARIMA 過程

(7.9) $$\Delta\Delta_s y_t = (1+\theta_1 L)(1+\theta_s L^s)\xi_t$$

となる．ここで $\xi_t$ はホワイト・ノイズである．

ところで(7.8)と(7.9)はそれぞれの自己相関関数を調べることによって結びつけることができよう．季節 ARIMA モデルに関しては，

(7.10 a) $$\rho(1) = \theta_1/(1+\theta_1^2)$$

(7.10 b) $$\rho(s-1) = \theta_1\theta_s/(1+\theta_1^2)(1+\theta_s^2)$$

(7.10 c) $$\rho(s) = \theta_s/(1+\theta_s^2)$$

(7.10 d) $$\rho(s+1) = \rho(s-1)$$

となり，他の自己相関はゼロである．(7.5)より導かれるモデルでは分散は，

(7.11) $$\gamma(0) = 2\sigma_\eta^2 + 2\sigma_\omega^2 + 4\sigma^2$$

であり，ゼロでない自己共分散は



(7.12 a) $$\gamma(1) = -(\sigma_\omega^2 + 2\sigma^2)$$

(7.12 b) $$\gamma(s-1) = \sigma^2$$

(7.12 c) $$\gamma(s) = -(\sigma_\eta^2 + 2\sigma^2)$$

(7.12 d) $$\gamma(s+1) = \gamma(s-1)$$

となる．この自己共分散を $\gamma(0)$ で割り，(7.10)と比較すると，

(7.13 a) $$\tilde{\sigma}_\eta^2 = \sigma_\eta^2/\sigma^2 = -(1+\theta_1)^2/\theta_1$$

(7.13 b) $$\tilde{\sigma}_\omega^2 = \sigma_\omega^2/\sigma^2 = -(1+\theta_s)^2/\theta_s$$

が得られる．

**例 2** 例 1 の航空乗客量のデータに(7.9)のモデルをあてはめてみよう．(7.13)を用いるとトレンド要素と季節要素の相対的分散の推定値はそれぞれ 0.90 と 0.37 となる．したがって，季節パターンは非常にゆるやかに変化していることがわかる．

### 過剰階差問題

ある時系列の階差を $h$ 回とると $h$ 次の非確率的トレンドを取り除くことになる．しかしもし(5.1)の撹乱項が定常的であるときには，階差モデルは MA 要素の随伴多項式が単位円周上に $h$ 個の根を持つことになって撹乱項は反転可能ではなくなる．

**例 3** 線型トレンド・モデル

(7.14) $$y_t = \alpha + \beta t + \epsilon_t$$

を考えよう．2 次の階差をとると

(7.15) $$\Delta^2 y_t = (1-L)^2\epsilon_t = \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2}$$

となる．このとき $\Delta^2 y_t$ は平均ゼロの定常過程となるが反転可能ではない．同じことは **6.5** 節の例 7 においても成立する．もし $\sigma_\eta^2 = \sigma_\xi^2 = 0$ ならば確率的トレンドは非確率的トレンドとなり，(5.36)は $\theta_1 = -2$，$\theta_2 = 1$ の ARIMA $(0,2,2)$ 過程となる．

季節階差は多項式トレンド項を取り除く役割もはたしている．それに加えてもちろん季節階差は非確率的季節要素も取り除く役割もはたしていることはいうまでもない．

**例 4** 次のモデル

(7.16) $$y_t = \alpha + \beta t + \sum_{j=1}^{11}\gamma_j z_{jt} + \epsilon_t, \qquad t = 1, \cdots, T$$

は航空乗客者数データへあてはめる一つの可能な選択であろう．ここで $z_{jt}$ は(6.13)で定義された季節ダミー変数であり，定数項が既に含まれているので11個のダミー変数が必要である．1次の階差に続いて12次の階差をとれば

(7.17) $$\Delta\Delta_{12} y_t = (1-L)(1-L^{12})\epsilon_t, \qquad t = 13, \cdots, T$$

となる．これは乗法的季節 ARIMA 過程(7.3)において $\theta_1 = \theta_{12} = -1$ としたモデルである．(7.15)と同様この確率過程も反転可能ではない．モデル(7.3)における $\theta_1$ と $\theta_2$ の最尤推定値がともに単位円周内にあることより(7.16)の非確率的モデルがこの場合不適切であることがわかる．

階差をとった系列が定常的であっても，厳密には反転可能でないときに過剰階差(over-differencing)が起きたという．ある系列において過剰に階差がとられたことを見逃した場合には，推定法が不適切となったり，有限標本予測法によらない予測を行なうことによって問題が生じる．推定問題についてはすでに5章で述べた．予測に関していえば，有限標本予測が(7.15)や(7.17)のように反転可能でないモデルで行なわれる限り情報は何も失われない．実際，その場合には任意の期間についての予測は(7.14)及び(7.16)に最小二乗法を適用して得られる予測と同一になる．このことについて詳しくは Harvey (1981 b) を参照されたい．

**例5** 平均を $\mu$ とするホワイト・ノイズ過程

(7.18) $$y_t = \mu + \epsilon_t, \qquad t = 1, \cdots, T$$

を考えよう．このとき $y_{T+l/T}$ の最小平均二乗線型推定量は $\bar{y}$ であり，その平均二乗誤差は総ての $l$ に対して $(1+T^{-1})\sigma^2$ である．そこで1次の階差がとられたとしよう．そしてさらに ARIMA (0, 1, 1) モデルと識別されたとして $t = 2, \cdots, T$ について状態空間形に表現しよう(4章 **4.1** 節の例1参照)．カルマン・フィルターを用いると，

(7.19) $$a_T = \begin{bmatrix} \Delta y_T \\ \theta \tilde{\epsilon}_T \end{bmatrix}, \qquad P_T = \begin{bmatrix} 0 & 0 \\ 0 & f_{T+1} - 1 \end{bmatrix}$$

であって $f_t$ は

(7.20) $$f_t = 1 + \theta^{2(t-1)} / [1 + \theta^2 + \cdots + \theta^{2(t-2)}], \qquad t = 2, \cdots, T+1$$

と定義される．ところで予測は $\Delta y_t$ ではなく $y_t$ に対して行なわれるので，状態空間表現を修正する必要がある．一つの方法は $y_t$ を $\phi = 1$ の ARMA (1, 1) 過程とみなすことであろう．このときには時刻 $t = T$ における状態変数ベクトルは $P_t$ を(7.19)で与えられるとして $a_T^{\dagger} = (y_T, \theta\tilde{\epsilon}_T)'$ となる．そこで多段予測の公式4章(5.1)と(5.2)より，



(7.21 a) $$\tilde{y}_{T+l/T} = y_T + \theta\tilde{\epsilon}_T, \qquad l = 1, 2, \cdots$$

(7.21 b) $$P_{T+1/T} = \begin{bmatrix} f_{T+1} & \theta \\ \theta & \theta^2 \end{bmatrix},$$

$$P_{T+2/T} = \begin{bmatrix} f_{T+1} + 2\theta + \theta^2 + 1 & \theta \\ \theta & \theta^2 \end{bmatrix}$$

などとなる．もし $\theta = -1$ であれば全ての予測期間に対し $P_{T+l/T}$ の左上要素は単に $f_{T+1}$ となり，(7.20)と4章(5.4)より $l$ 期先予測の平均二乗誤差は，

(7.22) $$\mathrm{MSE}(\tilde{y}_{T+l/T}) = \sigma^2(1 + T^{-1}), \qquad l = 1, 2, 3, \cdots$$

となる．これは $\bar{y}$ の平均二乗誤差に一致する．実際(7.21 a)と $\bar{y}$ は同一であることがわかる．

### モデル選択

適当な ARMA モデルをコレログラムや偏自己相関関数より識別する方法について **6.3** 節で述べた．同じ方法は階差の次数さえ決定されれば季節 ARIMA モデルにも適用されよう．そこで追加的に生じる問題は乗法項の識別だけであろう．この問題への一つのアプローチとしては(7.2)の非乗法モデルを用い，次に制約条件を捜す方法が考えられよう．すなわち，もしモデルが

(7.23) $$\Delta\Delta_{12} y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_{12}\epsilon_{t-12} + \theta_{13}\epsilon_{t-13}$$

と識別されたとすると次のステップとして制約条件 $\theta_{13} = \theta_1\theta_{12}$ が適切か否かを決める．そしてもし適切ならば乗法モデル(7.3)を採用することが考えられよう．

適切な階差次数を決定するにはいくつかの方法が考えられる．ボックス=ジェンキンズが好むアプローチはコレログラムに基づく方法である．定常過程では時差が大きくなるにつれて自己相関がゼロに近づくので，それがコレログラムに反映される．他方，非定常過程からの標本自己相関は通常時差が大きくなってもゼロに減衰することはない．その直観的理由は定常 AR (1) モデルを考えることで理解されよう．すなわち2章(2.30)で与えられる自己相関関数は $|\phi|$ が1に近づくにつれて，よりゆるやかに減衰するようになる．$\phi = 1$ の場合には確率過程は酔歩となって理論的自己相関関数は定義されないが標本相関関数はなお計算することができる．

もしコレログラムが高次に比較的大きな自己相関を示していれば多分観測値の1次階差が適当であろう．コレログラムが定常性の性質を示すまで階差をと

るのがボックス=ジェンキンズの方法である．この方法では月次や四半期の観測値については1次階差と季節階差の色々な組み合わせを試みる必要が生じる．ところでコレログラムを用いて階差の次数を決める方法は大標本理論に基づくものである．もし$T$が小さければ状況は必ずしも明瞭ではなくなる．**6.3**節で指摘したように，定常過程におけるコレログラムは標本自己相関関数自体が互いに相関を持つために理論値ほどには急速にゼロへ収束することはない．他方，非定常過程のコレログラムも$T$が小さければ急速にゼロへ収束しうる．そこで階差の次数に何らかの疑問があるときには，主な競合モデルをあてはめAICに基づいてモデルを選択することも一考に価するであろう．非確率的トレンド要素と確率的トレンド要素$f(t)$，$s(t)$を(7.2)に代入すると，

$$\Delta^d \Delta_s^D y_t = f(t) + s(t) + \phi^{-1}(L)\,\theta(L)\,\epsilon_t \tag{7.24}$$

となる．(7.24)の右辺が$n$個の母数を含んでいれば，適合度基準は

$$\mathrm{AIC}^\dagger = \tilde{\sigma}^2 \exp\{[\log|V| + 2(n+d+sD)]/T\} \tag{7.25}$$

となる．ただし，ここで$\sigma^2 V$は$\phi^{-1}(L)\theta(L)\epsilon_t$の共分散行列であって，$T$は階差をとる前の観測値数である．分散の推定値$\tilde{\sigma}^2$の公式で分母として用いられるのはこの$T$である．

**例6**　モデル(7.16)を航空乗客数データへあてはめると$\tilde{\sigma}^2 = 0.00328$となる．したがってモデル選択基準は，

$$\mathrm{AIC}^\dagger = 0.00328 \exp(2\times 13/144) = 384\times 10^{-5}$$

となる．他方，乗法ARIMAモデル(7.3)で$\log|V|$をゼロとしておくとき，

$$\mathrm{AIC}^\dagger = 0.00135 \exp(2\times 15/144) = 166\times 10^{-5}$$

である．したがって，ARIMAモデルの方が好ましいのは明らかである．ここで(7.3)を(7.6)と比較すると事実上二つの余分な母数を含んでいる．このことは(7.17)が(7.3)の特殊な場合であり，有限標本予測法を用いると(7.16)と同一の予測値をもたらすことによって合理化されよう．ARIMAモデルの優越性はモデルの診断によっても確認される．実際$Q(8) = 317.6$であるから非確率要素モデル(7.16)を用いることは問題外であろう．

非季節モデルにおける実際の階差次数は1あるいは2であろう．もし$d = 3$とすると2次の予測関数が得られるのでそれに適する時系列は多くない．しかしながら，定常性を達成するために多くの階差が必要となることもあり得る．例えば，一つにはトレンド項と誤差項の加法モデルにおいてトレンドが



$\alpha$, $\beta$, $\gamma$を母数とするロジスティック関数

$$f(t) = \alpha/(1+\beta e^{-\gamma t}) \tag{7.26}$$

で与えられる場合である．この関数は$t\to\infty$となるにつれて上限あるいは飽和水準に向い，ある状況では非常に有用なモデルとなる．ところがこのときには2次の階差モデルでは$l = 1$, 2を除いて予測はあまりよくならないであろう．

観測値が月次あるいは四半期のとき，通常は適当な季節階差が用いられる．$D$を1以上にするのはきわめてまれであって，$D = 1$を$d = 0$あるいは1に結びつけることが多い．もし$d = D = 1$ならば随伴予測関数は線型トレンドに固定季節要素を被せたものとなる．

モデルの診断に際しては，ボックス=ピアース検定とラグランジュ乗数検定は前と同様に行なうことができる．ただし残される季節的影響を含めるためにはボックス=ピアース検定の際に相当長い数のラグを用いなければならなくなる．そこでボックス=ジェンキンズは航空乗客数データの分析では$Q$－統計量の計算に$P = 48$としている．ところで残念なことに，高次の$P$を用いると一つや二つの異常に大きな自己相関の影響は容易に無視されるので，コレログラムを視覚的に調べることも特に重要であろう．この状況では累積ピリオドグラム法は検定法として$p = q = 0$の場合を除いて厳密とはいえないが，より一層有効な診断法となることが考えられる．

## 6.8　多変量モデルの作成

一変量モデルの作成法の多くはそのまま多変量モデルに拡張されるが，多変量の場合にはそれに加えていくつか新しい問題が生じる．とくにモデルの識別はより一層複雑なものとなる．この節では多変量モデル作成法の要約というよりも，むしろ一つの応用例を通じてそこに存在する問題の諸々の側面に焦点をあてることにする．例としては，Chan=Wallis(1979)を用いることとするが，そのデータは1849年から1909年にかけてのカナダのハドソン湾会社によって年間に取引されたミンクとジャコウネズミの皮数である．このデータに多変量の分析が求められる理由は，二種類の動物に餌—捕食者の関係があることが知られているので，相互の関係が両者の人口動態に影響するからである．

### 定　常　性

事前的データ分析によれば，ジャコウネズミの観測値は一次の階差によって定常化されるが，ミンクの観測値は原系列が定常化を示している．Chan=Wallis によれば，二つの時系列に別々の階差操作を行なうと系列間の位相関係を歪めることになる．そこでミンクの系列の階差をとることをせずに，双方の時系列を二次関数に回帰してその残差を分析しているが，この操作がミンク系列に過剰階差をほどこすよりも好ましいかは意見が分れよう．もし過剰に階差をとるとすれば厳密最尤推定法と有限標本予測法を行なわなければならない．

### 識別，推定および診断

さて $y_{1t}$ と $y_{2t}$ をそれぞれトレンドを除いたジャコウネズミとミンクの系列とすると，Chan=Wallis は次の1変量モデルをあてはめによって得た．

(8.1)　(i)　$(1-1.03L+0.68L^2-0.39L^3+0.34L^4)\,y_{1t}=\eta_{1t}$,

$$\begin{matrix}&(0.12)&(0.17)&(0.17)&(0.13)\end{matrix}$$

$$\tilde{\sigma}_1^2=0.0789,\qquad Q(16)=19.12$$

(8.2)　(ii)　$(1-1.36L+0.67L^2)\,y_{2t}=(1-0.70L)\,\eta_{2t}$.

$$\begin{matrix}&(0.13)&(0.09)&&(0.15)\end{matrix}$$

$$\tilde{\sigma}_2^2=0.0605,\qquad Q(17)=15.86$$

この結果は多変量時系列分析についてのいくつかの示唆を与えている．$y_{1t}$ について選ばれたモデルは純粋な自己回帰モデルであるが，尤度比検定からは多変量 ARMA(2, 1) モデルが最も満足できる定式化であることがわかった．推定結果は次のようになる．

(8.3)
$$\begin{bmatrix}1-1.22L+0.61L^2 & 0\\ (0.16)\ (0.12) & \\ 0 & 1-1.29L+0.62L^2\\ & (0.15)\ (0.13)\end{bmatrix}\begin{bmatrix}y_{1t}\\ y_{2t}\end{bmatrix}$$

$$=\begin{bmatrix}1-0.15L & -0.83L\\ (0.22) & (0.16)\\ 0.37L & 1-0.81L\\ (0.14) & (0.14)\end{bmatrix}\begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{bmatrix},$$



ただし，

$$\tilde{\Omega}=\begin{bmatrix}0.061 & 0.022\\ 0.022 & 0.053\end{bmatrix},\qquad Q(30)=\begin{bmatrix}28.05 & 18.15\\ 15.73 & 22.45\end{bmatrix}$$

であった．二つの系列を同時に推定することによって残差分散は1変量モデルの残差分散よりも小さくなっている．さらに $|\tilde{\Omega}|=0.00275$ であるのに対し，二つの1変量モデルからは $|\tilde{\Omega}_0|=\tilde{\sigma}_1^2\times\tilde{\sigma}_2^2=0.0789\times0.0605=0.00477$ となる．多変量モデルでは AIC 基準は

(8.4)　$$\mathrm{AIC}^{\dagger}=|\Omega|\exp(2n/T)$$

である．この基準に照らしてみると，一変量モデルでは $n=9$ より $\mathrm{AIC}^{\dagger}=0.00643$ となるのに対し，多変量モデルでは $n=11$ より $\mathrm{AIC}^{\dagger}=0.00397$ となるので後者の方がまさっている．ただし Chan=Wallis の分析では AIC は考えていない．

ところで(8.3)における $Q$ 行列はボックス=ピアース統計量の一つの一般化であって非対角要素は残差の相互相関より作られているが，この行列はモデルのあてはまりのよさに関する大ざっぱな診断に用いることができよう．統計量 $Q$ の分布はまだよく分っていないが個々の自己相関関数が $P$ 次まで計算されているときには，適合度の目やすとして個々の統計量を $\chi^2_{P-p-q}$ 分布を用いて検定することが考えられよう．「多変量風呂敷統計量」(multivariate portmanteau statistics) に基づく検定方法は Hosking (1980) が提案している．

さて(8.3)における二つの AR 多項式は非常に似かよっているので，両者が等しいとした制約条件下でモデルを推定すると，

(8.5)
$$\begin{matrix}(1-1.28L+0.63L^2)\,y_t=\\ (0.13)\ (0.11)\end{matrix}\begin{bmatrix}1-0.27L & -0.79L\\ (0.16) & (0.14)\\ 0.34L & 1-0.75L\\ (0.11) & (0.12)\end{bmatrix}\begin{bmatrix}\epsilon_{1t}\\ \epsilon_{2t}\end{bmatrix},$$

および

$$\tilde{\Omega}=\begin{bmatrix}0.061 & 0.023\\ 0.023 & 0.054\end{bmatrix}$$

となる．そして二つの AR 作用素が同一であるとする帰無仮説は尤度比検定では棄却されない．

次に2章(7.23)と(7.25)の関係を用いると，本章(8.5)のベクトル ARMA 過程に対応するシステムの最終型は1次自己回帰モデルである．そこでそれをあ

てはめてみると，

$$\begin{bmatrix} 1-0.79L & 0.68L \\ (0.07) & (0.09) \\ -0.29L & 1-0.51L \\ (0.07) & (0.09) \end{bmatrix} y_t = \epsilon_t, \tag{8.6}$$

および

$$\tilde{\Omega} = \begin{bmatrix} 0.061 & 0.022 \\ 0.022 & 0.058 \end{bmatrix}, \qquad Q(30) = \begin{bmatrix} 29.11 & 21.02 \\ 18.41 & 26.38 \end{bmatrix}$$

となる．ここで $|\tilde{\Omega}| = 0.00305$ となりモデル(8.5)よりも大きくなるが母数の数は少なくなっている．というのは，多項式 $|\Phi(L)|$ と $\Phi(L)^\dagger$ はともに $\Phi(L)$ より導かれたが，(8.5)の推定に際しては係数の制約条件を考慮していないためである．したがって(8.6)の方が母数は二つ少なくなり AIC$^\dagger$ は小さくなっている．

ところで(8.6)に対応した最終型を導くことも有益であろう．そこで

$$\Phi(L) = \begin{bmatrix} 1-0.79L & 0.68L \\ -0.29L & 1-0.51L \end{bmatrix} \tag{8.7}$$

であることから，その行列式多項式は

$$\begin{aligned} |\Phi(L)| &= (1-0.79L)(1-0.51L) - (0.68L)(-0.29L) \\ &= 1-1.30L+0.60L^2 \end{aligned} \tag{8.8}$$

となる．この方程式は(8.5)にあてはめられた AR 部分の多項式に相当近くなっている．さらに余因子行列は，

$$\Phi^\dagger(L) = \begin{bmatrix} 1-0.51L & -0.68L \\ 0.29L & 1-0.79L \end{bmatrix} \tag{8.9}$$

となる．ここで2章(7.25)を用いると(8.6)では $\Theta(L) = I$ となることから(8.9)を(8.5)の MA 部分の行列と比較すると母数は非常に類似している．

また(8.6)に対する $Q$ 統計量も満足できるものであり，Chan=Wallis もこのモデルがその単純さにもかかわらず種々の過剰階差のテストに合格したことを報告している．彼らによればこのモデルは時系列の相互作用を考慮し，生物学者が注目する重要な現象をとらえているとのことである．係数行列の非対角要素はジャコウネズミの増加がミンクの次年次での増加を生み，ミンクの増加はジャコウネズミの次年次での減少をもたらすことを意味している．



### モデルの性質*

ところで(8.5)の AR 部分の多項式の根は共役複素数になるが，それは約9.93年の周期の減衰振動を意味する．(3章(4.11)参照)．同様の振動が(8.6)についても予想される．モデル(8.6)の確率的性質を調べる最良の方法はそのスペクトラム密度行列を計算することであろう．Chan=Wallis は二つの系列のスペクトラム密度では10年周期に対応する周波数でかなりのスペクトルへの寄与が観察され，位相図ではこの周期期間でジャコウネズミの時系列がミンクの時系列を2〜4年先行していることを報告している．

### 予　　測

ベクトル ARMA 過程における予測は1変量の場合とまったく同様にして行なうことができる．漸化式(4.2)はスカラー量についてであったが，それをベクトルと行列におき換えることによって直ちに多変量の場合へ一般化される．たとえば(8.6)のベクトル AR(1)モデルにおいては $l$ 期先予測値は，

$$\tilde{y}_{T+l/T} = \begin{bmatrix} 0.79 & -0.68 \\ 0.29 & 0.51 \end{bmatrix} \tilde{y}_{T+l-1/T}, \qquad l = 1, 2, \cdots$$

および $\tilde{y}_{T/T} = y_T$ とすることによって求まる．そしてこの式は(4.4)の一般化になっている．この確率過程は定常であるから $l \to \infty$ のとき $\Phi^l \to 0$ となり，予測値は振動しながら減衰の形をとる．ここでもし $y_{1T} = y_{2T} = 1$ としたならば，

$$\tilde{y}_{T+1/T} = \begin{bmatrix} 0.79-0.68 \\ 0.29+0.51 \end{bmatrix} = \begin{bmatrix} 0.11 \\ 0.80 \end{bmatrix},$$

$$\tilde{y}_{T+2/T} = \begin{bmatrix} 0.79\times 0.11 - 0.68\times 0.80 \\ 0.29\times 0.11 + 0.51\times 0.80 \end{bmatrix} = \begin{bmatrix} -0.46 \\ 0.44 \end{bmatrix}$$

などとなる．

ここで全く同じ予測値は(8.6)を解くことによって得た最終型から求めることもできる．そのモデルは ARMA(2, 1)過程であるので各時系列に対する予測系列は(8.8)に基づいた同一の差分方程式に支配される．すなわち，このことによって二つの予測系列の振動周期が同一であることが確認されることとなる．

さらに予測誤差の共分散行列は **6.4** 節の結果を一般化すれば求めることができる．代入を繰り返すことによって，

$$\tilde{y}_{T+l/T} = \boldsymbol{\Phi}^l y_T \sum_{j=0}^{l-1} \boldsymbol{\Phi}^j \boldsymbol{\epsilon}_{T+l-j}, \qquad l = 1, 2, \cdots \tag{8.10}$$

となる．したがって $\boldsymbol{\Phi}$ を条件とした時の $\tilde{y}_{T+l/T}$ の平均二乗誤差は

$$\mathrm{MSE}\,(\tilde{y}_{T+l/T}) = \sum_{j=0}^{l-1} \boldsymbol{\Phi}^j \boldsymbol{\Sigma} \boldsymbol{\Phi}'^j \tag{8.11}$$

である．この表現は $\boldsymbol{\Phi}$ を推定することによって $0(T^{-1})$ の項がつけ加わることが知られているが，これについては例えば Baillie (1979) を参照されたい．

## ノート

**6.3** ARMA モデルにおけるラグランジュ乗数検定は Godfrey (1979) によって導かれ，Poskitt=Tremayne (1980, 1981) によって展開されている．AIC による識別については Akaike (1974) や Hannan (1980) などが論じている．

**6.7** 乗法 ARIMA モデルと加法型要素モデルの関係は Harrison (1967)，Cleveland=Tiao (1976)，Engle (1978 a)，Nerlove *et al.* (1979) などが調べている．Roy (1977)，Hasza (1980)，Fuller (1976, p. 367) は非定常過程におけるコレログラムの性質を調べている．Harvey (1980) は AIC を使って原系列と一次階差系列による回帰モデルを区別する方法について述べている．ARIMA モデルは特別な場合として議論されている．

## 練習問題

**1.** 5 章問題 **7** の情報が与えられているときに，どのようにして仮説 $\rho(2) = 0$ を検定することが可能であろうか．

**2.** ARIMA (0, 2, 3) モデルにおいて 1 期先，2 期先，3 期先予測量の表現を求めよ．一般に $l > 1$ に対する予測関数は二次の差分方程式の解

$$\tilde{y}_{T+l/T} = \alpha + \beta l$$

となることを示せ．ただし $\alpha$，$\beta$ は $\tilde{y}_{T+2/T}$ と $\tilde{y}_{T+3/T}$ に依存するものとする．ここでもし $y_T = 10$，$y_{t-1} = 9, \theta_1 = -1.0, \theta_2 = 0.2, \theta_3 = -0.5, \epsilon_T = 0.5, \epsilon_{T-1} = 1.0, \epsilon_{T-2} = -0.2$ とすると，$l = 1, 2, 3$ に対して $\tilde{y}_{T+l/T}$ の値を決め，$\alpha$，$\beta$ の値を求めよ．

**3.** ジョーンズ氏はあるデータの組を分析するに際し，観測値が 6 次の定常自己回帰モデル

$$y_t = \pi_1 y_{t-1} + \cdots + \pi_6 y_{t-6} + \delta + \epsilon_t$$

から生成された（$\delta$ は定数）として適当な方法で未知母数の値を推定して，

$$\hat{\pi}_1 = 0.4,\ \hat{\pi}_2 = -0.36,\ \hat{\pi}_3 = 0.32,\ \hat{\pi}_4 = -0.29,\ \hat{\pi}_5 = 0.26,\ \hat{\pi}_6 = -0.23$$

を得た．これらの推定値から 6 個の母数は実は ARMA (1, 1) 過程の 2 個の母数によるものであることを説明できるであろうか．2 母数モデルに対し 6 母数モデルを採用していることに対し助言を与えるとすれば，どのようなことだろうか．

上の数値より ARMA (1, 1) モデルの母数を求め $\rho(1), \cdots, \rho(4)$ を計算しコレログラムをスケッチせよ．どのような理由によって AR (1) モデルの定式化を直ちに棄却できるであろうか．［ヨーク大学試験問題］



**4.** 5 章の練習問題 **3** で定義された MA (2) 過程について，$l = 1, 2, 3$ に対する予測値を計算せよ．もし $\theta_1$ と $\theta_2$ の値が真の値であるとして予測の平均二乗誤差を求めよ．

**5.** フォン・ノイマン比を用いて観測値 2.0, 1.2, −0.1, −0.5, 0.3, 0.4 が独立であるという仮説を，正の相関があるという対立仮説に対して検定せよ．

**6.** 平均ゼロの定常 AR (1) モデルを回帰分析法によって推定すると，

$$R^2 \cong \hat{\phi}^2$$

となることを示せ．

**7.** 以下に挙げる非確率トレンド要素を図に描き，(5.1) 型のモデルをどのように推定したらよいか考えよ．

(a) $f(t) = \exp(\alpha + \beta/t)$

(b) $f(t) = \exp(\alpha + \beta\gamma^t)$

**8.** AR (2) モデルにおいて $\phi_1 = 0.5$，$\phi_2 = -0.2$ でありしかも $y_{T-1} = 1, y_T = 2, \sigma^2 = 3$ とするとき，$l = 1, 2, 3$ に対し予測量とその平均二乗誤差を計算せよ．

**9.** (7.3) において $\epsilon_t$ を計算する漸化式を求めよ．そして最尤推定量 $\hat{\tilde{\theta}}_1$ と $\hat{\tilde{\theta}}_{12}$ の漸近共分散行列の表現を導け．さらにもし $|\theta_1|$ が 1 に近ければ $\tilde{\theta}_1$ と $\tilde{\theta}_{12}$ は近似的に無相関になり $\tilde{\theta}_1$ と $\tilde{\theta}_{12}$ の漸近分散はそれぞれ近似的に $(1-\theta_1^2)/T$ と $(1-\theta_{12}^2)/T$ となることを示せ．

**10.** 次にあげる定常過程における自己共分散関数を作れ．

(a) $y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_{12} \epsilon_{t-12}$

(b) $y_t = (1 - \phi L^4)^{-1}(1 + \theta L^4)\epsilon_t$

**11.** (7.10) と (7.12) を導け．

**12.** (7.3) において標本相関が $r(1) = -0.34$，$r(12) = -0.39$ とするとき，$\theta_1$ と $\theta_2$ の推定値を求めよ．

**13.** AR (1) モデルで $\phi = 0.5$ とするとき，一次の階差系列の自己相関関数を描け．

**14.** (6.5) において仮説 $\phi_4 = 0$ を検定するラグランジュ乗数統計量を求めよ．(6.8) に対して同様の検定を構成せよ．

**15.** Nelson (1972) は次のモデルを米国の GNP データにあてはめた．

$$\Delta y_t = 0.615 \Delta y_{t-1} + 2.76 + \epsilon_t$$

このとき予測関数と GNP の予測経路を決めよ．

**16.** $r(1)=0.7$, $r(2)=0.5$ として，始めの2つの標本偏相関を計算せよ．さらにそのモデル選択についての意味を述べよ．



# 7章
# 時系列回帰分析

## 7.1 逐次最小二乗法

1章(2.6)で与えられる線型回帰モデルにおける未知母数係数 $\beta$ のある推定値が最初の $t-1$ 個の観測値から計算されているとしよう．次に $t$ 番目の観測値が利用可能となるとき，1章(2.13)の直接最小二乗法の公式に従って積率行列の逆行列を求めることなしに推定値を計算することができる．それは次にあげる逐次更新公式によって実行される．ここで $b_t$ を最初の $t$ 個の観測値に基づく最小二乗推定量，$X_t=(x_1,\cdots,x_t)'$ としよう．このとき

(1.1) $$b_t=b_{t-1}+(X'_{t-1}X_{t-1})^{-1}x_t(y_t-x'_tb_{t-1})/f_t$$

と表現される．ただし

(1.2) $$(X'_tX_t)^{-1}=(X'_{t-1}X_{t-1})^{-1}-(X'_{t-1}X_{t-1})^{-1}x_tx'_t(X'_{t-1}X_{t-1})^{-1}/f_t$$

および

(1.3) $$f_t=1+x'_t(X'_{t-1}X_{t-1})^{-1}x_t,\qquad t=k+1,\cdots,T$$

としよう．この逐次更新公式は4章付論に与えた行列の逆転公式を用いて容易に導くことができる．またこの節の最後で述べる割引逐次最小二乗法のアルゴリズムの特殊な場合としても(1.1)～(1.3)を導くこともできる．ところで逐次最小二乗法はカルマン・フィルターの特殊な場合とみなすこともできる．1章(3.6)の古典的線型回帰モデルは4章(3.1)の観測方程式で $\alpha_t=\beta$，$z_t=x_t$，$\xi_t=\epsilon_t$，$h_t=1$ とみなし得る．遷移方程式は $\alpha_t=\alpha_{t-1}$ となるように $T=I$，$Q=0$ とおけばよい．したがって予測方程式は恒等式 $a_{t/t-1}=a_{t-1}$，

$P_{t|t-1}=P_{t-1}$ となり，最小二乗更新公式は4章(3.15)と(3.17)において $P_t=(X_t'X_t)^{-1}$ とすることによって得ることができる．

未知母数 $\beta$ の最小二乗推定値を計算するには最低限 $k$ 個の観測値が必要である．もし $X_k$ が正則ならば，最初の $k$ 個の観測値に基づく推定値は

(1.4) $$b_k=(X_k'X_k)^{-1}X_k'y_k^*=X_k^{-1}y_k^*$$

となる．ただしここで $y_k^*=(y_1,\cdots,y_k)'$ である．これを初期値とすれば推定値 $b_{k+1}$，…，$b_T$ の計算に行列の逆転を用いる必要はない．最後の推定値 $b_T$ は $T$ 個の観測値に基づく標準的最小二乗推定量と同等になる．

逐次公式を始めるもう一つの方法は，$t=0$ において $b_0=0$ として出発し，積率行列 $(X_0'X_0)^{-1}$ をある大きな数 $\kappa$ に対し $\kappa I$ と置くことが考えられよう．$b_0=0$ と置くことはアド・ホック(恣意的)であるが $\kappa$ を大きくとることによって推定値への影響は $k$ 個の観測値を得た後では無視されよう．$\beta$ を確率変数とすれば，これらの初期値は弱い形の事前分布の母数とみなすこともできるが，回帰分析の場合には漸化式をこのように始めることを単なる近似としてとらえることもできよう．

### 逐次残差

未知母数 $\beta$ の最小二乗推定量が逐次的に計算されると $T-k$ 個の予測誤差

(1.5) $$\tilde{v}_t=y_t-x_t'b_{t-1},\qquad t=k+1,\cdots,T$$

が得られる．この誤差は平均ゼロで分散は $\sigma^2 f_t$ となり，さらに互いに無相関である(Harvey(1981 a，2.6節)参照)．基準化された予測誤差

(1.6) $$v_t=\tilde{v}_t/f_t^{1/2},\qquad t=k+1,\cdots,T$$

は逐次残差(recursive residuals)として知られ，回帰モデルの定式化や安定性を診断する方法としてしばしば用いられる．その方法については Harvey(1981 a，5.2節)で論じられている．ここで注目すべきこととしては，もし回帰モデルの誤差項が正規分布に従えば逐次残差も正規分布に従う，すなわち **7.1** 節の(2.6)において $\epsilon_t\sim NID(0,\sigma^2)$ ならば $v_t\sim NID(0,\sigma^2)$ となることであろう．

逐次残差は残差平方和の更新公式においてその特色を表わし，

(1.7) $$SSE_t=SSE_{t-1}+v_t^2,\qquad t=k+1,\cdots,T$$

であって，しかも

(1.8) $$SSE_t=(y_t^*-X_tb_t)'(y_t^*-X_tb_t),\qquad t=k,\cdots,T$$

となる．さらに $SSE_k=0$ であるから，$T$ 個の観測値すべてからの残差平方和



は逐次残差の和に等しく，

(1.9) $$SSE_T=\sum_{t=k+1}^{T}v_t^2$$

となっている．

### 割引逐次最小二乗法

回帰モデルの未知母数 $\beta$ の割引最小二乗推定量(discounted least squares estimator：DLSE)とは，$\delta$ を $0<\delta\leq 1$ の範囲内にある定数として，

(1.10) $$S(\beta)=\sum_{t=1}^{T}\delta^{T-t}(y_t-x_t'\beta)^2$$

を最小化するベクトル $\beta=b^\dagger$ である．もし $\delta=1$ であれば $b^\dagger$ は最小二乗(OLS)推定量となる．1よりも小さい $\delta$ を用いると標本に指数ウェイトを乗ずることになる．標準回帰モデルの枠組では，もし $\sigma_t^2=\sigma^2\delta^{t-T}$ であれば，$\beta$ の最良線型不偏推定量(BLUE)になる．このことは観測値の時間を遡るにつれて撹乱項の分散が大きくなることを意味するわけであるが，割引最小二乗推定量の導出に当って現在から遠く離れた時点の観測値を，何らかの意味で信頼度が乏しいとしているのではない．むしろ割引最小二乗法を用いる根拠は，より実際的であって最小二乗法よりもよい予測を与えることがあるとの理由である．

割引最小二乗法の例は Gilchrist(1976)が与えている．そこではとくに6章(5.1)のモデルで多項式トレンドの推定問題への応用が強調され，大域的トレンドではなく局所的トレンドの推定に割引最小二乗法が用いられている(1章の練習問題2参照)．この方法は ARIMA 時系列モデルのフィットにも用いることができる．Leskinen=Terasvirta(1976)はこの方法を1954年から1974年にかけてのフィンランドのアルコール飲料消費データへ $(0,1,1)\times(0,1,1)_{12}$ の季節 ARIMA を適用する際に用いている．使用した割引率は12ヵ月について0.9であったが，そこから得られた係数推定値は直接最小二乗法によって得られた推定値が時間とともに変化していたことを示唆している．興味深いことには，割引最小二乗推定値は1963年から1974年の間のみの観測値から計算された最小二乗推定値にきわめて類似していた．

回帰モデルにおける割引最小二乗推定量は逐次計算が可能である．時刻 $t$ における $\beta$ の推定量は

(1.11) $$b_t^{\dagger}=\left[\sum_{j=1}^{t}\delta^{t-j}x_jx_j'\right]^{-1}\sum_{j=1}^{t}\delta^{t-j}x_jy_j,\qquad t=k,\cdots,T$$

で与えられる．ここで

(1.12) $$\sum_{j=1}^{t}\delta^{t-j}x_jx_j'=P_t^{-1}=\delta P_{t-1}^{-1}+x_tx_t'$$

に注意して行列の逆転公式(4章(A.3))を用いると，

(1.13) $$P_t=\delta^{-1}P_{t-1}-\frac{\delta^{-1}P_{t-1}x_tx_t'P_{t-1}}{\delta+x_t'P_{t-1}x_t},\qquad t=k+1,\cdots,T$$

となる．もし $\delta=1$ ならば(1.13)は(1.2)に退化する．$b_t^{\dagger}$ の更新公式は，

(1.14) $$\sum_{j=1}^{t}\delta^{t-j}x_jy_j=\delta\sum_{j=1}^{t-1}\delta^{t-1-j}x_jy_j+x_ty_t$$

となることに注意して式を変形すると，

(1.15) $$b_t^{\dagger}=b_{t-1}^{\dagger}+\frac{P_{t-1}x_t(y_t-x_t'b_{t-1}^{\dagger})}{\delta+x_t'P_{t-1}x_t}$$

$$=b_{t-1}^{\dagger}+P_tx_t(y_t-x_t'b_{t-1}^{\dagger}),\qquad t=k+1,\cdots,T$$

が得られる．$\delta=1$ ならば表現(1.1)を得る．

逐次推定値により母数がどのように変化するかその軌跡を調べることができる．割引率を用いると推定値がモデルの構造の変化に対しより敏感に反応することが利点として考えられるが，反面このことは推定値の安定性が多少犠牲になることを意味している．母数推定値の敏感性と安定性のトレード・オフが存在するために，通常 $\delta$ の値は1からあまり離れないようにとられる．

## 7.2 系列相関のある撹乱項

線型回帰モデルの撹乱項がARMA$(p,q)$過程に従う場合，

(2.1 a) $$y_t=x_t'\beta+u_t,\qquad t=1,\cdots,T,$$

(2.1 b) $$u_t=\phi_1u_{t-1}+\cdots+\phi_pu_{t-p}+\epsilon_t+\theta_1\epsilon_{t-1}+\cdots+\theta_q\epsilon_{t-q}$$

と表現される．ここで(2.1 a)は1章(2.6)と同一である．このようにモデルを定式化すると疑似最尤推定は条件付最小二乗のアルゴリズムを一般化することで実行できる．その際の技術的問題についてはHarvey(1981 a，6節)に述べられている．ここでのより困難な問題はある撹乱項に対して適当な時系列モデルを見つけることであるが，それについては前章で述べた方法を用いることがで



きよう．$\hat{\beta}$ を $\beta$ のある一致推定量とすると，$\hat{\beta}$ を用いたときの残差系列

(2.2) $$\hat{u}_t=y_t-x_t'\hat{\beta},\qquad t=1,\cdots,T$$

の分布は大標本では真の撹乱項の分布と同一になろう．したがって標本自己相関とモデル診断のための統計量についての漸近的結果はそのまま適用される．すなわち標準的なボックス=ジェンキンズ流のモデル識別法と診断法を最小二乗法による $\beta$ の初期推定値の下での残差系列にそのまま適用することができよう．

以上の議論は6章(5.1)で与えたトレンド項と誤差項の加法モデルを扱う際にも用いることができる．もし6章(5.2)のように $f(t)$ が多項式であれば，最小二乗推定量は最尤推定量と同一の漸近分布を持つことになる．しかし小標本では最小二乗法はそれほどよいとはいえず，モデル作成の主たる目的が予測にあるならば，撹乱項のARMA過程を推定しなければならなくなる．多項式回帰における最小二乗法の漸近有効性の証明は，Fuller(1976, pp. 388-93)に与えられている．もちろん標本平均の有効性はその特殊ケースになっている．

トレンド要素と撹乱項の加法モデルで $f(t)$ が三角関数の多項式となるときにも最小二乗法によって有効な推定法が与えられる．したがってこの結果を，トレンド項が多項式で季節性が非確率要素である6章(7.16)にも適用することができる．

### 厳密最尤推定法*

モデル(2.1)の尤度関数は **5.3** 節での議論において $y_t$ を $y_t-x_t'\beta$ に置き換えることによって得られる．その尤度関数を $\beta$，$\phi$，$\theta$ について最大化することによって最尤法を得る．

もう一つのアプローチとしては未知母数 $\phi$ と $\theta$ を条件とした一般化最小二乗推定法によって尤度関数を $\beta$ について最大化することが考えられよう．この方法では1章(2.14)での撹乱項の共分散行列を計算する必要がないことのために実用的であろう．撹乱項が純粋のAR$(p)$過程に従うときには，比較的取扱いが容易である(Harvey(1981 a，6.1節，6.4節)参照)．撹乱項がMA過程やARMA過程に従う場合には，それほど自明ではないが，一つの解決手段としてモデルを状態空間形に表現しカルマン・フィルターによって一般化最小二乗推定量を計算することが考えられよう．

例えばMA(1)過程に従う撹乱項を持つ回帰モデルを考えてみよう．未知母

数ベクトル $\beta$ が漸化式 $\beta_t = \beta_{t-1}$ を満たし，状態変数ベクトルを $\alpha_t = (\beta_t', u_t, \theta\epsilon_t)'$ とする．このとき状態変数は遷移方程式，

$$\alpha_t = \begin{bmatrix} I & 0 \\ 0 & T^* \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} 0 \\ R^* \end{bmatrix} \epsilon_t \tag{2.3}$$

かつ

$$T^* = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad R^* = \begin{bmatrix} 1 \\ \theta \end{bmatrix} \tag{2.4}$$

となり（4章(1.7)を参照），観測方程式は，

$$y_t = z_t' \alpha_t = (x_t' \ 1 \ 0)\alpha_t \tag{2.5}$$

となる．この定式化は状態変数ベクトルの一部分が非確率要素となるので状態空間モデルとしては典型的とはいえないが，それにもかかわらず **4.2** 節で述べたカルマン・フィルターには何の変化もない．ここで唯一問題となるのは，適当な初期値の選択についてであろう．逐次最小二乗法のアルゴリズムと同様に考えれば，$a_k = (b_k', 0, 0)'$ とおいて最初の $k$ 個の観測値を用いればよいかもしれない．$P_k$ 行列の作成は MA(1) 過程については比較的容易であるが，高次の ARMA モデルについては相当厄介な問題となる．そこで別の便利な方法としては逐次公式を $t = 0$ より $q_0 = 0$ かつ

$$P_0 = \begin{bmatrix} \kappa I & 0 \\ 0 & P_0^* \end{bmatrix} \tag{2.6}$$

とすることが考えられよう．ここで $\kappa$ は大きいが任意の有限な定数であって $P_0^*$ は 4 章(4.3)で与えられる．$k$ 回逐次公式を適用すればその後は状態変数ベクトル $a_k$ とその共分散行列 $P_k$ は上述の公式より直接計算した値とほぼ一致する．

一般化最小二乗推定量を計算する副産物としてカルマン・フィルターによって $T-k$ 個の基準された予測誤差，

$$v_t = (y_t - z_t' a_{t/t-1}) / (z_t' P_{t/t-1} z_t)^{1/2}, \qquad t = k+1, \cdots, T \tag{2.7}$$

が求められる．母数の値が既知の一般化最小二乗法からの予測誤差あるいは「一般化逐次残差」(generalized recursive residuals) は $NID(0, \sigma^2)$ に従う．この残差は観測値を $L'L = V^{-1}$ を満たす下三角行列 $L$ の変換によって得られた逐次残差と同一である．母数 $\phi$ と $\theta$ を与えた時の $\sigma^2$ の最尤推定量は一般化逐次残差を用いると，



$$\tilde{\sigma}^2(\phi, \theta) = T^{-1} \sum_{t=k+1}^{T} v_t^2 \tag{2.8}$$

と表わされる．この推定量は集約対数尤度関数

$$\log L_c(y; \phi, \theta) = -\frac{1}{2} T \log 2\pi - \frac{1}{2} T \log \tilde{\sigma}^2(\phi, \theta) - \frac{1}{2} \log |V| \tag{2.9}$$

の構成に用いられる．この(2.9)を巡る一つの残される問題は行列式 $|V|$ の評価であろう．この問題についての最も簡単な解決法と考えられるのは対応する ARMA モデルにおける行列 $P_t$ と $P_{t/t-1}$ についての漸化式と同じ補助漸化式を作っておくことであろう．$t = 1, \cdots, T$ に対して $P_{t/t-1}$ 行列の右上要素を $f_t$ の評価に用いれば対数尤度関数の行列式部分は，

$$\log |V| = \sum_{t=1}^{T} \log f_t \tag{2.10}$$

となる．ただし補助漸化式を出発するために必要な初期行列 $P_0^*$ はいずれにしても一般化最小二乗法によるカルマン・フィルターの有機的要素であるので計算する必要がある．((2.6)参照)．

以上述べたアルゴリズムによって $\beta$ の二段階推定量を作ることができる．未知母数 $\phi$ と $\theta$ の一つの一致推定量は最小二乗残差のコレログラムより作られ，それを用いてカルマン・フィルターより厳密な意味での一般化最小二乗推定量を構成できる．さらに近似的にも一般化最小二乗推定量に基づくいくつかの二段階推定法を制約条件付最小二乗アルゴリズムより導くこともできる．(Harvey (1981 a, 6.5 節) 参照)．これらの推定量は漸近的性質は最尤推定量と同等であるが小標本では非常に異なることもありえよう．この問題についてのモンテ・カルロ実験に基づく若干の結果は Harvey=McAvinchey (1981) に報告されている．

最後に予測についてであるが，一般化最小二乗法の定式化では $l$ 期先予測は **4.5** 節の公式を用いて行なえばよい．ただし 4 章(5.4)の公式は母数 $\phi$ と $\theta$ について条件付平均二乗誤差である．いい換えるならば，ここでは $\beta$ の推定に伴う標本変動は考慮されていないことを意味している．

### 累積ピリオドグラム

回帰モデルを最小二乗法によって推定するとき，いわゆる風呂敷統計量やその修正統計量（6章(2.4)）を基にして攪乱項の系列相関についての一般的検定

方式が考えられよう．特に一階の自己相関$r(1)$に関心があるとすればダービン=ワトソン(Durbin=Watson)統計量による限界検定(bound test)を行なうことができよう(Harvey(1981 a, 6.3節)参照)．

系列相関の一般的検定方法としての累積ピリオドグラム法についてはすでに**6.2**節で述べた．6章(2.10)においてピリオドグラムが最小二乗残差から作られるときに，一つの限界検定方式が可能であることをDurbin(1969)は示している．この方法は比較的容易に実行できる．この場合には，一本の直線$s = c_0 + i/n$の代りに二本の直線を引く．もし$s_2$の軌跡が上側の直線を越えると帰無仮説は低周波数領域が過大であるとの対立仮説に対し棄却される．下側の直線を越えなければ帰無仮説は棄却されないが，他方，下側の直線を越えるが上側の直線を越えないときには検定は不定とする．この検定はちょうど正の系列相関に対するダービン=ワトソン検定に対応している．両側検定は45°線の下側に同様の二本の直線を引くことによって行なうことができる．

外側の二直線は$n^* = \frac{1}{2}(T-k)$とすると$\frac{1}{2}k - \frac{1}{2} \leq i \leq n^*$に対し

$$s = \pm c_0 + i/n^*$$

であり内側の二本の直線は

$$s = \pm c_0 + \left[i - \frac{1}{2}(k-1)\right]/n^*$$

で与えられる．すなわち，ここで必要とされるのは標準的累積ピリオドグラム検定の数表のみである．というのは，内側と外側の二直線の差は回帰モデルの自由度$k-1$にのみ依存しているので，上限と下限の詳しい数表は必要としないからである．上限と下限は$T$と$k$がともに奇数であれば精密になり，そうでない場合にもDurbin(1969, p.10)が指摘しているように，$n^*$が小さくなければ近似精度はよく，実際上には何の問題もない．

**例1** 図7.1はもともとKlein(1950, pp. 74-75, 135)が消費を企業利潤と賃金へ回帰した結果の最小二乗残差による累積ピリオドグラムである．$T = 21$, $k = 3$であるから$n^* = 9$となる．5%の有意水準の片側検定では，$c_0 = 0.32538$なので，これより図に2本の直線を引いた．図7.1をみると$s_i$は始め上昇するので残差において低周波数が支配的であることを示唆している．ただし$s_i$は内側の直線は越えているものの外側の直線は越えていない．すなわち，検定結果は不定である．



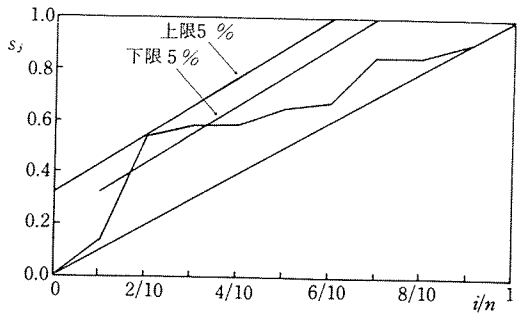



**図7.1** アメリカ合衆国1921～41年の間の消費量の賃金と利潤への回帰残差の累積ピリオドグラム$s_i$(観測数値21)

出典：Durbin(1969, p. 13).

## 7.3 時変係数回帰モデル

回帰方程式の構造が時間とともに変化するものとしよう．未知母数ベクトル$\beta_t$がある確率過程によって生成されると仮定することによってモデルの一つの定式化が可能となる．その候補としては基本的に三つのモデルのクラスが考えられる．第一には母数がある固定された未知の平均の回りを変動することが考えられる．このクラスのモデルは確率的係数(random coefficient)回帰モデルと呼ばれる．その扱いは比較的容易であって，すでに多くの応用例が報告されているが，例えばTheil(1971, pp. 622-627)を参照されたい．ところで母数が確率的に変動するときには，それが全く偶然的に変動するよりも時間の経過とともに徐々に変動すると考える方が自然であろう．そこで，より一般的に$\beta_t$が多変量ARMA過程に従うモデルが考えられる．$\beta_t$の確率過程が定常であると仮定すると$\beta_t$は平均の回りで変動するので，しばしばこのモデルに「定常性への再帰」(return to normality)といった言葉が冠される．

時変係数回帰モデルへの第三のアプローチとしては母数がある非定常過程より生成されるとする考え方がありうる．このクラスの重要な例としては母数が多変量酔歩(random walk)過程に従う場合がある．母数はもはや固定された平均を持つように制約されないので，モデルは時間とともに徐々に発展する．したがって標本期間の最後の観測値は，最初の観測値と非常に異なってくること

がありうる．モデルの背後の関係に急速な変化があるとみられるときには，以上の性質はきわめて魅力的となるかもしれない．

以下では，以上挙げた三つの時変係数回帰モデルのクラスに対する最尤推定法を開発する．定常性への再帰モデルと酔歩係数回帰モデルについてはカルマン・フィルターによって尤度関数がうまく導かれ，その枠組の中から二段階推定量が構成されよう．

### 確率的係数モデル

確率的係数モデルにおいて，時刻 $t$ における未知母数ベクトルが平均 $\bar{\beta}$，共分散行列 $Q$ の多変量正規分布から生成されているとしよう．すなわちすべての $t$ に対して

(3.1 a) $$y_t = x_t'\beta_t,$$

(3.1 b) $$\beta_t = \bar{\beta} + \epsilon_t, \qquad \epsilon_t \sim NID(0, Q)$$

としよう．ここで $x_t$ の第1要素が常に1をとる変数であれば(3.1 a)に攪乱項をつけ加える必要はない．定数項に対応する(3.1 b)の攪乱項によって方程式の誤差項は含まれることになる．

そこで(3.1 b)を(3.1 a)に代入すると，

(3.2) $$y_t = x_t'\bar{\beta} + u_t, \qquad t = 1, \cdots, T$$

を得る．ここで $u_t = x_t'\epsilon_t$ である．モデル(3.2)は分散が不均一な標準的回帰モデルであり，攪乱項は互いに独立で平均ゼロ，分散

(3.3) $$\sigma_t^2 = x_t'Qx_t, \qquad t = 1, \cdots, T$$

の正規分布に従うことになる．ここでもし $Q$ が対角行列ならばモデルの扱いは簡約化されて，$q_1$, $q_2$, …, $q_k$ を $Q$ の対角要素とすると(3.3)は，

(3.4) $$\sigma_t^2 = q_1 + q_2 x_{2t}^2 + \cdots + q_k x_{kt}^2, \qquad t = 1, \cdots, T$$

と表現される．

モデルを(3.2)と(3.4)に定式化すれば比較的簡単な二段階推定法が導かれる．$q_1$, …, $q_k$ の推定量はまず $t$ 番目の最小二乗残差の二乗 $\hat{u}_t^2$ を $x_{2t}^2$, …, $x_{kt}^2$ に回帰することによって得られる．次に推定量 $\hat{q}_1$, …, $\hat{q}_k$ はスコアリング法を用いることによって求めることができる．フィッシャー情報行列は $\bar{\beta}$ と $q$ に関してブロック対角であり，スコアリング法による反復計算は $\bar{\beta}$ について(3.5 a)の加重最小二乗法に帰着することになる．また $q$ についてのスコアリング法の反復計算も加重最小二乗法



(3.5 b) $$q^* = \hat{q} + \left[E\left(-\frac{\partial^2 \log L}{\partial q \partial q'}\right)\right]^{-1} \frac{\partial \log L}{\partial q}$$
$$= \left[\sum_{t=1}^{T} \hat{\sigma}_t^{-4} w_t w_t'\right]^{-1} \sum_{t=1}^{T} \hat{\sigma}_t^{-4} w_t (y_t x_t' \tilde{\beta})^2$$

から $\bar{\beta}$ の加重最小二乗(weighted least squares)推定量

(3.5 a) $$\tilde{\bar{b}} = \left[\sum_{t=1}^{T} \hat{\sigma}_t^{-2} x_t x_t'\right]^{-1} \sum_{t=1}^{T} \hat{\sigma}_t^{-2} x_t y_t$$

を作る．ただし $\hat{\sigma}_t^2 = \hat{q}_1 + \hat{q}_2 x_{2t}^2 + \cdots + \hat{q}_k x_{kt}^2$ とする．この加重最小二乗推定量 $\tilde{\bar{b}}$ は $\bar{\beta}$ の一つの一般化最小二乗推定量であり，適当な正則条件の下で $\hat{q}_1$, …, $\hat{q}_k$ は一致性を持ち $\tilde{\bar{b}}$ は漸近的に有効となる．母数 $\bar{\beta}$ と $q = (q_1, \cdots, q_k)'$ の最尤推定量に帰着される．ここで $w_t = (1\ x_{2t}^2, \cdots, x_{kt}^2)'$ である．

### 定常性への再帰モデル*

方程式(3.1 a)において未知母数が定常多変量 AR(1)過程

(3.6) $$\beta_t - \bar{\beta} = \Phi(\beta_{t-1} - \bar{\beta}) + \epsilon_t$$

によって生成されるとしよう．ここで $\epsilon_t \sim NID(0, Q)$ であって $\Phi$ は $k \times k$ の母数行列である．$\Phi$ は $k^2$ 個の母数を含んでいるが，このモデルでは一般性を失なうことなしに対角行列としておく．このときには $\beta_t$ の定常性の条件は単に $\Phi$ の各要素の絶対値が1より小さいことを意味する．母数 $\beta_t$ がこのような過程に従っていることは，それが時間とともに平均 $\bar{\beta}$ の回りで緩やかに変化することを意味している．Schaefer *et al.*(1976)は市場シェアのリスク問題に関連してこの種の母数変化を検討し受け入れているし，初期の応用例としてはRosenberg(1973)をあげることができる．

確率的係数モデル(3.1 b)は(3.6)において，とくに $\Phi = 0$ とすることによって得ることができる．ただし自己回帰要素の導入によって潜在的にはきわめて有用な定式化が導かれ，さらに $\beta_t$ が任意の定常 ARMA 過程に従う場合も扱うことができる．任意の多変量 ARMA 過程は多変量 AR(1)過程として表現することができ，カルマン・フィルターによってその尤度関数は評価できる．そこでこの方法はこの節でのモデルにも適用可能である．しかしながら，$\beta_t$ が任意の ARMA 過程に従う場合を原理的に定式化できてもその構造を識別し推定する実際問題は(3.6)よりもはるかに複雑になる．

さて(3.6)を(3.1 a)に代入すると回帰モデル(3.2)における攪乱項は，

(3.7) $$u_t = x_t'(\beta_t - \bar{\beta}) = x_t'\delta_t, \qquad t = 1, \cdots, T$$

と表現される．$u_t$ の期待値はゼロであるが分散不均一でしかも系列相関があるために，確率的係数回帰モデルとも異なっている．$(u_1, \cdots, u_T)'$ の共分散行列の要素は

(3.8) $$E(u_t u_s) = x_t'\Gamma(\tau)x_s, \qquad t, s = 1, \cdots, T$$

で与えられる．ここで $\tau = t - s$ であり $\Gamma(\tau)$ は $\beta_t - \bar{\beta}$ の $\tau$ 次自己共分散行列である．もし $\Phi$ と $Q$ がともに対角行列であれば，(3.8)は

(3.9) $$E(u_t u_s) = \sum_{i=1}^{k} x_{it}x_{is}q_i\phi_i^{\tau}/(1-\phi_i^2), \qquad t, s = 1, \cdots, T$$

となる．表現(3.9)によって次の二段階推定法が考えられよう．まず $\tau = 0, 1$ に対して $E(u_t u_s)$ を $\hat{u}_t\hat{u}_s$ でおき換えるとすると，$\hat{u}_t^2$ から $x_{2t}^2$，$\cdots$，$x_{kt}^2$ への回帰式および $\hat{u}_t\hat{u}_{t-1}$ から $x_{2t}x_{2t-1}$，$\cdots$，$x_{kt}x_{kt-1}$ への回帰式が考えられる．こうして $2k$ 個の係数推定値を求め，それを $\Phi$ と $Q$ の $2k$ 個の未知母数について解くことができる．そこで $\hat{\gamma}_{i0}$ を $\tau = 0$ に対する回帰式による第 $i$ 番目の回帰係数推定値，$\hat{\gamma}_{i1}$ を $\tau = 1$ に対する第 $i$ 番目の回帰係数推定値とすると，$\hat{\phi}_i = \hat{\gamma}_{i1}/\hat{\gamma}_{i0}$ によって $\phi_i$ が推定される．母数 $\phi_i$ の推定値が得られると直ちに $q_i$ の推定値を得ることができる．この $q_1$，$\cdots$，$q_k$ 及び $\phi_1$，$\cdots$，$\phi_k$ の推定は確率的係数回帰モデルにおける二段階推定量の第一段階の自然な一般化になっている．第二段階では，確率的係数回帰モデルと同じようにして $\bar{\beta}$ の一般化最小二乗推定量を構成する．モデルの状態空間表現を求めてさえいれば，(3.8)で与えられる共分散行列の逆転を行なうことなくカルマン・フィルターによって一般化最小二乗推定量を求めることができる．

このモデルの状態空間表現は **7.2** 節で求めた ARMA 過程に従う攪乱項付回帰モデルと同様にして求めることができる．すべての $t$ について $\bar{\beta}_t = \bar{\beta}$ とするときに状態変数ベクトルを $\alpha_t = (\bar{\beta}_t', \delta_t')'$ とおき，$z_t = (x_t', x_t')'$ とする．遷移方程式は(2.3)と同様に $T^* = \Phi$，$R^* = I$ とする．通常は $t = 0$ より逐次公式を $P_0$ を(2.6)として始めるとつごうがよい．この場合には，行列 $P_0^*$ は $\Gamma(0)$ と等しいが，もし $\Phi$ と $Q$ が対角行列であれば $\Gamma(0)$ も対角行列となりその $i$ 番目の要素は $q_i/(1-\phi_i^2)$ となる．

最尤推定については二つのアプローチがありうる．第一に一般化最小二乗法に基づき尤度関数を $\bar{\beta}$ について集約することが考えられよう．$\epsilon_t$ の共分散行列を $\sigma^2\bar{Q}$ と再定義する．ただし $\bar{Q}$ の(1, 1)要素は 1 である．そこで一般化最小二乗法に基づくカルマン・フィルターの漸化式は $\sigma^2$ と独立になるので，$\sigma^2$ について尤度関数を集約化して(2.9)より母数 $\Phi$ と $\bar{Q}$ に関して尤度を最大化すればよいこととなる．



第二のアプローチでは尤度関数の集約化は考えない．$y_t^{\dagger} = y_t - x_t'\bar{\beta} = x_t'\delta_t$ は定常確率過程であるから，1章(2.21)と同様にして $T$ 個のインベーション(予測誤差)を用いて分解される．そこで状態空間表現は，

(3.10) $$y_t^{\dagger} = x_t'\delta_t,$$
$$\delta_t = \Phi\delta_{t-1} + \epsilon_t, \qquad t = 1, \cdots, T$$

で与えられる．状態空間ベクトル $\delta_t$ の初期値は定常性の条件から自動的に導かれる．すなわち本節では，$\delta_0 \sim N(0, \Gamma(0))$ となる．ところで本節ではすでに述べたように $\Gamma(0)$ は対角行列となっている．

**酔歩係数回帰モデル**

酔歩(random-walk)母数係数の回帰モデルとは，

(3.11 a) $$y_t = x_t'\beta_t + \epsilon_t, \qquad t = 1, \cdots, T$$

において未知母数ベクトルが確率過程

(3.11 b) $$\beta_t = \beta_{t-1} + \eta_t$$

に従うモデルのことである．ここで $\eta_t \sim NID(0, \sigma^2 Q)$ とする．このモデルは前に述べた自己回帰母数の回帰モデルとは異なり，母数は時間とともに発展(evolve)する．したがって，このモデルはかなりの構造変化に対応すると考えることができよう．

さて $k \times k$ 行列 $Q$ は母数が変化し得る程度を決定する．もし $Q = 0$ であれば，すべての $t$ について $\beta_t = \beta_{t-1}$ となってモデルは古典的回帰モデルに退化する．もし $Q$ が正定符号行列であるならば $k$ 個の母数は時間とともに変化する．以下の議論では便宜上一つの母数が確率変数ならば他の母数も確率変数とする．むろんこのことが常に正しい必要はない．

この酔歩係数回帰モデルに対しては，これまでのように尤度関数を構成できないのは明らかであろう．この問題についての Cooley=Prescott (1976) の方法では，$\beta_T$ を固定された母数とみなす．このとき代入を繰り返すと各 $\beta_t$ は $\beta_T$ によって表現される．例えば

$$\beta_T = \beta_{T-1} + \eta_T = \beta_{T-2} + \eta_{T-1} + \eta_T$$

あるいは

$$y_t = x_t'\beta_T + \zeta_t, \qquad t = 1, \cdots, T$$

となる．より一般的にはモデルを

$$y_{T-2} = x_{T-2}'\beta_T + [\epsilon_{T-2} - x_{T-2}'(\eta_{T-1} + \eta_T)],$$

ただし

$$\zeta_t = \epsilon_t - x_t'(\eta_{t+1} + \eta_{t+2} + \cdots + \eta_T) \tag{3.12}$$

と表現することができる．そうするとこのモデルは標準的一般化最小二乗法と見なすことができるかもしれない．母数ベクトル $\beta_T$ を固定した下で攪乱項は平均ゼロ，共分散行列 $\sigma^2 V$ を持つ．ここで行列 $V$ の $(i,j)$ 要素は，

$$v_{ij} = [\min(T-i, T-j)] \cdot x_i' Q x_j + \delta_{ij} \tag{3.13}$$

で与えられる．ただし $i \neq j$ に対し $\delta_{ij}$ はゼロで，$i = j$ に対し $\delta_{ij}$ は 1 である．そこで $Q$ が与えられたときに未知母数 $\beta_T$ の一般化最小二乗推定量が計算される．次に尤度関数を未知母数 $\beta_T$ と $\sigma^2$ について集約化し，さらに表現 (2.9) を $Q$ の要素に関して最大化することによって最尤推定が可能となろう．Cooley=Prescott (1976) は $T \times T$ 行列 $V$ の逆行列を計算せずに最尤推定が可能であることを主張している[訳注]．

ところで母数 $\beta_T$ を固定したものとみなすことはきわめてアド・ホック (恣意的) である．実際には Cooley=Prescott は $\beta_{T+1}$ を固定したとしている．しかしながらいろいろな $\beta_t$ の条件の取り方の有効性を考えるよりもカルマン・フィルターに基づく別のアプローチを考えることができよう．本節の状況では以下に説明する方法はきわめて自然な方法であろう．すなわち，$\alpha_t = \beta_t$，$z_t = x_t$，$\xi_t = \epsilon_t$ と記号を修正することによって，このモデルを **4.3** 節の状態空間形に表現できる．唯一残される問題は初期値の設定である．

初期値の設定問題は逐次最小二乗法にも現われたが，そこでは最初の $k$ 個の観測値を用いて初期値を設定することで解決された．同じ方法がここでも適用できよう．最初の $k$ 個の観測値を

$$y_t = x_t'\beta_k + \zeta_t, \qquad t = 1, \cdots, k$$

と表わしておく．ここで $\zeta_t$ は $T = k$ とした (3.12) で定義される．このとき，$\beta_k$ の最小分散不偏推定量は，

$$b_k = (X_k' V_k^{-1} X_k)^{-1} X_k' V_k^{-1} y_k^* = X_k^{-1} y_k^* \tag{3.14}$$

訳注　この方法については問題があることが次の論文によって指摘されている．Tsurumi, H. and T. Shiba (1981) "On Cooley and Prescott's Time Varying Parameter Model," *The Economic Studies Quarterly*, Vol. XXXII.



で与えられる．ここで $X_k$，$y_k^*$ は (1.4) で与えられるとともに，$\sigma^2 V_k$ は $(\zeta_1, \cdots, \zeta_k)'$ の共分散行列である．$b_k - \beta_k$ の共分散行列は $\sigma^2 P_k$ で与えられるが，ここで

$$P_k = (X_k' V_k^{-1} X_k)^{-1} \tag{3.15}$$

である．この $b_k$ と $P_k$ によってカルマン・フィルターの初期値が設定される．観測値をすべて利用した後には $Q$ を条件とした $\beta_T$ の最小分散不偏推定量が求められる．もし $Q = 0$ かつ $P_k = (X_k' X_k)^{-1}$ であれば，以上の方法は逐次最小二乗アルゴリズムに還元される．もし漸化式を時刻 $t = 0$ において $a_0 = 0$ かつ十分大きいスカラー $\kappa$ を用いて $P_0 = \kappa I$ とすれば $b_k$ と $P_k$ を明示的に求める必要はなくなる．

さて (3.14) と (3.15) を導くにあたって母数 $\beta_k$ を固定しておく必要はない．実際にカルマン・フィルターから尤度関数を求めるためには，$b_k$ を固定することが必要である．このことは $\beta_k$ が固有の事前分布 $\beta_k \sim N(b_k, \sigma^2 P_k)$ を持つことを意味し，予測誤差分解より $y_{k+1}$，$\cdots$，$y_T$ の尤度関数はカルマン・フィルターによる予測誤差によって，

$$\log L(y_{k+1}, \cdots, y_T) = -\frac{1}{2}(T-k)\log 2\pi - \frac{1}{2}(T-k)\log\sigma^2 - \frac{1}{2}\sum_{t=k+1}^{T}\log f_t - \frac{1}{2}\sigma^{-2}\sum_{t=k+1}^{T} v_t^2/f_t \tag{3.16}$$

と表現される．一見すると (3.16) は一般化最小二乗法から得られた $T$ 個の観測値についての尤度関数よりも制約されているようにみえる．しかしながら，そうではない．なぜなら $y_1$，$\cdots$，$y_k$ は固定されているがすべての母数が確率変数として扱れているからである．

尤度関数はスカラー母数 $\sigma^2$ について集約することができる．そこでさらに $Q$ の最尤推定量は関数

$$\log L_c = \log\tilde{\sigma}^2(Q) + (T-k)^{-1}\sum_{t=k+1}^{T}\log f_t$$

を最小化することによって得ることができる．ただし，

$$\tilde{\sigma}^2(Q) = (T-k)^{-1}\sum_{t=k+1}^{T} v_t^2/f_t$$

である．

**例 1**　1916-66 年の年データを用いて，Garbade (1977) が最小二乗法によって推定した貨幣需要関数は次式で与えられる．

$$\widehat{\Delta \log M1_t} = 0.0020 - 0.0833\,\Delta \log Rcp_t + 0.4947\,\Delta \log y_t \tag{3.17}$$

$$R^2 = 0.360 \quad s = 0.051 \quad DW = 1.43$$

ここで変数としては 1 人当り狭義の貨幣供給量 ($M1$)，利子率を表わす商業手形利回り ($R_{cp}$)，1 人当り実質所得 ($y$) を用いている．Garbade は次に (3.11 b) で与えられる酔歩係数回帰モデルを推定している．行列 $Q$ は対角とし，上で述べたカルマン・フィルターを用いて三つの要素の推定値 $q_{11}$，$q_{22}$，$q_{33}$ を計算した．その結果は $\tilde{q}_{11} = 0.1982$，$\tilde{q}_{22} = 2.6791$，$\tilde{q}_{33} = 5.4574$，$\tilde{\sigma} = 0.0337$ であった[1]．

行列 $Q$ を推定して Garbade はさらに **4.5** 節で述べたアルゴリズムを用いて $\beta_t$ の平滑化推定値を求めている．彼の論文の 62 ページのグラフによれば，$\beta_{1t}$ と $\beta_{2t}$ には明瞭な変化が検出されるが，他方，所得弾力性 $\beta_{3t}$ は 1930 年代中頃に約 0.5 前後となって安定化する．ただし，このような母数値の変化はデータを生成せしめている確率過程の構造の変化よりも，むしろ動学モデルの定式化を誤ったことによるものとも考えられる．実際，母数推定値の軌跡はきわめて有力なモデル診断法であることが Salmon (1980) によって主張されている．

### 過去の観測値の割引き

本章 **7.1** 節で未知母数が時間に関して安定的であるかどうか疑わしいときの回帰モデルの推定法として割引最小二乗法を提案した．酔歩係数回帰モデルについてはカルマン・フィルターのアルゴリズムを用いてモデルの未知母数を推定するアプローチが考えられた．その場合には行列 $Q$ は割引率に対応してあらかじめ固定しておけば，それが母数の推定値の中に入ってくる．

もし $x_t$ ら定数項のみから成るとすれば (3.11) は 6 章 (5.32) に帰着される．そして (5.32) における最適予測は指数型加重移動平均の形をとることになる．すなわち (3.11) に基づく割引最小二乗法の考え方は標準的時系列手法の自然な一般化とみなすことができよう．

## 7.4　周波数領域における回帰分析

古典的な線型回帰モデルは，その従属変数と独立変数に対して有限のフーリ

1) Garbade の方法と本文で述べた方法とは細かい一つの相違点がある．Garbade のアルゴリズムでは $P_k$ は (3.15) を用いて計算されるわけではない．ここで述べた方法はもともと Harvey and Phillips (1976 b) によって提案されたものである．



エ変換をほどこすことによって，周波数領域において解釈することができる．この変換によって時間でなく周波数によって目盛りづけられた $T$ 個の観測値が得られる．そして変換された従属変数を変換された独立変数の上に回帰することにより，いわゆるスペクトラム回帰分析 (spectral regression) が行なわれる．この回帰分析を行なうには多くの理由があるが，その第一は帯状スペクトラム回帰分析 (band spectrum regression) と呼ばれる方法の適用を可能にすることである．帯状スペクトラム回帰分析とは周波数領域においてある特定の周波数を除いて回帰分析を行なうものである．この方法は季節性とか説明変数における誤差を含むモデルなどへの多くの興味ある応用が可能である．たとえば，もし観測値が強い季節性を示す場合には，変換された変数のうち季節的周波数の周辺にある観測値を除くことによって，変数間の関係の分析における季節性の影響を弱める試みに有用であるかもしれない．観測上の誤差の問題に関しては，周波数領域の回帰分析から高周波数の観測値を除くことによって，最小二乗推定量の好ましからざる結果が弱められることが示唆される．経済変数はゆっくり変化しがちであり，一方，典型的な誤差はホワイト・ノイズであるとみなされるので，変数における歪みは高周波数においてより深刻であると考えられる．他方，低周波数は，このような種類の測定誤差に関しては比較的頑健である．したがってスペクトラム回帰分析をこれらの周波数に限って行なうということによって利益がもたらされるかもしれないのである．これらの考えについてのさらにより進んだ議論は，Engle (1974 a, 1978 b) に見い出すことができる．

スペクトラム回帰分析に対して関心が払われる第二の理由は，時間領域において撹乱項が系列相関を持っているとしても，周波数領域では変換された撹乱項は近似的に無相関になることをあげることができよう．もっとも，系列相関の除去は不均一分散の導入という不利益な結果をもたらす．しかしながら，変換後の撹乱項の分散がもとの撹乱項のスペクトラム密度と密接な関係があることから，いくつかの便益をもたらし，これは推定と検定の双方に含意を持つこととなる．とくにスペクトラムの推定量は，それが撹乱項の確率過程についての特定の定式化に依存しないという意味において，ノン・パラメトリック (non-parametric) であるということができよう．

### 周波数領域への変換

行列表示による一般化線型回帰モデルは以下のように書ける．

(4.1) $$y = X\beta + u$$

ここで $u$ は $T\times 1$ の撹乱項ベクトルで，平均はゼロおよび共分散行列は $\sigma^2 V$ である．(たとえば1章(2.12)を参照)．ここで $W$ を3章(2.19)で定義された $T\times T$ のフーリエ行列 $Z$ の転置行列としよう．ここで(4.1)に与えられている観測量に $W$ を前から掛けることにより，以下の結果を得る．

(4.2) $$\dot{y} = \dot{X}\beta + \dot{u}.$$

ここで $\dot{y} = Wy$，$\dot{X} = WX$ そして $\dot{u} = Wu$ である．3章(2.19)で定義されたフーリエ行列は $T$ が偶数の場合であるが，$T$ が奇数の場合の変換行列も同様の形式をとる．そして $T$ が奇数の場合の適当な形は，**3.2** 節に戻ることによって簡単に導くことができるであろう．

このように観測量を変換することによって，線型回帰モデルの周波数領域における解釈が導かれる．しかしながら，(4.2)は(4.1)に比べてより多くのまたはより少なく情報を含んでいるわけではない．実際に，もし(4.1)の撹乱項がホワイト・ノイズであるならば，(4.2)における撹乱項もホワイト・ノイズである．このことは $W$ の直交性から直ちに導かれる．なぜなら

$$E(\dot{u}\dot{u}') = E(Wuu'W') = \sigma^2 WW' = \sigma^2 I$$

が成立するからである．つまり周波数領域において変換された観測量に最小二乗法をほどこすことは，$\beta$ の最良線型不偏推定量をもたらすことになるが，この推定量は変換されないもとの観測量から計算される最小二乗推定量と同一である．正式な論証は以下のように直接的に与えられる．

(4.3) $$b = (\dot{X}'\dot{X})^{-1}\dot{X}'\dot{y} = (X'W'WX)^{-1}X'W'Wy = (X'X)^{-1}X'y.$$

変換された観測量について興味のある点は，連続した二つの観測量ごとに異なった周波数に対応していることである．したがって観測量の集合から適当な観測量を取り除くことによって，簡単に特定の周波数要素が除かれることになる．これが帯状スペクトラム回帰分析として知られているものである．変換された観測値は実数なので，どのような標準的な回帰分析計算パッケージもその計算を行なうのに使うことができるであろう．さらに古典的な回帰モデルに関して該当するすべての結果が，修正無しに適用することができよう．すなわち，たとえば $u_t \sim NID(0,\sigma^2)$ の場合には，ある周波数に対応する観測値が残りの観測値と同じモデルに従うか否かを検定するのに，$F$ 統計量を用いることができよう．このような検定は，通常の回帰モデルにおける共分散分析と同一のものである．



それぞれの変数に対するフーリエ変換の影響について考えることによって，スペクトラム回帰において何が起っているのかについて，より深い洞察を得ることができよう．ここで簡単化のためにモデルがただ一つの説明変数を含んでいるとすると，周波数領域では以下のように表わされる．

(4.4) $$\dot{y}_s = \beta\dot{x}_s + \dot{u}_s, \qquad s = 1,\cdots,T$$

すると，もとの説明変数に関する観測量のピリオドグラムは以下のように与えられる．

(4.5) $$P_j = \dot{x}_{2j}^2 + \dot{x}_{2j+1}^2, \quad (j = 1,\cdots,n-1),$$
$$P_n = \dot{x}_T^2.$$

比例定数を除けば，ピリオドグラムは標本スペクトラム密度と同じである．以後の展開への観点からは，これらの二つのうち後者について考えるのがより容易であろう．そこで $I_x(\lambda_j) = p_j/4\pi$，$j = 1,\cdots,n-1$，かつ $I_x(\lambda_n) = p_n/2\pi$ と置くことにしよう．ここで $I_x(\lambda_0)$ は $I_x(\lambda_0) = \dot{x}_1^2/2\pi = T^{1/2}\bar{x}/2\pi$ と置くことにする．標本相互スペクトラム $I_{xy}^*(\lambda_j)$ の実数部分も同様に定義される．ここで $I_{xy}^*(\lambda_j) = (\dot{x}_{2j}\dot{y}_{2j} + \dot{x}_{2j+1}\dot{y}_{2j+1})/4\pi$，$j = 1,\cdots,n-1$ である．したがって $\beta$ の最小二乗推定量は次のように書くことができる．

(4.6) $$b = \sum_{s=1}^{T}\dot{x}_s\dot{y}_s \Big/ \sum_{s=1}^{T}\dot{x}_s^2 = \sum_{j=0}^{n}\psi_j I_{xy}^*(\lambda_j) \Big/ \sum_{j=0}^{n}\psi_j I_x(\lambda_j).$$

ここで $j = 0$ および $j = n$ の時は $\psi_j = 0.5$ であり，それ以外のときは $\psi_j = 1$ である．

ところで，標本相互スペクトラムの虚数部分は奇関数であるので，両側の和をとることにより消えてしまう．ゆえに(4.6)は次のように書き改めることができる．

(4.7) $$b = \sum_{j=-n}^{n-1} I_{xy}(\lambda_j) \Big/ \sum_{j=-n}^{n-1} I_x(\lambda_j).$$

ここで $I_{xy}(\lambda_j)$ は標本相互スペクトラムである．観測値が奇数の場合には，(4.7)における和は $-n$ から $n$ についてとられる．

モデルが $k > 1$ の独立変数を含む一般の場合においては，$I_x(\lambda_j)$ は $k\times k$ の標本スペクトラムおよび相互スペクトラム行列 $I_{xx}(\lambda_j)$ によって置き換えられる．同様に $I_{xy}(\lambda_j)$ は $k\times 1$ の相互スペクトラムのベクトルであり，(4.7)は以下のように書き換えられる．

$$b=\left[\sum_{j=-n}^{n-1} I_{xx}(\lambda_j)\right]^{-1}\sum_{j=-n}^{n-1} I_{xy}(\lambda_j). \tag{4.8}$$

### 系列相関

回帰モデルにおける攪乱項が，平均はゼロであるが定常な確率過程によってもたらされていると想定しよう．このモデルの周波数領域への変換は，以下のような共分散行列を持つ攪乱項ベクトル $\dot{u}$ をもたらす．

$$\dot{\Gamma}=E(\dot{u}\dot{u}')=WE(uu')W'=W\Gamma W'.$$

もとの攪乱項が定常であるので，$\Gamma$ はトエプリッツ(Toeplitz)行列である．そしてこの型の行列は，有限フーリエ変換によって近似的に直交化されるということを示すことができる．すなわちもし $\dot{\gamma}_{ij}$ が $\dot{\Gamma}$ の $ij$ 要素を表わすとすれば，以下のような近似が与えられる．

$$\dot{\gamma}_{ij}\simeq\begin{cases}2\pi f_u(\lambda_j), & i=j\\ 0, & i\neq j.\end{cases} \tag{4.9}$$

ここで $f_u(\lambda)$ はもとの攪乱項のスペクトラム密度である．もし攪乱項を生成している確率過程が循環過程(circular)であるならば，この表現は正確なものである．この仮説が成立しない場合でも，近似誤差はもし $T$ が大きければ無視し得るものである．このことについては，たとえばFishman(1969)を参照されたい．

もし(4.9)が厳密に成立し，攪乱項のスペクトラム密度が既知であれば，周波数領域で $\beta$ の一般化最小二乗推定量が加重最小二乗法により計算できる．もちろん $f_u(\lambda)$ が未知であるようなより一般的な場合には，それがまず推定されなければならない．この推定が一致性をもってなされるという前提の下では，通常の標準的な漸近理論を適用して，結果として得られる実行可能な一般化最小二乗推定量が漸近的に有効であるということを示すことができる．

この $f_u(\lambda)$ を推定するために必要な追加的な計算はそれほど大きな重荷ではない．$\beta$ の一致推定量，たとえば最小二乗推定量が与えられれば，対応した残差は正しい残差の一致推定量となる．ここでもし真の攪乱項が既知ならば，$f_u(\lambda)$ はそのピリオドグラムを平滑化することにより一致性をもって推定される．攪乱項の一致推定量から計算されたピリオドグラムに対して，同様な平滑化操作を適用すると，やはり $f_u(\lambda)$ の一致推定値を得ることができる．このような方法によって $f_u(\lambda)$ の推定量を求めるのに必要な計算は，それほどたいしたことではない．なぜなら，もし $e=y-Xb$ を時間領域における最小二乗残差のベクトルとすれば，$\dot{e}=\dot{y}-\dot{X}b=We$ が成立するからである．したがって最小二乗残差のピリオドグラムは，周波数領域の残差から直接的に得ることができる．以上から観測量を周波数領域に変換した際の漸近的に有効な $\beta$ の推定量は，二つの段階を経て計算することができる．第一段階では，最小二乗法を変換されたモデル(4.2)に適用することにより，$\beta$ の一致推定量を得る．もとの攪乱項のスペクトラム密度関数は，この第一段階の結果として生じた周波数領域の残差を二乗し，連続した二つずつを加え合せ，そして平滑化することにより直接的に推定することができる．第二段階では，加重最小二乗を用いることにより有効な推定量が得られる．もし望むならば，推定量の小標本特性を改善するという観点から，この手続きを繰り返すこともできよう．そこに含まれる計算のコストは，最初にフーリエ変換を行なう際にかかるコストに比べて相対的に小さなものである．



さて一般化最小二乗推定量は，(4.7)における最小二乗推定量とほとんど同じような形に書ける．単一の説明変数の方程式に関しては，実行可能な一般化最小二乗法は以下のように与えられる．

$$\tilde{b}=\sum_{j=-n}^{n-1}\hat{f}_u^{-1}(\lambda_j)I_{xy}(\lambda_j)\Big/\sum_{j=-n}^{n-1}\hat{f}_u^{-1}(\lambda_j)I_x(\lambda_j). \tag{4.10}$$

ここで $\hat{f}_u(\lambda_j)$ は $f_u(\lambda)$ の一致推定量である．$\hat{f}_u(\lambda)$ はピリオドグラムの平滑化によってもたらされているので，$I_x(\lambda_j)$ および $I_{xy}(\lambda_j)$ の平滑化を考慮することもまた自然であろう．さらにスペクトラム密度関数の連続性についての議論によって，スペクトラム密度関数や相互スペクトラム密度関数を評価する点の数を減少させても，推定量にほとんど影響を持たないということが知られている．もし $N$ がスペクトラム密度の推定量において用いられた周波数の種類の数であり，$\lambda_j=2\pi i/N$ とすると，(4.10)の代りに以下のような推定量が考えられるであろう．

$$\tilde{b}^{\dagger}=\sum_{i=-N}^{N-1}\hat{f}_u^{-1}(\lambda_i)\hat{f}_{xy}(\lambda_i)\Big/\sum_{i=-N}^{N-1}\hat{f}_u^{-1}(\lambda_i)\hat{f}_x(\lambda_i). \tag{4.11}$$

適当な条件の下に $\tilde{b}^{\dagger}$ は $\tilde{b}$ と同じ漸近分布を持つ．上記の二つの推定量の選択は，主に計算上の簡便さに依存するであろう．もっとも，ホワイト・ノイズの攪乱項の下での推定量の性質についての結論からEngle=Gardner(1976)は(4.10)を強く推奨している．

上記(4.11)を一般化最小二乗法と解釈すると，以下のような結果を意味する．

$$\text{(4.12)} \qquad \text{Avar}\,(\tilde{b}^{\dagger}) = T^{-1}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} f_u^{-1}(\lambda) f_x(\lambda)\, d\lambda\right)^{-1}.$$

この種の表現を評価することは，時間領域における対応した形を評価するよりも一般的に容易である．したがって周波数領域で検討をすることは，しばしば実際的かつ理論的に便利である．この点に関しては，ここではこれ以上追求しない．しかし興味ある読者は，より詳しいことについて Engle=Gardner (1976) または Nicholls=Pagan (1979) を参照されたい．

## 7.5 系列相関のノン・パラメトリックな扱い

もし回帰モデルにおける攪乱項を生成している確率過程について定式化の誤りがあった場合には，推定量の効率に重大な損失を生じるであろう．この種の定式化の誤りの最も一般的な例は，より一般的な ARMA モデルが適当なときに，AR(1)過程によって $u_t$ が生成されていると想定するものである．このような状況のときには，AR(1)過程の下での推定量は最小二乗推定量よりもさらに非効率になる．いい換えると，攪乱項を恣意的にモデル化すると修正をまったく行なわない最小二乗法よりも実際にはかえって悪くなる可能性がある．このことについてのいくつかの証拠は Engle (1974 b) 及び Engle=Gardner (1976) などに見い出すことができる．

このような問題を避ける一つの方法は，(4.10)または(4.11)のようなスペクトラム密度関数の推定量を用いることである．この方法で計算された推定量は漸近的に有効推定量であるが，攪乱項を生成している確率過程を定式化する必要がないという意味においてノン・パラメトリックである．ノン・パラメトリックな手法によって母数 $\beta$ の推定量を求めるもう一つの方法は，攪乱項を AR $(p)$ 過程としてモデル化することである．ここで $p$ は $T\to\infty$ のときに $p/T$ が一定値に留まるように選ばれる．この方法は本質的には(4.10)の時間領域における対応物であり，この推定量は比較的簡単に計算できる．AR 部分の母数，$\phi_1$, …, $\phi_p$ の一致(かつ有効)推定量を得るためには最小二乗残差 $e_t$ をそのラグ値 $e_{t-1}$, …, $e_{t-p}$ 上に回帰すればよい．そして $\beta$ の推定量は

$$y_t - \hat{\phi}_1 y_{t-1} - \cdots - \hat{\phi}_p y_{t-p}, \qquad t = P+1, \cdots, T$$

を，以下で与えられる量の上に回帰することによって得られる．



$$x_t - \hat{\phi}_1 x_{t-1} - \cdots - \hat{\phi}_p x_{t-p}, \qquad t = P+1, \cdots, T.$$

この自己回帰最小二乗推定量は，しばしば ALS$(P)$ と呼ばれる．

もし $u_t$ のスペクトラム密度関数が **3.6** 節の最後で説明した自己回帰法によって推定されるならば，時間領域における推定量と周波数領域における推定量の性質は非常に似たものとなろう．もっとも両者の関係は(4.9)によってもたらされる近似および $\hat{f}_u(\lambda)$ を得るための平滑化のために，厳密に一致はしないであろう．ある特殊な場合にはこのことが重要な違いをもたらすかもしれないが，多くの実際的な目的においては二つの推定量は同一と考えてよいであろう．

このノン・パラメトリック推定量は漸近的に有効であるけれども，小標本においても有効であるということを必ずしも意味しない．Engle=Gardner (1976) に示されているモンテ・カルロ実験の結果によれば，$T = 100$ の場合に，スペクトラム(密度に基づく)推定量の分散は最善で有効推定量の漸近分散の 2 倍となる．このことは，適切に倹約された攪乱項の定式化に基づく，小標本においてより効率的推定量の開発へ多くの余地があることを意味する．実際に，もし攪乱項に定式の誤りがある場合でも，小標本においてはそれに基づく推定量がノン・パラメトリック推定量よりも，より効率的であるということが十分起こり得る．Engle=Gardner は，この現象についてもいくつかの証拠を示している．彼らはある一つの実験において，攪乱項は実際には AR(2)過程によって生成されたのにもかかわらず，$T = 100$ の場合に ALS(1)は，スペクトラム推定量より劇的によい結果をもたらしていることを示している．このことを裏づける証拠は，Gallant=Goebel (1976) によっても与えられている．彼らはある特定の(非線型)回帰モデルにおいて，もととなる MA(4)攪乱項について，ALS(2)推定量がスペクトラム推定量よりもより小さい平均二乗誤差をもたらすことを示している．

上記の結果は，ノン・パラメトリック回帰分析は，とくにそれが周波数領域で行なわれる場合には，いつも推奨できるわけではないということを意味している．ここで多少なりとも妥協するとすれば，AIC のような簡単な当てはまりの基準によって選ばれる次数 $P$ に基づく ALS 法を採用することであろう．この方法によればモデル化と推定は半自動的なやり方として組み込むことができ，適切な ARMA 過程を同定するという作業をはぶくことができる．

ノート

**7.2** Harvey=Phillips (1979) を参照せよ.

**7.3** *The Annals of Economic and Social Measurement* (1973) 所収の論文, Harvey (1978 b) 及び Pagan (1980) を参照せよ. 定常性への回帰モデルの最尤推定量および二段階推定量の性質についてのモンテ・カルロ研究が, Harvey=Phillips (1981) に報告されている.

**7.4** Hannan (1963) 及び Harvey (1978 a) を参照せよ.

**7.5** Akaike (1969), Amemiya (1973) を参照せよ.

## 練習問題

**1.** MA (1) 型の攪乱項を持つ回帰モデルにおいて, 条件付き二乗和のアルゴリズムを用いて, どのようにして近似的最尤推定値を得るのか説明せよ. 一階の微係数を計算するための漸化式の表現を書き表わせ.

推定量の大標本共分散行列を導け.

**2.** 説明変数についての適当な値が与えられたとして, ARMA (1, 1) に従う攪乱項を持つ回帰モデルにおける従属変数の将来の値の予測をいかに行なうか説明せよ. このモデルにおいて, 推定された母数に条件付予測の平均二乗誤差を計算する表現を書き表わせ.

**3.** ARMA (1, 1) に従う攪乱項を持つ回帰モデルにおいて, 母数 $\phi$ および $\theta$ が既知であると想定しよう. モデルの状態空間表現を求めることにより, いかに $\beta$ の一般化最小二乗推定量が計算されるか示せ. 対応する $x_t$ の値が与えられたときに, $y_t$ の将来の値の予測を行なうに当って, 状態空間モデルをどのように用いることができるかを説明せよ. ここで $\beta$ が推定されなければならないときに, 予測の信頼区間は厳密なものであろうか.

**4.** 緩やかに変化する季節要素を伴った回帰モデルは, 以下の形式に書けるとしよう.

$$y_t = s_t + x_t'\beta + \epsilon_t$$
$$s_t = s_{t-s} + \eta_t, \qquad t = 1, \cdots, T.$$

ここで $\epsilon_t$ と $\eta_t$ は独立に正規分布をするホワイト・ノイズ過程である. このようなモデルをいかに推定するか説明せよ. そして適当な $x_t$ の値を知っているという状況下で, $y_t$ の将来の値の予測をどのように作るか説明せよ. もし $s_t$ が確率的でなく決定論的である場合には, 推定及び予測に対しての以上の結論はどのようになるのであろうか.



**5.** 観測値が以下のようなモデルから生成されたと想定しよう.

$$y_t = \alpha + \beta x_t + \epsilon_t, \qquad t = 1, \cdots, T.$$

ここで $\alpha$ はゼロでない母数である. もしモデルを誤って定数項がないモデルとして定式化した場合には, $\beta$ の最小二乗推定量が偏りを持つことを示せ. もし $\beta$ が (3.11 b) によって示唆されたようなとき, 変母数であると想定され, さらに再び定数項なしと想定されたモデルが推定されたならば, 予測は前よりもよくなるであろうか. ここでの答は定式化の誤りのあるモデル一般について何らかの示唆を与えるであろうか.

# 文　　献

Akaike, H. (1969), Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics*, 21, pp. 243–247.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, AC-19, pp. 716–723.

Amemiya, T. and Wu, R. Y. (1972), The effect of aggregation on prediction in the autoregressive model, *Journal of the American Statistical Association*, 67, pp. 628–632.

Amemiya, T. (1973), Generalized least squares with an estimated autocovariance matrix, *Econometrica*, 41, pp. 723–732.

Anderson, B. D. O. and Moore, J. B. (1979), *Optimal Filtering*, Prentice-Hall.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, John Wiley, New York.

Ansley, C. F. (1979), An algorithm for the exact likelihood of a mixed autoregressive moving average process, *Biometrika*, 66, pp. 59–65.

Ansley, C. F. and Newbold, P. (1980), Finite sample properties of estimators for autoregressive moving average models, *Journal of Econometrics*, 13, pp. 159–184.

Baillie, R. T. (1979), Asymptotic prediction mean squared error for vector autoregressive models, *Biometrika*, 66, pp. 675–678.

Beach, C. M. and MacKinnon, J. G. (1978), A maximum likelihood procedure for regression with auto-correlated errors, *Econometrica*, 46, pp. 51–58.

Box, G. E. P. and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, revised edition, Holden-Day, San Francisco.

Box, G. E. P. and Pierce, D. A. (1970), Distribution of residual autocorrelations in autoregressive integrated moving average time series models, *Journal of the American Statistical Association*, 65, pp. 1509–1526.

Brewer, K. R. W. (1973), Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models, *Journal of Econometrics*, 1, pp. 133–154.

Burman, J. P. (1980), Seasonal adjustment by signal extraction, *Journal of the Royal Statistical Society, Series A*, 143, pp. 321–337.

Chan, W. Y. T. and Wallis, K. F. (1978), Multiple time series modelling:

another look at the mink–muskrat interaction, *Applied Statistics*, 27, pp. 168–175.

Cleveland, W. S. and Tiao, G. C. (1976), Decomposition of seasonal time series: a model for the census X-11 program, *Journal of the American Statistical Association*, 71, pp. 581–587.

Cooley, T. F. and Prescott, E. C. (1976), Estimation in the presence of stochastic parameter variation, *Econometrica*, 44, 167–184.

Crowder, M. H. (1976), Maximum likelihood estimation for dependent observations, *Journal of the Royal Statistical Society, Series B*, 38, pp. 45–53.

Davidson, J. (1981), Problems with the estimation of moving average processes, *Journal of Econometrics* (forthcoming).

Dent, W. T. and Min, A. S. (1978), A Monte Carlo study of autoregressive integrated moving average processes, *Journal of Econometrics*, 7, pp. 23–55.

Duncan, D. B. and Horn, S. D. (1972), Linear dynamic regression from the viewpoint of regression analysis, *Journal of the American Statistical Association*, 67, pp. 815–821.

Durbin, J. (1969), Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals, *Biometrika*, 56, pp. 1–15.

Engle, R. F. (1974a), Band spectrum regression, *International Economic Review*, 15, pp. 1–11.

Engle, R. F. (1974b), Specification of the disturbance term for efficient estimation, *Econometrica*, 42, pp. 135–146.

Engle, R. F. (1978a), Estimating structural models of seasonality (with discussion), in A. Zellner (ed.), *Seasonal Analysis of Economic Time Series*, Bureau of the Census, Washington DC, pp. 281–308.

Engle, R. F. (1978b), Testing price equations for stability across spectral frequency bands, *Econometrica*, 46, pp. 869–881.

Engle, R. F. and Gardner, R. (1976), Some finite sample properties of spectral estimators of a linear regression, *Econometrica*, 44, pp. 149–166.

Fishman, G. S. (1969), *Spectral Methods in Econometrics*, Harvard University Press.

Fuller, W. A. (1976), *Introduction to Statistical Time Series*, John Wiley, New York.

Fuller, W. A. and Hasza, D. P. (1978), Properties of prediction for autoregressive time series, Dept. of Statistics, Iowa State University (unpublished paper).

Galbraith, R. F. and Galbraith, J. F. (1974), On the inverse of some patterned matrices arising in the theory of stationary time series, *Journal of Applied Probability*, 11, pp. 63–71.

Gallant, A. R. and Goebel, J. J. (1976), Nonlinear regression with autocorrelated errors, *Journal of the American Statistical Association*, 7, pp. 961–967.

Garbade, K. (1977), Two methods for examining the stability of regression coefficients, *Journal of the American Statistical Association*, 72, pp. 54–63.

Gardner, G., Harvey, A. C. and Phillips, G. D. A. (1980), An algorithm for exact maximum likelihood estimation by means of Kalman filtering, *Applied Statistics*, 29, 311–322.

Gilchrist, W. (1976), *Statistical Forecasting*, John Wiley, New York.



Godfrey, L. G. (1978), Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables, *Econometrica*, 6, pp. 1293–1302.

Godolphin, E. J. (1976), On the Cramér–Wold factorization, *Biometrika*, 63, pp. 367–380.

Goldberg, S. (1958), *Difference Equations*, John Wiley, New York.

Granger, C. W. J. and Hatanaka, M. (1964), *Spectral Analysis of Economic Time Series*, Princeton University Press.

Granger, C. W. J. and Morris, M. J. (1976), Time series modelling and interpretation, *Journal of the Royal Statistical Society, Series A*, 139, pp. 246–257.

Granger, C. W. J. and Newbold, P. (1977), *Forecasting Economic Time Series*, Academic Press, New York.

Griffiths, L. J. and Prieto-Diaz, R. (1977), Spectral analysis of natural seismic events using autoregressive techniques, *IEEE Transactions on Geo-Science Electronics*, GE-15, pp. 13–25.

Hannan, E. J. (1963), Regression for time series, in M. Rosenblatt (ed.), *Time Series Analysis*, John Wiley, New York, pp. 14–37.

Hannan, E. J. (1969), The identification of vector mixed autoregressive moving average systems, *Biometrika*, 56, pp. 223–225.

Hannan, E. J. (1970), *Multiple Time Series*, John Wiley, New York.

Hannan, E. J. (1980), The estimation of the order of an ARMA process, *Annals of Statistics*, 8, pp. 1071–1081.

Harrison, P. J. (1967), Exponential smoothing and short-term sales forecasting, *Management Science*, 13, pp. 821–842.

Harrison, P. J. and Stevens, C. F. (1976), Bayesian forecasting (with discussion), *Journal of the Royal Statistical Society, Series B*, 38, pp. 205–247.

Hart, B. I. (1942), Significance levels for the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics*, 13, pp. 445–447.

Harvey, A. C. (1978a), Linear regression in the frequency domain, *International Economic Review*, 19, pp. 507–512.

Harvey, A. C. (1978b), The estimation of time-varying parameters from panel data, *Annales de l'INSEE*, 30–31, pp. 203–226.

Harvey, A. C. (1980), On comparing regression models in levels and first differences, *International Economic Review*, 21, pp. 707–720.

Harvey, A. C. (1981a), *The Econometric Analysis of Time Series*, Philip Allan.

Harvey, A. C. (1981b), *Finite Sample Prediction and Overdifferencing*, Discussion Paper, London School of Economics.

Harvey, A. C. and McAvinchey, I. D. (1981), On the relative efficiency of various estimators of regression models with moving average disturbances, in G. Charatsis (ed.), *Proceedings of European Meeting of the Econometric Society*, Athens, 1979.

Harvey, A. C. and Pereira, P. L. V. (1980), The estimation of dynamic models with missing observations, Paper presented to *World Congress of the Econometric Society*, Aix-en-Provence, August 1980.

Harvey, A. C. and Phillips, G. D. A. (1976a), The maximum likelihood

estimation of autoregressive-moving average models by Kalman filtering, University of Kent (unpublished paper).

Harvey, A. C. and Phillips, G. D. A. (1976b), Testing for stochastic parameters in regression, Paper presented at the *European Meeting of the Econometric Society*, Helsinki, August 1976.

Harvey, A. C. and Phillips, G. D. A. (1977), A comparison of estimators in the ARMA(1, 1) model, University of Kent (unpublished paper).

Harvey, A. C. and Phillips, G. D. A. (1979), The estimation of regression models with autoregressive-moving average disturbances, *Biometrika*, 66, pp. 49–58.

Harvey, A. C. and Phillips, G. D. A. (1981), The estimation of regression models with time-varying parameters, in *Proceedings of a Symposium in Honour of Oskar Morgenstern*, Vienna, May 1980.

Hasza, D. P. (1980), The asymptotic distribution of the sample autocorrelations for an integrated ARMA process, *Journal of the American Statistical Association*, 75, pp. 349–352.

Hosking, J. R. M. (1980), The multivariate portmanteau statistic, *Journal of the American Statistical Association*, 75, pp. 602–608.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Jones, R. H. (1980), Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics*, 22, pp. 389–395.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Transactions ASME Journal of Basic Engineering*, 82, pp. 35–45.

Kang, K. M. (1975), A comparison of estimators for moving average processes, Australian Bureau of Statistics (unpublished paper).

Kendall, M. G. (1973), *Time Series*, Griffin.

Klein, L. R. (1950), *Economic Fluctuations in the United States, 1921–41*, John Wiley, New York.

Kuznets, S. S. (1961), *Capital and the American Economy: Its Formation and Financing*, National Bureau of Economic Research, New York.

Leskinen, E. and Terasvirta, T. (1976), Forecasting the consumption of alcoholic beverages in Finland, *European Economic Review*, 8, pp. 349–369.

Ljung, G. M. and Box, G. E. P. (1978), On a measure of lack of fit in time series models, *Biometrika*, 66, pp. 297–303.

Nelson, C. R. (1972), The prediction performance of the FRB–MIT–PENN model of the U.S. Economy, *American Economic Review*, 62, pp. 902–917.

Nelson, C. R. (1974), The first order moving average process, *Journal of Econometrics*, 2, pp. 121–141.

Nelson, C. R. (1976), Gains in efficiency from joint estimation of systems of autoregressive-moving average processes, *Journal of Econometrics*, 4, pp. 331–348.

Nelson, C. R. and Shea, G. S. (1979), Hypothesis testing based on goodness-of-fit in the moving average time series model, *Journal of Econometrics*, 10, pp. 221–226.

Nerlove, M., Grether, D. M. and Carvalho, J. L. (1979), *Analysis of Economic Time Series*, Academic Press, New York.

Nicholls, D. F. and Pagan, A. R. (1977), Specification of the disturbance for



efficient estimation – an extended analysis, *Econometrica*, 45, pp. 211–217.

Osborn, D. R. (1976), Maximum likelihood estimation of moving average processes, *Annals of Economic and Social Measurement*, 5, pp. 75–87.

Pagan, A. (1980), Some identification and estimation results for regression models with stochastically varying parameters, *Journal of Econometrics*, 13, pp. 341–363.

Parzen, E. (1969), Multiple time series modelling, in P. R. Krishaiah (ed.), *Multivariate Analysis, Vol. II*, Academic Press, New York, pp. 389–409.

Poskitt, D. S. and Tremayne, A. R. (1980), Testing the specification of a fitted autoregressive-moving average model, *Biometrika*, 67, pp. 359–363.

Poskitt, D. S. and Tremayne, A. R. (1981), *A Time Series Application of th Use of Monte Carlo Methods to Compare Statistical Tests*, Discussion Paper, University of York.

Rosenberg, B. (1973), Random coefficient models: the analysis of a cross section of time series by stochastically convergent parameter regression, *Annals of Economic and Social Measurement*, 2, pp. 399–428.

Roy, R. (1977), On the asymptotic behaviour of the sample autocovariance function for an integrated moving average process, *Biometrika*, 64, pp. 419–42

Salmon, M. (1980), Structural constancy in model selection, Paper presented to *World Congress of the Econometric Society*, Aix-en-Provence, August 1980.

Sargan, J. D. and Bhargava, A. (1980), *Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle*, Discussion paper, London School of Economics.

Schaefer, S. *et al.* (1975), Alternative models of systematic risk, in E. Elton and M. Gruber (eds), *International Capital Markets: An Inter and Intra Country Analysis*, North-Holland Publishing Co., pp. 150–161.

Schweppe, F. (1965), Evaluation of likelihood functions for Gaussian signals, *IEEE Transactions on Information Theory*, 11, pp. 61–70.

Shenton, L. R. and Johnson, W. L. (1965), Moments of a serial correlation coefficient, *Journal of the Royal Statistical Society, Series B*, 27, pp. 308–320.

Silvey, S. D. (1970), *Statistical Inference*, Chapman and Hall.

Theil, H. (1971), *Principles of Econometrics*, John Wiley, New York.

Theil, H. and Goldberger, A. S. (1961), On pure and mixed statistical estimation in economics, *International Economic Review*, 2, pp. 65–78.

Wallis, K. F. (1974), Seasonal adjustment and relations between variables, *Journal of the American Statistical Association*, 69, pp. 18–31.

Wilson, G. T. (1973), The estimation of parameters in multivariate time series models, *Journal of the Royal Statistical Society, Series B*, 35, pp. 76–85.

Yamamoto, T. (1976), A note on the asymptotic mean square error of predicting more than one step ahead using the regression method, *Applied Statistics*, 25, pp. 123–127.

# 索　　引


### ア行

赤池の情報量基準（AIC）　21, 90, 136, 163, 189
ARIMA　→自己回帰和分平均移動
ARMA　→自己回帰平均移動
位相（phase）　60, 77, 78, 93, 94
位相図（phase diagram）　78
1次差分作用素(first difference operator)　28
一様最強力不偏検定　154
一致推定量　18, 88, 131
一般化回帰モデル　109
一般化最小二乗（CLS）推定量　13, 106, 216, 217
一般化線型回帰モデル　13, 213
移動平均　71, 76
移動平均過程　36, 71, 127
　1次の——　23
　季節自己回帰——　177
　$q$次——　36
　無限次——　48
　ベクトル値自己回帰——　48
移動平均フィルター　76, 81
移動平均母数　42
入れ子（nested）型　136
　非——（non-nested）　20, 136
入れ子型仮説（nested hypothesis）　20
ウィンドウ（窓）　88
　矩形——　88
　スペクトラム・——　90
　——の幅　88
　ラグ・——　89, 90
ウォルド（Wold）の分解定理　70
AIC　→赤池の情報量基準
AIC統計量　163
ALS法　219
X-11法　181
エルゴード性（ergodic）　25
帯幅（bandwidth）　88

### カ行

階差（差分）　151
$\chi^2$分布　19, 146
ガウス=ニュートン（Gauss=Newton）法　10, 18
ガウス=マルコフ（Gauss=Markov）定理　107
　拡張——　107
撹乱項　4
確率過程（stochastic process）　3, 23
確率的係数モデル　206
確率変数　23
過剰階差　185, 186
過剰母数　158
仮説検定　136
片側検定　157
片側対立仮説　154





偏り　140
可変的季節パターン　179
加法（additive）モデル　175, 178
カルマン・フィルター　9, 104, 109, 112, 131, 134, 201
カルマン・ゲイン　102
頑健性　123
観測不能な要素（unobservable components）　9, 42, 104, 115
観測方程式（measurement equation）　103
幾何級数分布ラグモデル　96
季節ARMA過程　177
季節階差（seasonal differencing）　180, 181
季節性　77, 176
季節調整　77, 82, 181
　——フィルター　82
季節要素　176
規則的循環　55
疑似循環変動（pseudo-cyclical behavior）　74
季調済み系列　181
逆予測（backcasting）　130
共通因子（common factor）　42
共分散行列　46
　自己——　46
　漸近——　135
行列の逆転公式　111, 120
クラメール表現（cramérs' representation）　68
計量経済モデル　1
系列相関　2, 154, 200
系列間の位相関係　190
ゲイン（gain）　77, 81, 93, 94
ケチの原理（principle of parsimony）　48, 158, 163
欠足観測値　42, 44
欠足観測値（missing observation）問題　43
系列相関　200
検出力　154
減衰的周期変動　35
倹約的な表現（parsimonious representation）　3
コイック（Koyck）型　96
更新方程式（updating equation）　104, 109, 110, 112
更新問題　104
高速フーリエ変換　91
固定的季節パターン　180
コヒーレンス（choerence）　93, 95
　標本——　97
コレスキー（Cholesky）分解　15
コレログラム（correlogram）　27, 126, 153, 159 187
混合過程　38, 75, 127
混合推定法（mixed estimation）　107

### サ行

最小二乗推定量　64
　加重——　207
最小二乗法（OLS）　4
最小分散不偏推定量（MVUE）　12
最小平均二乗推定　106
最小平均二乗推定量（MMSE）　14, 163, 165
最小平均二乗線型推定量（MMSLE）　104, 107
最適予測　14
　——量　104, 110
最尤推定　10, 113, 123
　厳密——　125, 130
最尤推定量　10
最尤法　10
最良線型不偏推定量（BLUE）　12, 106
雑音（noise）　103
三角関数　58
三角関数多項式　100
残差　161
　一般化逐次——　202
　逐次——　198
　直接——　118
　予測誤差からの——　118

残差二乗和関数 123
残差の診断 (diagonostic checking) 7
時間集計 (temporal aggregation) 43, 45
時間領域 1, 40, 89
識別性 (identification) 51
時系列
一変量—— 8
——回帰分析 9, 197
集計—— 45
多変量—— 8, 45
——データ 1
——モデル 1
自己回帰移動平均 (ARMA) 3
——モデル 7, 9, 151
——過程 4, 8, 29, 71
自己回帰過程 30, 73
ベクトル値—— 145
自己回帰行列 145
自己回帰スペクトラム推定 92
自己回帰的最終型 (autoregressive final form) 52
自己回帰母数 42
自己回帰モデル 2
1次—— 16, 30
時変係数—— 205
2次—— 31
$p$ 次—— 33
無限次—— 37
自己回帰和分移動平均 (ARIMA) 7
——過程 7, 71, 151
——モデル 7, 9
自己共分散 24
自己共分散関数 (autocovariance funcion) 26, 34, 37, 40
自己相関関数 (autocorrelation function) 26, 35, 37, 40, 126
自己相関行列 (autocorrelation matrix) 46
事後分布 109
二乗和関数 125
指数型加重移動平均 (EWMA) 173
——予測 173
事前情報 108
事前分布 109
実現系列 (realization) 24
実数確率過程 57
時変母数 (time-varying parameter) 104
周期 (period) 59
周期的関数 59
周期的変動 1
周期的要素 54
集計問題 42
周波数 59
——応答関数 (frequency response function) 73, 75, 79, 83
季節—— 82, 181
基本—— 80
高—— 58
調和—— 65, 80
低——領域 58
ナイキスト (Nyquist) —— 61, 80
——領域 1, 53, 156, 213
循環過程 (circular) 216
循環変動 36
峻 (スパイク) 71, 74
条件付確率 14
条件付二乗和(CSS)関数 9, 128, 134
状態空間形 104, 201
状態空間モデル 9, 103
状態変数 (state variables) 103
小標本特性 18
乗法型モデル 178
乗法的季節 ARMA 過程 178
乗法的季節 ARIMA 過程 182
情報量行列 134
初期推定値 131
振幅 59
信頼区間 136, 139
推定 123, 151
推定効率 132
随伴多項式 (associated polynomials) 29, 75, 178
酔歩 (ランダム・ウォーク)
——係数回帰モデル 209



——過程 115, 118, 171
——母数係数 209
スコアリング法 17
スペクトグラム (spectrogram) 63
スペクトラム 56, 74
帯状——回帰分析 213
——回帰分析 213
クォドラチュア・—— 96
コ・—— 96
相互—— 93, 96
多変量——行列 93
——の推定 85
パワー・—— 56, 69, 74
標本相互—— 215
標本——密度 85, 86, 91, 215
標本コ・—— 97
標本クォドラチュア・—— 97
——分布関数 70
——密度 57
——密度関数 84
有理—— 75
離散—— 70
スペクトラム表現
確率過程の—— 66, 68
複素—— 69
スペクトラム分析 9, 55
相互—— 55
多変量—— 93
正規分布 12, 123
正則条件 18
制御工学 103
正弦 (サイン) 関数 59
正の系列相関 204
制約条件 19
制約付非線型探索法 140
遷移方程式 (transition equation) 103
漸近的 $t$ 比 137
漸近的 $F$ 比 137
漸近的特性 134
漸近的に有効 11
漸近的情報行列 11
漸近的標準誤差 11
漸近分布 127
線型回帰 11
線型過程 29
線型時間不変 (linear time invariant) フィルター 71
線型動学システム 104
線型フィルター 55, 76, 77, 79, 82, 94
非—— 77
線型推定量 12
相関 97
相互共分散 (cross covariances) 46
相互相関関数 47

タ行

対数尤度関数16
大標本特性 123
多期間予測 116
一般—— 116
ダービン=ワトソン検定 204
多変量 ARMA $(p, q)$ 過程 48
多変量最小二乗推定量 146
多変量自己回帰移動平均過程 50
多変量正規分布 13, 145
多変量風呂敷統計量 (mulitvariate portmanteau statistics) 191
多変量ホワイト・ノイズ過程 48
多変量モデル 189
ダミー変数 66, 179
逐次最小二乗法 197
長期景気変動仮説 84
長期的循環変動 84
頂 (ピーク) 71, 74, 80, 89
直交関係 99
定常 (stationary) 3, 24
——ARMA 過程 4, 42
——解 119
——過程 3
——季節モデル 178
——時系列 3
同時——(jointly stationary) 46
定常性(stationality) 24, 30, 38, 46, 50
強—— 25

共分散—— 25
弱—— 25
——の条件 33
非—— 168
定常性への再帰モデル 205, 207
適合度 163
——の尺度 163
伝達関数 (transfer function) 72, 83
同時方程式モデル体系 8
トエプリッツ (Toeplitz) 行列 216
独立性検定 152
トレンド 4
確率的—— 175
局所的—— 5, 170
——項と誤差項の加法モデル 170
周期的—— 58
周期的——要素 55
循環的—— 64
大域的—— 170
大域的——モデル 6
——付酔歩 172
——付酔歩過程 183
非確率的—— 175

ナ行

二段階推定法 208
二段階推計量 18
ニュートン=ラフソン (Newton=Raphson) 法 18
ネイマン=ピアソン (Neyman=Pearson) 補題 19
ノン・パラメトリック 218
——回帰分析 219

ハ行

パルゼン (Parzen) 法 92
反転可能条件 133
反転可能性 37, 38, 49
非確率的季節要素 179
非決定論的過程 29
非独立性 151
標本自己共分散 27
標本自己相関関数 27
標本相関関数 152
標本分散 27
標本平均 27
標本偏自己相関関数 159, 160
ピリオドグラム (periodogram) 58, 63, 85, 215
累積—— 157, 161, 203
累積——法 204, 215
累積——検定 156
フィッシャー情報行列 206
フィルター 71
フォン・ノイマン比 153
不規則循環 55
複素フーリエ変換 57
不偏推定量 86
フーリエ解析 55, 60
フーリエ逆変換 70
フーリエ係数 64
フーリエ表現 61
フーリエ変換 100
風呂敷 (portmanteau) 検定 154
風呂敷統計量 154-56, 161, 203
多変量—— 191
平滑化 (smoothing) 76, 117
——推定量 117
——定数 6
——方程式 117
——問題 105
平均 (ensemble) 24
平均二乗誤差基準 142
ベイズ的 (Bayesian) 109
ヘッセ (Hessian) 行列 18
ペナルティ (罰金) 163
母数過剰 53
母数の識別問題 38
母数の推定 10
ボックス=コックス(Box=Cox)変換 170
ボックス=ジェンキンズ 7, 151, 158
ボックス=ジェンキンズ法 151
ボックスの計算法 140
ボックス=ピアス (Box=Pierce) 161





——検定 161
——修正 162
ボックス=リュング (Box=Ljung) 検定 162
ホルト=ウィンターズ (Holt=Winters) の方法 6, 176
ホワイト・ノイズ 4
ホワイト・ノイズ過程 68, 157

マ行

マン=ワルド (Man=Wald) の定理 134
見かけ上無関係なベクトル値 ARMA 過程 148
見かけ上無関係な回帰方程式 148
見せかけの周期変動 83
無条件不偏 ($u$ 一不偏) 105
モデル識別 151, 193
モデル選択 157, 187
モデルの診断 21, 151, 189
漏れ (leakage) 91
モンテ・カルロ (Monte Carlo) 実験 141, 219

ヤ行

有意水準 153
有効推定量 10
尤度関数 15, 17, 113
集約対数—— 203
条件付—— 124, 127, 128
対数—— 10, 16, 203
尤度比検定 18, 138
尤度比検定統計量 138
尤度比統計量 147
尤度方程式 (likelihood equations) 10
ユール・ウォーカー (Yule=Walker) 方程式 126
ユール=スルツキー (Yule= Sutsky) 効果 83, 84
余弦 (コサイン) 関数 59
予測 2, 163, 193
一期先——量 168, 173
最適—— 165
線型——量 165
多段—— 168
有限標本——(finite sample prediction) 164, 166, 167
予測関数 164, 172
随伴——(eventual forecasting function) 173
予測区間 175
予測誤差 105, 112
——の分解 10, 13, 124
——の分散 14
予測方程式 (prediction equation) 104, 109, 110
予測問題 104

ラ行

ラグ作用素 28
ラグ変数 3
ラグランジュ乗数 (Lagrange multiplier) 検定 19, 20, 138, 155
ラジアン (角速度) 32, 59
両側検定 204

ワ行

和の定理 43
和分過程 170
割引最小二乗推定量 (DLSE) 199
ワルド検定 (Wald Test) 18, 19, 137, 146

**訳者略歴**

国友直人

1950年　東京に生れる
1975年　東京大学経済学部卒業
1981年　スタンフォード大学 Ph. D(経済学)
現　在　東京大学経済学部助教授
(〒113　東京都文京区本郷7-3-1・東京大学)

山本　拓

1945年　東京に生れる
1968年　慶応義塾大学工学部卒業
1974年　ペンシルベニア大学 Ph. D(経済学)
現　在　横浜国立大学経済学部助教授
(〒240　神奈川県横浜市保土ヶ谷区常盤台156・横浜国立大学)

時系列モデル入門

1985年9月15日　初　版
1988年10月15日　2　刷

[検印廃止]

訳　者　国友直人
山本　拓　©

発行所　財団法人　東京大学出版会
代表者　菅野卓雄
113 東京都文京区本郷7-3-1 東大構内
電話 (811)8814・振替東京6-59964
組　版　株式会社永昌美術
印　刷　三栄印刷株式会社
製　本　永沢製本所

ISBN 4-13-042022-4

| | | |
|---|---|---|
| 統計入門 | 中村隆英他 | A5判・290頁<br>2200円 |
| 経済統計入門 | 中村隆英他 | A5判・310頁<br>2200円 |
| ベイズ統計入門 | 繁桝算男 | A5判・240頁<br>2800円 |
| 計量経済学の新展開 | 竹内啓編 | A5判・290頁<br>2800円 |
| マクロ経済学研究 | 吉川洋 | A5判・300頁<br>3600円 |
| マクロ経済学と経済政策 | 貝塚啓明他編 | A5判・230頁<br>2400円 |
| 最小二乗法による実験データ解析 | 中川徹<br>小柳義夫 | A5判・230頁<br>2400円 |
| 正規分布特性と応用 | 柴田義貞 | A5判・290頁<br>2800円 |
| 物理現象のフーリエ解析 | 小出昭一郎 | A5判・180頁<br>2400円 |