新版へのまえがき

『時系列解析入門』を出版してから 15 年が経過した．この間，データ環境，計算環境，データ活用技術は飛躍的に発展し，データ駆動型の社会が実現しようとしている．数理・データサイエンス・AI を現代社会におけるリテラシーと捉え，専門分野によらずすべての学生を対象にした教育も始まろうとしている．このようなデータサイエンスの発展の背景として，計算環境の著しい変化がある．現在では，ネット上に無料で利用できる関数が大量に存在し，データ解析を試みる人は R や Python などの適当な関数を探すことによって大規模データの解析さえもできるようになっている．このような時代背景を考慮して，本書の改訂にあたっては R による解析例を取り入れる形にした．

本書は初版以来，読者が個別の問題に対応して自らモデルを考案し，それを実装できるようになることを目標としてきた．今後の時系列解析の発展にはそのような能力を持った人の育成は不可欠であるが，それに加えて，既存のソフトを使って手早く実際のデータに適用してみることができるようにすることも重要である．幸い，著者が以前在籍した統計数理研究所において本書初版の前身である『FORTRAN77 時系列解析プログラミング』に掲載されていた FORTRAN プログラムをベースに，統計計算言語 R のパッケージ TSSS (Time Series analysis with State Space model) が開発された．TSSS の利用によって本書で取り上げたほとんどのモデリングの方法や計算法は簡単に実現できるようになった．そこで，本書ではほとんどの例題を TSSS パッケージの関数で計算するとともに，その利用法についても説明するようにした．今回全面的に採用した R 関数を活用し，方法の習得と並行して実際のデータ解析を行うことによって，時系列解析の理解が一段と進むことを期待している．

改訂にあたっては，このほか ARMA モデルの次数を自動的に探索する方法や，乱数生成法にメルセンヌ・ツイスタを加えるなどいくつかの修正を行った．また第 15 章のタイトルをモンテカルロ・フィルタから粒子フィルタに変更した．著者はモンテカルロ・フィルタという名称を提案してきたが，この方

法の開発から四半世紀を経て，粒子フィルタという名称が世界的に定着したからである．

この新版への改訂は TSSS パッケージの開発によってはじめて実現できたものであり，開発にあたられた嵯峨優美さんと中野純司教授に心からお礼を申し述べたい．またこの新版の出版は岩波書店編集局自然科学書編集部の濱門麻美子さんと高野照子さんの提案によるものであり，出版に至るまでにお世話になったお二人に感謝する．実はこの改訂作業は英訳版のほうが先行して行われた．初版の英訳版は 2010 年に Chapman & Hall/CRC から出版されていたが，第 2 版の出版を機に R による例題を取り込むことになった．英訳版第 2 版の出版を提案いただいた Editor の David Grubbs 氏とこの英訳版初版・第 2 版出版にあたってお世話になった岩波書店編集局ライツマネジメント部の森川裕美さんにも感謝する．

2020 年 10 月，東京

北川源四郎

初版まえがき

21 世紀の現代社会は，地球環境，経済，安全性などさまざまな面で深刻な問題を抱えている．これらの問題の多くは，複雑・不確実そしてダイナミックな現象に関するものであり，そのような複雑なシステムの解明，予測そして情報抽出のための方法が求められている．一方，統計科学の分野では，情報量規準 AIC の導入を契機として，統計的モデリングの方法が急速な発展をとげ，大量データの蓄積と計算機能力の飛躍的な発展とも相まって，数多くの新しいモデルが開発され，実用化されてきた．

本書は，このように急速に進歩をとげつつある統計的モデリングのなかでも，とりわけ代表的な時系列モデリングの方法について解説することを目的としている．時系列解析は，ある現象の変動を過去の動きとの関連で捉えようとするものである．統計科学が対象とする多くの現象が時系列であることからも，またこのような現象の捉え方が，統計的モデリングの本質である条件つき分布の構成の代表例であることからも，時系列解析の方法を学ぶことは重要である．本書では，時系列解析に用いられる主要なモデルを紹介しているが，とくに状態空間モデルに基づく統一的な方法を説明していることが特徴の一つである．また本書では，時系列のモデリングに必要な，最小二乗法，最尤法，逐次フィルタによる効率的な推定の方法を示すとともに，情報量規準 AIC に基づく一貫したモデル評価・選択の方法を紹介している．

なお，本書は，1993 年に「岩波コンピュータサイエンス」シリーズの一つとして刊行された『FORTRAN77 時系列解析プログラミング』をベースとし，改訂を加えた入門書であることを注記しておきたい．前著の刊行から十余年が経過する間，統計科学が確実に前進をとげてきたことと並行して，情報処理のインフラとも言うべきコンピュータの劇的な発展・普及という大きな変化があったことはご承知の通りである．あらゆる環境にパーソナルコンピュータが進出する一方で，多くの分野で大量データが利用可能となったことから，それらの資源を活用し，データを適切にモデリングし，適切な情報を取り出す手法が以前にも増して求められている．

その一方で，実際の解析にあたっては，かつてのように大型計算機上でFORTRANプログラムを動かすというケースよりは，さまざまなプラットフォームにおいて，多種多様なプログラム言語による解析が行われる場合が増えてきた．したがって，本書ではFORTRANのソースコードとその説明部分を削除してコンパクトにする一方で，モンテカルロ・フィルタに関する新たな章を設けることによって，より自由なモデリングを目指す新しい時代の要請にもできる限り対応可能なように心掛けたつもりである．ただし，新版で削除した計算プログラムとその説明は，Webサイト(http://www.ism.ac.jp/~kitagawa/)からダウンロードできるように配慮した．また，産業技術総合研究所の松本則夫博士はこれらのプログラムをWeb上で実行できる解析システムを開発し，次のサイトで公開している(http://www.aist.go.jp/RIODB/gxwell/GSJ/analysis/index.html)．これらの利用によって，本書で取り上げたモデルや解析法のほとんどすべてを読者自身のデータでためしてみることができる[1)]．また，本書では，新たに各章末ごとに幾つかの問題を設けた．これらが，本文で取り上げた代表的なモデルを理解する上で，読者にとって助けとなれば幸いである．

最後になるが，本書に至るまでの時系列解析の研究において，著者はこれまで多くの方々のお世話になった．統計数理研究所の赤池弘次元所長を始め，多くの諸先輩同僚の方々には，日ごろの議論や会話を通して多大の影響を受けたことはいうまでもない．また，本書で取り上げた非定常モデルの多くは，ハワイ大学のGersch教授，東京海洋大学の大津皓平氏，北海道大学の高波鐵夫氏，産業技術総合研究所の松本則夫氏はじめ多くの方々との共同研究の成果として得られたものである．また，これらの方々には本書の例題として使用した貴重なデータを快く提供していただいた．本書の原稿の作成および計算にあたっては福島靖子さんと小野節子さんのお世話になった．本書の出版にあたっては，岩波書店編集部の吉田宇一氏と松永研氏にはひとかたならぬお世話になった．これらの方々には心からのお礼を申し述べたい．

2005年1月

北川源四郎

1） 計算環境が大幅に変わり，上記の2つのサイトは利用できなくなっているので，計算ソフトに興味がある読者は新版で採用したRの関数の利用をお願いしたい．(2020年10月)

R と時系列解析パッケージ TSSS

R はオープンソースの統計計算と可視化のためのプログラム言語および開発実行環境で，Linux，Mac あるいは Windows で動かすことができる．本書の例題で用いているデータと解析プログラムの多くは R のパッケージ TSSS (Time Series analysis with State Space model) に含まれている．

R の取得

R は，下記の CRAN (The Comprehensive R Archive Network) のウェブページにおいて Linux，(Mac) OS X あるいは Windows のいずれかの OS を選択して取得できる．

https://cran.r-project.org/
https://cran.ism.ac.jp/ (日本のミラーサイト)

R の時系列解析パッケージ TSSS

TSSS は統計数理研究所において嵯峨優美氏および中野純司教授によって，本書初版の前身である『FORTRAN77 時系列解析プログラミング』に掲載されていた FORTRAN プログラムを基礎に開発されたものである．

TSSS パッケージも CRAN からダウンロードできるが，統計数理研究所の下記のサイトには R も含め，インストールの詳しい説明がある．

https://jasp.ism.ac.jp/ism/TSSS/

TSSS には以下の関数が用意されている．

`arfit`	1 変量 AR モデルのあてはめ
`armachar`	1 変量 ARMA モデルによる特徴抽出(インパルス応答関数，自己共分散関数，PARCOR，スペクトル，特性根の計算)
`armafit`	1 変量 ARMA モデルのあてはめ
`armafit2`	多数の 1 変量 ARMA モデルの自動的推定
`boxcox`	Box-Cox 変換の計算
`crscor`	相互共分散関数と相互相関関数の計算
`fftper`	FFT によるピリオドグラムの計算
`klinfo`	K-L 情報量の計算
`lsar`	局所定常 AR モデルによる時系列の局所定常区間への分割
`lsar.chgpt`	局所定常 AR モデルによる時系列の変化点の推定
`lsqr`	ハウスホルダー変換を用いた最小二乗法による三角関数回帰モデルの推定
`marfit`	ユール-ウォーカー法による多変量 AR モデルのあてはめ
`marlsq`	最小二乗法による多変量 AR モデルのあてはめ
`marspc`	クロススペクトル，コヒーレンシーと相対パワー寄与率の計算
`ngsim`	非ガウス型状態空間モデルによるシミュレーション
`ngsmth`	非ガウス型平滑化
`pdfunc`	確率密度関数の計算
`period`	ピリオドグラムの計算
`pfilter`	粒子フィルタと粒子平滑化
`pfilterNL`	非線形状態空間モデルの粒子フィルタと粒子平滑化
`polreg`	多項式回帰モデルの推定
`season`	季節成分を含む時系列の季節調整
`simssm`	ガウス型状態空間モデルによるシミュレーション
`trend`	トレンド推定
`tsmooth`	状態空間モデルによる時系列の予測と補間
`tvar`	時変係数 AR モデルの推定
`tvspc`	時変係数 AR モデルによる時変スペクトルの計算
`tvvar`	時変分散モデルの推定
`unicor`	自己共分散関数と自己相関関数の計算

TSSS パッケージには本書で用いられる下記のデータも含まれている．

`BLSALLFOOD`	アメリカの食品産業の従事者数
`Haibara`	地下水位データ
`HAKUSAN`	船舶の航行中の多変量データ
`MYE1F`	地震波データ
`Nikkei225`	日経 225 平均株価データ
`NLmodel`	非線形状態空間モデルのためのテストデータ
`PfilterSample`	平均がシフトする時系列の人工データ
`Rainfall`	各月日について東京で 2 年間に雨が降った回数
`Sunspot`	太陽黒点数データ
`Temperature`	東京の日最高気温データ
`WHARD`	あるハードウェアの卸売高データ

目　次

新版へのまえがき

初版まえがき

R と時系列解析パッケージ TSSS

第 1 章　時系列データの解析とその準備

1.1　時系列データ ……… 1
1.2　時系列の分類 ……… 6
1.3　時系列解析の目的 ……… 8
1.4　時系列データの前処理 ……… 9
1.4.1　変数変換 ……… 9
1.4.2　差分(階差) ……… 10
1.4.3　前期比，前年同期比 ……… 12
1.4.4　移動平均 ……… 13
[章末問題] ……… 16

第 2 章　共分散関数

2.1　時系列の分布と定常性 ……… 17
2.2　定常時系列の自己共分散関数 ……… 21
2.3　多変量時系列と散布図 ……… 25
2.4　相互共分散関数および相互相関関数 ……… 27
[章末問題] ……… 31

第 3 章　スペクトルとピリオドグラム

3.1　スペクトル ……… 33
3.2　ピリオドグラム ……… 37

3.3　ピリオドグラムの平均と平滑化 …… 42
3.4　ピリオドグラムの計算法 …… 46
3.5　FFT によるピリオドグラム計算 …… 47
［章末問題］ …… 50

第 4 章　モデリング

4.1　確率分布と統計的モデル …… 51
4.2　K-L 情報量とエントロピー最大化原理 …… 56
4.3　K-L 情報量の推定と対数尤度 …… 59
4.4　最尤法によるパラメータの推定 …… 61
4.5　AIC（赤池情報量規準） …… 64
4.6　データの変換 …… 68
［章末問題］ …… 73

第 5 章　最小二乗法

5.1　回帰モデルと最小二乗法 …… 75
5.2　ハウスホルダー法に基づく最小二乗法の解法 …… 77
5.3　AIC による次数選択 …… 79
5.4　データの追加と分割処理 …… 83
5.5　AIC による変数選択 …… 84
［章末問題］ …… 86

第 6 章　ARMA モデルによる時系列の解析

6.1　ARMA モデル …… 87
6.2　インパルス応答関数 …… 88
6.3　自己共分散関数 …… 90
6.4　AR 係数と PARCOR の関係 …… 91
6.5　パワースペクトル …… 93
6.6　特性方程式 …… 96
6.7　多変量 AR モデル …… 99

[章末問題] ………… 105

第7章　ARモデルの推定

7.1　ARモデルのあてはめ ………… 109
7.2　ユール-ウォーカー法とレビンソンのアルゴリズム ………… 111
7.3　最小二乗法によるARモデルの推定 ………… 113
7.4　PARCOR法によるARモデルの推定 ………… 115
7.5　AR係数の推定量の誤差分布 ………… 118
7.6　数 値 例 ………… 119
7.7　ユール-ウォーカー法による多変量ARモデルの推定 ………… 122
7.8　最小二乗法による多変量ARモデルの推定 ………… 123
[章末問題] ………… 130

第8章　局所定常ARモデル

8.1　局所定常ARモデル ………… 131
8.2　任意個の区間への自動分割 ………… 133
8.3　変化時点の精密な推定 ………… 139
8.4　変化時点の事後確率 ………… 144
[章末問題] ………… 146

第9章　状態空間モデルによる時系列の解析

9.1　状態空間モデル ………… 147
9.2　カルマンフィルタによる状態の推定 ………… 150
9.3　平滑化のアルゴリズム ………… 153
9.4　状態の長期予測 ………… 153
9.5　時系列の予測 ………… 154
9.6　時系列モデルの尤度計算とパラメータ推定 ………… 158
9.7　欠測値の補間 ………… 161
[章末問題] ………… 165

第 10 章　ARMA モデルの推定

10.1　ARMA モデルの状態空間表現 ………… 167
10.2　AR モデルの初期状態 ………… 168
10.3　ARMA モデルの初期状態 ………… 171
10.4　ARMA モデルの最尤推定 ………… 172
10.5　パラメータの初期値について ………… 175
[章末問題] ………… 177

第 11 章　トレンドの推定

11.1　多項式回帰モデル ………… 179
11.2　トレンド成分モデル――構造の確率的変化のモデル ………… 182
11.3　トレンドモデル ………… 187
[章末問題] ………… 192

第 12 章　季節調整モデル

12.1　季節成分モデル ………… 193
12.2　標準的季節調整モデル ………… 196
12.3　定常 AR 成分を含む分解 ………… 200
12.4　曜日効果項を含む分解 ………… 206
[章末問題] ………… 210

第 13 章　時変係数 AR モデル

13.1　時変分散モデル ………… 213
13.2　時変係数 AR モデル ………… 216
13.3　時変スペクトルの推定 ………… 222
13.4　時変係数 AR モデルのシステムノイズの仮定 ………… 225
13.5　係数の急激な変化について ………… 226
[章末問題] ………… 228

第 14 章　非ガウス型モデル

14.1　非ガウス型モデルの必要性 ………… 229
14.2　非ガウス型状態空間モデルと状態推定 ………… 230
14.3　状態推定公式の数値的実現 ………… 232
14.4　非ガウス型トレンドモデル ………… 235
14.5　非対称な分布——時変分散モデル ………… 239
14.6　非ガウス型状態空間モデルの応用 ………… 244
14.6.1　混合ガウス分布による異常値の処理 ………… 244
14.6.2　非定常離散系列 ………… 245
14.6.3　時変分散を直接推定する方法 ………… 246
14.6.4　非線形状態空間モデル ………… 246
[章末問題] ………… 247

第 15 章　粒子フィルタ・平滑化

15.1　非線形・非ガウス型の状態空間モデルと分布の近似 ………… 249
15.2　粒子フィルタ ………… 252
15.2.1　1 期先予測 ………… 252
15.2.2　フィルタ ………… 253
15.2.3　粒子フィルタのアルゴリズム ………… 253
15.2.4　モデルの尤度 ………… 254
15.2.5　リサンプリング法について ………… 254
15.2.6　数 値 例 ………… 256
15.3　粒子平滑化 ………… 258
15.3.1　粒子近似による平滑化 ………… 258
15.3.2　非線形平滑化 ………… 262
[章末問題] ………… 264

第 16 章　シミュレーション

16.1　一様乱数の生成 ………… 267
16.2　白色雑音の生成 ………… 270
16.2.1　χ^2 分布 ………… 271
16.2.2　コーシー分布 ………… 273

16.2.3　任意の分布に従う乱数の生成 …… 273
16.3　ARMA モデルのシミュレーション …… 274
16.4　状態空間モデルによるシミュレーション …… 276
16.5　非ガウス型状態空間モデルによるシミュレーション …… 280
[章末問題] …… 284

付録 A　非線形最適化のアルゴリズム …… 285

付録 B　レビンソンのアルゴリズムの導出 …… 287

付録 C　カルマンフィルタと平滑化のアルゴリズムの導出 …… 290
C.1　カルマンフィルタ …… 290
C.2　平 滑 化 …… 292

付録 D　粒子フィルタのアルゴリズム …… 294

参考文献　297
章末問題解答　299
索　引　313

第 1 章 時系列データの解析とその準備

時系列データにはさまざまな特徴を持つものがあるので，時系列解析の第一歩として，時系列のグラフ表示を行い，その特徴を十分把握することが重要である．グラフ表示によって，次に進むべき分析の方法あるいは用いるべきモデルに関する示唆が得られることが多い．本章の前半では第 2 章以降の数値例で用いられる時系列データを図示し，その特徴を明らかにするとともに，一般的な時系列の分類や時系列モデリングの目的についても説明する．また後半では，変数変換や階差など時系列の前処理の方法を説明する．本書ではほとんどの数値例の計算に統計計算環境 R を用いている．それらの例については TSSS パッケージを中心に R 関数の利用法についても説明する．

1.1 時系列データ

時間の経過とともに不規則に変動する現象の記録が**時系列**(time series)である．気圧，気温や雨量などの気象情報，地震波の記録，株価や為替レートなどの経済現象の記録，脳波や心電図などの医学データ，自動車，船舶，航空機の操縦の記録など，我々の生活の身近なところだけを取り上げてみても数多くの時系列を思い浮かべることができる．

時系列解析の第一歩として，まずデータを図示してみることが重要である．これにより，時系列のおおまかな特徴を捉えることができるばかりでなく，必要な前処理や，その後どのような解析を行うべきかの方針をたてることができる．

本書の全体にわたって，以下のデータセットを例題として用いる．これらのデータは R パッケージ TSSS の中に含まれており，以下に示すように関数

data によって取り込むことができる.

```
> data( HAKUSAN ) # 船舶の航行中の多変量データ
> data( Sunspot ) # 太陽黒点数データ
> data( Temperature ) # 東京の日最高気温データ
> data( BLSALLFOOD ) # アメリカの食品産業の従事者数
> data( WHARD ) # あるハードウェアの卸売高データ
> data( MYE1F ) # 地震波データ
> data( Nikkei225 ) # 日経 225 平均株価データ
> data( Rainfall ) # 各月日について東京で 2 年間に雨が降った回数
> data( Haibara ) # 地下水位データ
```

また，これら以外のデータセットを解析したい場合には，関数 read.csv などによって csv 形式のデータを読み込むことができる．ここで，関数 as.ts は読み込んだデータを時系列オブジェクトとするためのものである．また，引数(ひきすう) start によって時系列の始まりの時点を，また frequency によって単位時間あたりの観測点数を登録することができる．デフォルト値は start=1, frequency=1 である．月次データの場合は frequency=12 とし，start= c(1998,5) のようにすればデータの開始年と月を登録できる．

```
> sunspot <- as.ts( read.csv( "sunspot.csv" ) )
>
> blsfood <- as.ts( read.csv( "blsfood_new.csv" ), start = c(1967,1),
frequency = 12 )
```

いったん R にデータが読み込まれると，関数 plot によって，そのグラフを描くことができる．この関数では引数 ylim および main によって縦軸の上限・下限および図のタイトルを指定することができる．

```
> par( mar=c(2,2,3,1)+0.1 )
> plot( HAKUSAN[,1],main="(a) 船舶の方向角速度" )
> plot( Sunspot,main="(b) 太陽黒点数" )
> plot( Temperature,main="(c) 東京の日最高気温" )
> plot( BLSALLFOOD,main="(d) 食品産業従事者数" )
> plot( WHARD,main="(e) 卸売高" )
> plot( MYE1F,main="(f) 地震波(東西方向)" )
> plot( Nikkei225,main="(g) 日経 225 平均株価" )
> plot( Haibara,main="(h) 地下水位(上段)と気圧(下段)" )
```

```
> plot( HAKUSAN[,c(2,3,4)],main="(i) 船舶の横揺れ(上段)，縦揺れ(中段)，舵角
(下段)" )
```

図 1.1 に以降の数値例で用いられる典型的な時系列を図示する．少し長くなるが，それぞれのデータの特徴を説明しておく．

(a) 4 変量の HAKUSAN データの第 1 列で外洋を航行中の船舶の**方向角速度**を 1 秒ごとに記録した時系列である．針路を一定に保つ保針制御が行われているので，方向角速度は 0 の前後を変動している．(大津皓平氏提供)

(b) 毎年の**太陽黒点数**(Wolf sunspot number)を記録した時系列である．ほぼ 11 年周期の同様なパターンで増減を繰り返している．

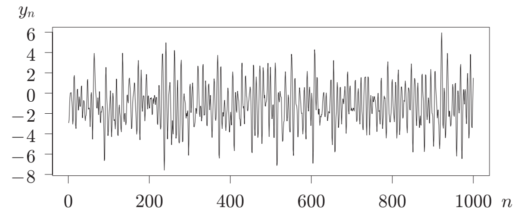
(c) 東京の毎日の**最高気温**を 16 カ月にわたって記録したデータである．支配的な年周期(トレンド)が見られるが，年周期のまわりでは不規則な変動をしている．(東京管区気象台公表)

(d) アメリカの食品産業に従事する労働者の人数を毎月調べた時系列である．経済時系列に特有な，徐々に変化するトレンド成分と毎年同じような変動を繰り返す季節成分を含んでいる．(合衆国 Bureau of Labor Statistics (BLS)公表)

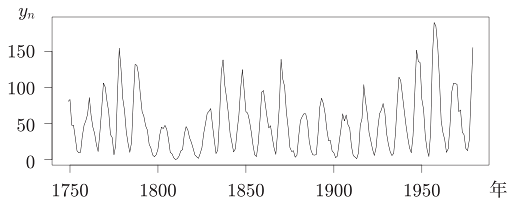
(e) あるハードウェアの毎月の**卸売高**を記録したデータである．毎年ほぼ一定の割合で増加する経済データに特有な特徴が見られ，それにともなってトレンドのまわりの変動も徐々に大きくなっている．(合衆国 Bureau of Labor Statistics (BLS)公表)

(f) **地震波**の東西成分を約 0.02 秒間隔で記録した時系列で，P 波(縦波)および S 波(横波)の到着にともなって変動の大きさが著しく変化している．また詳しく調べてみると，振幅だけでなく変動の仕方自体も変化していることがわかる．(高波鐵夫氏提供)

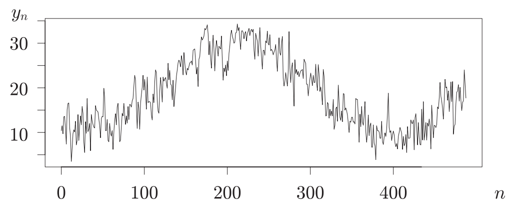
(g) 日経 225 平均**株価データ**で，1988 年 1 月 4 日から 1993 年 12 月 30 日までの日次終値を示す．1989 年末まではほぼ単調に増加しているが，バブル崩壊後は，大きな変動を繰り返しながら徐々に下降している．株



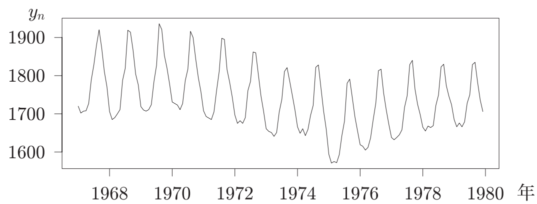
(a) 船舶の方向角速度



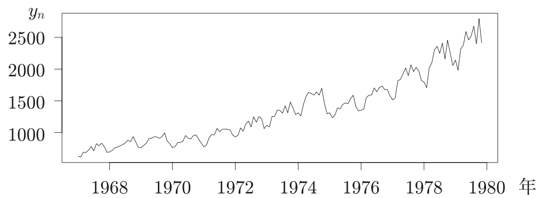
(b) 太陽黒点数



(c) 東京の日最高気温

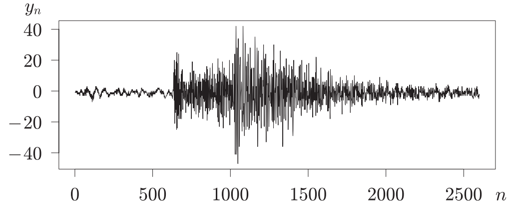


(d) 食品産業従事者数

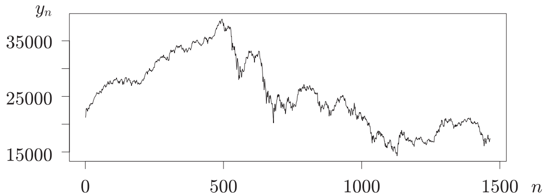


(e) 卸売高

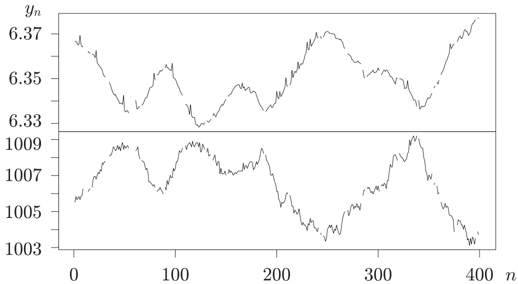
図 1.1　いろいろな時系列の例



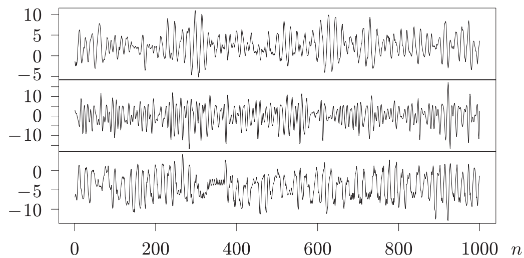
(f) 地震波(東西方向)



(g) 日経 225 平均株価



(h) 地下水位(上段)と気圧(下段)



(i) 船舶の横揺れ(上段)，縦揺れ(中段)，舵角(下段)

図 1.1　いろいろな時系列の例(続き)

価データの解析は対数変換後，差分をとってから行うことが多い．

(h) 東海地方の観測点で**地下水位**と**気圧**の変動を 10 分ごとに観測した 2 変量時系列である．折れ線の空白の部分は何らかの原因で観測が行われなかったところを表している．また，水位データのところどころに上方に飛び出した値が見られるが，これは観測機器の問題による異常値と考えられる．長期間にわたる継続観測によって得られた貴重なデータを有効に利用するためには，このようなデータでも解析できる方法を開発することが必要である．(高橋誠氏，松本則夫氏提供)

(i) 太平洋を航行中の船舶の**横揺れ**，**縦揺れ**および**舵角**を 1 秒ごとに記録した多変量時系列である．横揺れと舵角は，ともに十数秒程度の周期的な変動をしていることがわかる．一方，縦揺れは 10 秒以下の短い周期で変動している．(大津皓平氏提供)

1.2　時系列の分類

図 1.1 に見られるように時系列にはさまざまな種類のものがある．これらの時系列は以下のようにいろいろな観点から類別することができる．

連続時間時系列と離散時間時系列

アナログレコーダー等で連続的に記録されたデータは**連続時間**(continuous time)時系列と呼ばれる．これに対し，1 時間おきに計測された気圧のように，ある時間間隔で観測されたものを**離散時間**(discrete time)時系列と呼ぶ．離散時間時系列には**等間隔**(equally-spaced)に観測されたものと**不等間隔**(unequally-spaced)なものとがある．図 1.1 に示した時系列は実線で連続的に結ばれているが，実際にはすべて離散時間時系列である．計算機を用いて解析する場合には，離散的な観測値を取り扱うことがほとんどであるので，今後この本では，等間隔な離散時間時系列のことを単に時系列と呼ぶことにする．

1 変量時系列と多変量時系列

図 1.1 の(a)〜(g)のように各観測時点で 1 種類の情報だけが得られたもの

が**1変量**(univariate)時系列である．これに対し，(h)や(i)などのように2つ以上の情報を同時に記録したものが**多変量**(multi-variate)時系列である．ただし，ある現象と同時に起こる現象はいくらでもあるので，時系列が1変量であるか多変量であるか，また多変量の場合でもどのような変数を同時に考慮するかは解析の対象によって一意的に決まるものではなく，計測上の制限，目的やそれまでの解析結果などによって決められるものである．また，モデリングの立場からみると，解析に用いる変数の選択自体がきわめて重要な統計的問題である．

定常時系列と非定常時系列

時系列は時間とともに不規則な変動をしているが，時系列解析ではそのような不規則な変動を確率的なモデルで表現する．このとき，一見したところ不規則な現象でも時間的に変化しない一定の確率的モデルの実現値とみなすことができる場合がある．このような時系列は**定常**(stationary)時系列と呼ばれる．図1.1の中では(a)が典型的な定常時系列と考えられる．一方，(c)，(d)，(e)，(g)のように平均が時間とともに変動していたり，(f)のように平均のまわりの変動の仕方が時間的に変化しているものを**非定常**(nonstationary)時系列という．

ガウス型時系列と非ガウス型時系列

時系列の分布が正規分布(ガウス分布)に従うものが**ガウス型**(Gaussian)時系列，そうでない場合が**非ガウス型**(non-Gaussian)時系列である．この本で取り扱う多くのモデルは，ガウス分布に従うものと仮定したガウス型時系列モデルである．(b)のように変動のパターンに上下非対称性が見られ，そのままでは周辺分布が正規分布とはみなせない時系列でも，データに適切な変換を施すことによって近似的にガウス型時系列とみなせるようになることもある．

線形時系列と非線形時系列

線形なモデルの出力として表現できるような時系列は，**線形**(linear)時系列と呼ばれる．これに対し非線形なモデルが必要なものは，**非線形**(nonlinear)

時系列と呼ばれる.

欠測値と異常値

以上のような時系列の分類法とは異なるが，実際の時系列解析において注意すべきものとして欠測値と異常値がある．(h)に見られるように，何らかの理由で観測値が記録されなかった場合，その部分を**欠測値**(missing value, missing observation)と呼ぶ．また，観測している対象自体の異常な振舞い，観測装置の異常，記入やデータ転送時のミス等によって明らかに異常なデータがある場合，これらは**異常値**(outlier，外れ値)と呼ばれる．(h)の水位データに何カ所か見られる上方へ突出した観測値は異常値と考えられる.

1.3　時系列解析の目的

本書では，以上のような時系列の統計解析のための方法を説明する．本書で考える問題は以下のように 4 つの問題に大別できるが，これらを総称して**時系列解析**(time series analysis)と呼ぶことにする.

記述・可視化(description, visualization)

時系列を図示したり，標本自己共分散関数，標本自己相関関数，ピリオドグラムなどの基本的な記述統計量を用いて時系列の特徴を簡潔に表現すること．時系列解析では大量の数値が出力されるので，グラフによって表現することが効果的である.

モデリング(modeling)

与えられた時系列に対し，その変動の仕方を表現する時系列モデルを推定し，時系列の確率構造を解析すること．時系列にはさまざまな特徴を持つものがあるので，解析の対象や目的に応じて適切な時系列モデルを選択し，そのモデルに含まれるパラメータを推定することが必要である.

予測・シミュレーション(prediction, simulation)

時系列が時間的な相関を持つことを利用して，現在までに得られた情報から今後の変動を予測すること．とくに本書では，推定されたモデルを利用して予測したりシミュレーションを行う方法を考慮する．

信号抽出(signal extraction)

時系列から目的に応じて必要な信号や情報を取り出すこと．対象の特徴や目的に応じて適切なモデリングを行うことが重要である．

1.4　時系列データの前処理

時系列データに非定常性や非対称性が見られる場合には，第 2 章以降の解析に進む前に，データの前処理(1 次処理)を行うことがある．本節では定常化のためのいくつかの方法を紹介する．ただし，第 11 章以降では，このようなデータの前処理によらず，非定常な時系列を直接解析する方法を取り上げる．

1.4.1　変数変換

図 1.1 の(e)や(g)などの金額や人数などの正値のデータの多くは，値の増加にともなって，変動の大きさも増大するという特徴を持っている．このような場合，原系列 y_n の代わりに，その対数変換 $z_n = \log y_n$ を用いると，分散がより一様になったり，誤差分布がより正規分布に近くなる場合がある．

図 1.2 に卸売高データとその対数変換データを示す．対数変換によって，トレンドまわりの変動がより一様になり，またトレンドも直線的になっていることがわかる．

```
> data( WHARD )
> log_WHARD <- log10( WHARD )
> par( mfrow = c(2,1), mar = c(2,4,1,1)+0.1 )
> plot( WHARD, type = "l" )
> plot( log_WHARD, type = "l" )
```

対数変換を特殊な場合として含む，より一般の Box-Cox 変換の自動的な決

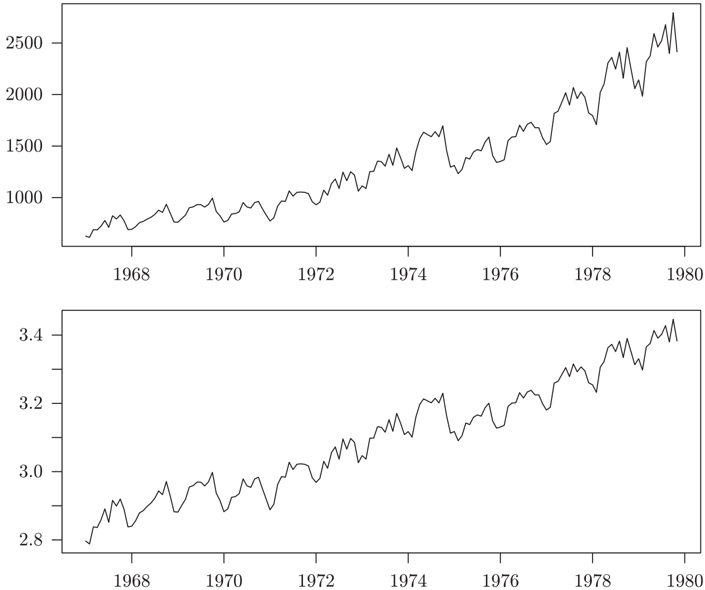


図 1.2　卸売高データ(上段)と対数変換データ(下段)

定など，変数変換の詳細については 4.6 節で解説する．また，ある事象が起こる確率や割合のような，$(0,1)$ 上の値をとる時系列 y_n の場合には，ロジット変換

$$z_n = \log\left(\frac{y_n}{1-y_n}\right) \tag{1.1}$$

によって $(-\infty,\infty)$ 上の値をとる時系列 z_n が得られる．変換後の z_n のほうが分布の偏りが少なく，モデリングが容易になることが多い．

1.4.2　差分(階差)

時系列 y_n が図 1.1 の(c)，(d)，(e)，(g)のようにトレンドを含む場合，まず差分系列

$$z_n = \Delta y_n = y_n - y_{n-1} \tag{1.2}$$

を求め，z_n を解析することがある．これは y_n が直線で $y_n = a + bn$ と表され