

まえがき

この本は、初学者のために書かれた時系列分析の入門書です。

パラメタ推定など技術的な面は一部省略されている代わりに、時系列データの構造や、様々な時系列モデルの特徴、ライブラリを用いた分析の手順や、結果の解釈の仕方に紙数を割きました。

Box-Jenkins 法と状態空間モデルを共に学べるのも大きな特徴です。初めて学ぶ方でも、古典的な内容からステップアップするため無理なく学べるはずです。

ビッグデータという言葉が代表するように、とても多くのデータが分析に使われるようになりました。しかし、どれほど多くのデータを持っていても、決して手に入らないデータがあります。

それは未来のデータです。

100 テラバイトの売り上げデータを持っていても、高頻度証券データを取得したとしても、1 ミリ秒先の未来のデータは、私たちの手元にありません。

まだ手に入っていない未来のデータについて言及することが予測であり、時系列分析は、予測を行う強力なツールです。そして、状態空間モデルは、現代の時系列分析の事実上の標準ともいえるフレームワークです。

この本では「R」や「Stan」といった無料のソフトウェアを用いて、時系列データを効率的に分析する方法も説明します。

統計学は、無から有を生み出す錬金術ではありません。

常に予測ができるわけではなく、どうしても予測の出しようがないこともあります。また、時系列データ特有の問題として、素朴な回帰分析などを適用すると、誤った考察を導いてしまうこともあります。

時系列データを分析する際の注意点、逆に、時系列データだからこそ持っている情報を利用する方法。両者をバランスよく学び、時系列データを有効活用してください。

統計学は便利な道具です。統計学を教える書籍も便利な道具であるべきです。本書が皆さんにとって、有用なツールとなることを願います。

本書の構成

この本の難易度

この本では、時系列分析のアイデアを伝えることに注力しました。統計学にそれほど詳しくなくても最後まで読めるように配慮されています。数式の量も、この分野の標準教科書と比べるとかなり抑えられているはずです。

しかし、時系列分析は統計学の応用編ともいえる分野です。推定や検定、回帰分析や最小二乗法といった言葉のある程度知っている、という方がこの本を読まれると良いでしょう。

この本の読み方

節のタイトルにアスタリスク (*) がついているものは、テクニカルな話題となるため初学者が読むのはやや難しい可能性があります。

アスタリスクの有無にかかわらず、少し難しいかなと思った箇所は（数式も含めて）どんどん飛ばしていき、自分がわかる部分だけをかいつまんで読むというのも、良いやり方だと思います。特に数式は、必ずその解釈を日本語で書くようにしているので、たとえ飛ばしたとしてもある程度は理解できるはずです。

また、理論の説明をした後でソフトを使って実装するという説明の仕方で統一されています。難しいと感じた理論は軽く読み流したうえで、自分で実装しながら都度読み返すという進め方をとることもできます。

計算のためのソフトウェアはすべて無料で手に入れることができます。ソースコードは著者の Web サイト (<https://logics-of-blue.com/>) から無料でダウンロードできます。

この本の構成

本書は 6 部構成となっています。目次をかなり詳細に書いてあるため、自分が今どこにいるのか、次に何を学ぶのかがわかるようになっています。

「第 1 部 時系列分析の考え方」では、時系列分析とは何かという基本から説明をします。

特に時系列データの構造、時系列モデルという考え方を理解してください。後

ほど学ぶ分析手法のほぼすべてで必要とされる考え方です

「第2部 Box-Jenkins 法とその周辺」では ARIMA モデルと呼ばれる古典的な時系列モデルを中心とした分析の方法を説明します。

Box-Jenkins 法は比較的古い手法とはいえ現代でも十分に実用的です。

また、Box-Jenkins 法は時系列分析の基礎を学ぶ格好の教材ともいえます。定常性や和分過程といった時系列データ特有の考え方に加え、モデル選択や残差のチェック、予測精度の評価といった分析における一般的な流れを学ぶことができるからです。ここまでを読了できれば、時系列分析の基本的な用語や考え方が身についているはずで

「第3部 時系列分析のその他のトピック」は、独立した3つの章で構成されています。

1つは時系列データに対して回帰分析を適用した際の問題点「見せかけの回帰」について。

2つ目は多変量時系列データの分析手法としての VAR モデルについて。

3つ目は分散不均一なデータへの分析手法として ARCH・GARCH モデルを解説します。

「第4部 状態空間モデルとは何か」でこの本のメインテーマの1つである状態空間モデルの導入をします。

「第5部 状態空間モデルとカルマンフィルタ」では、カルマンフィルタを用いて、線形ガウス状態空間モデルを推定します。カルマンフィルタや散漫カルマンフィルタ、平滑化などの考え方と計算方法を解説します。

(散漫)カルマンフィルタと平滑化に関しては、ローカルレベルモデルと呼ばれる単純なモデルを例に挙げて、ライブラリを使わずに自分の手で実装します。基礎を学んだ後、現実に近い問題を、ライブラリを用いて分析していきます。

「第6部 状態空間モデルとベイズ推論」では、MCMC を用いて非ガウシアン・非線形のモデルを構築します。ベイズ推論の基礎から、一般化状態空間モデルの実装方法までを解説します。

目次

まえがき	i
本書の構成	ii
第 1 部 時系列分析の考え方	17
1 章 時系列分析の基礎	18
1-1 推測統計学の考え方	18
1-2 時系列データとは	18
1-3 時系列データをどのように取り扱うべきか	19
1-4 母集団と確率分布・標本と確率変数	20
1-5 データ生成過程(DGP)の考え方	20
1-6 DGP と時系列モデル	21
2 章 時系列データの構造	23
2-1 自己相関とコレログラム	23
2-2 季節成分・周期成分	23
2-3 トレンド	24
2-4 外因性	24
2-5 ホワイトノイズ	25
2-6 時系列データの構造	25
3 章 数式による時系列データの表記方法	26
3-1 データの表記方法	26
3-2 確率変数の表記方法	26
3-3 期待値・分散	27
3-4 自己共分散・自己相関	27
3-5 偏自己相関	28
3-6 ホワイトノイズ	30
3-7 独立で同一の分布に従う系列 : iid 系列	30

3-8	ランダムウォークと確率的トレンド	31
第2部	Box-Jenkins法とその周辺	32
1章	Box-Jenkins法というフレームワーク	33
1-1	Box-Jenkins法における分析の手順	33
1-2	この本における説明の手順	33
1-3	Box-Jenkins法のメリット・デメリット	34
2章	定常過程とデータの変換	35
2-1	分析しやすいデータの特徴：定常性	35
2-2	定常過程が分析しやすいデータである理由	36
2-3	優秀なモデル：ARMAモデル	37
2-4	分析しにくいデータ：非定常過程	37
2-5	差分系列と単位根・和分過程	37
2-6	対数差分系列とその解釈	39
2-7	対数変換とその解釈	39
2-8	季節階差	40
2-9	定常過程と非定常過程のデータ例	41
3章	ARIMAモデル	42
3-1	自己回帰モデル(ARモデル)	42
3-2	予測と条件付期待値	43
3-3	ARモデルと条件付確率分布とデータ生成過程	43
3-4	ARモデルとホワイトノイズ・ランダムウォーク	44
3-5	移動平均モデル(MAモデル)	45
3-6	ARモデル・MAモデルのコレログラム	46
3-7	ARモデルとMAモデルの関係*	49
3-8	ARモデルの定常条件*	50
3-9	MAモデルの反転可能性*	50

3-10	自己回帰移動平均モデル(ARMA).....	51
3-11	自己回帰和分移動平均モデル(ARIMA).....	52
4章	ARIMA モデルの拡張.....	54
4-1	季節性 ARIMA モデル(SARIMA).....	54
4-2	ラグ演算子による SARIMA モデルの数理的表現*.....	55
4-3	外生変数付き ARIMA モデル(ARIMAX).....	57
4-4	まとめ：ARIMA と時系列の構造.....	58
5章	モデルの同定と評価の考え方.....	59
5-1	モデルの同定.....	59
5-2	パラメタの推定.....	59
5-3	ARMA 次数の決定：モデル選択.....	60
5-4	尤度と対数尤度・最尤法・最大化対数尤度.....	61
5-5	モデル選択と AIC.....	61
5-6	予測と当てはめ.....	62
5-7	なぜ AIC を使うのか.....	63
5-8	差分をとるか判断する：単位根検定.....	63
5-9	単位根検定：KPSS 検定.....	64
5-10	単位根検定：ADF 検定.....	65
5-11	評価：モデルの定常性・反転可能性のチェック*.....	66
5-12	自動 ARIMA 次数決定アルゴリズム.....	67
5-13	評価：残差の自己相関のテスト.....	67
5-14	評価：残差の正規性のテスト.....	68
5-15	テストデータによる予測精度の評価.....	69
5-16	ナイーブ予測との比較.....	70
6章	R による時系列データの取り扱い.....	72
6-1	R のインストール.....	72

6-2	RStudio のインストール	73
6-3	RStudio の簡単な使い方	73
6-4	四則演算	74
6-5	変数	75
6-6	関数とヘルプ	76
6-7	ベクトル	77
6-8	行列	78
6-9	データフレーム	80
6-10	リスト	82
6-11	パッケージのインストール	84
6-12	時系列データ <code>ts</code> 型	85
6-13	拡張された時系列データ <code>xts</code> 型	88
6-14	ファイルからのデータ取り込み	90
6-15	グラフ描画	92
6-16	単位根検定	94
7 章	R による ARIMA モデル	95
7-1	この章で使うパッケージ	95
7-2	分析の対象	95
7-3	対数変換	96
7-4	差分系列の作成方法	97
7-5	季節成分の取り扱い	99
7-6	自己相関とコレログラム	101
7-7	訓練データとテストデータに分ける	102
7-8	ARIMA モデルの推定	103
7-9	補足：差分系列と ARIMA の次数の関係	104
7-10	自動モデル選択 <code>auto.arima</code> 関数	106

7-11	定常性・反転可能性のチェック*	108
7-12	残差のチェック	109
7-13	ARIMAによる予測	111
7-14	ナイーブ予測	113
7-15	予測の評価	113
7-16	発展：非定常過程系列への分析	115
第3部	時系列分析のその他のトピック	117
1章	見せかけの回帰とその対策	118
1-1	この章で使うパッケージ	118
1-2	ホワイトノイズへの回帰分析	118
1-3	単位根のあるデータ同士の回帰分析	120
1-4	定常AR過程への回帰分析	123
1-5	残差の自己相関と見せかけの回帰	125
1-6	Durbin-Watson検定	125
1-7	シミュレーションによる見せかけの回帰	127
1-8	見せかけの回帰を防ぐ方法	130
1-9	単位根検定	131
1-10	一般化最小二乗法：GLS	132
1-11	RによるPrais-Winsten法	134
1-12	パッケージを使ったGLS	136
1-13	差分系列への回帰分析	137
1-14	共和分	138
1-15	共和分検定	141
2章	VARモデル	144
2-1	VARモデルの使い時	144
2-2	VARモデルの構造	145

2-3	Granger 因果性検定.....	146
2-4	インパルス応答関数.....	147
2-5	この章で使うパッケージ.....	147
2-6	分析の対象.....	148
2-7	RによるVARモデル.....	151
2-8	VARモデルによる予測.....	155
2-9	RによるGranger因果性検定.....	156
2-10	Rによるインパルス応答関数.....	158
3章	ARCH・GARCHモデルとその周辺.....	161
3-1	なぜ分散の大きさをモデル化したいのか.....	161
3-2	自己回帰条件付き分散不均一モデル(ARCH).....	161
3-3	一般化ARCHモデル(GARCH).....	163
3-4	GARCHモデルの拡張.....	164
3-5	この章で使うパッケージ.....	165
3-6	シミュレーションによるデータの作成.....	165
3-7	fGarchパッケージによるGARCHモデル.....	169
3-8	rugarchパッケージによるGARCHモデル.....	169
3-9	ARMA-GARCHモデルの作成.....	170
3-10	RによるGJRモデル.....	175
第4部	状態空間モデルとは何か.....	178
1章	状態空間モデルとは何か.....	179
1-1	状態空間モデルとは何か.....	179
1-2	状態空間モデルのメリット・デメリット.....	180
2章	状態空間モデルの学び方.....	181
2-1	状態空間モデルを学ぶとは、何を学ぶことか.....	181
2-2	データの表現とパラメタ推定は分けて理解する.....	182

2-3	データの表現：状態方程式・観測方程式	183
2-4	パラメタ推定：その分類	183
2-5	パラメタ推定：カルマンフィルタと最尤法	183
2-6	パラメタ推定：ベイズ推論と HMC 法	184
第 5 部	状態空間モデルとカルマンフィルタ	186
1 章	線形ガウス状態空間モデルとカルマンフィルタ	187
1-1	表現：状態方程式・観測方程式	187
1-2	状態推定：予測とフィルタリング	187
1-3	状態推定：平滑化	188
1-4	パラメタ推定：最尤法	189
1-5	線形ガウス状態空間モデルを推定する流れ	190
2 章	表現：状態方程式・観測方程式による表現技法	191
2-1	線形回帰モデルと状態方程式・観測方程式	191
2-2	自己回帰モデルと状態方程式・観測方程式	192
2-3	ローカルレベルモデル	193
2-4	この章で扱う具体例	193
2-5	ローカルレベルモデルと線形回帰モデルの比較	193
2-6	ローカルレベルモデルによる予測	195
2-7	ローカルレベルモデルと ARIMA モデルの関係	195
2-8	ローカル線形トレンドモデル	196
2-9	ローカル線形トレンドモデルと線形回帰の比較	197
2-10	ローカル線形トレンドモデルによる予測	198
2-11	行列による線形ガウス状態空間モデルの表現*	198
2-12	補足：トレンドモデル	199
2-13	周期的変動のモデル化	200
2-14	基本構造時系列モデル	201

2-15	外生変数と時変係数モデル	202
3 章	状態推定：カルマンフィルタ	204
3-1	カルマンフィルタとカルマンゲイン	204
3-2	カルマンゲインの求め方	205
3-3	日本語で読むカルマンフィルタ	205
3-4	数式で見るカルマンフィルタ	207
4 章	状態推定：散漫カルマンフィルタ	209
4-1	状態の初期値の問題	209
4-2	散漫初期化という解決策	210
4-3	日本語で読む散漫カルマンフィルタ	210
4-4	数式で見る散漫カルマンフィルタ	211
4-5	初期値がもたらす影響	212
4-6	過程誤差・観測誤差の分散がもたらす影響	213
5 章	状態推定：平滑化	214
5-1	平滑化の考え方	214
5-2	日本語で読む平滑化	214
5-3	数式で見る平滑化	215
5-4	フィルタ化推定量と平滑化状態の比較	218
6 章	パラメタ推定：最尤法	219
6-1	ローカルレベルモデルで推定するパラメタの種類	219
6-2	パラメタ推定の原理	219
6-3	カルマンフィルタと対数尤度	220
6-4	散漫カルマンフィルタと散漫対数尤度	221
6-5	数式で見る対数尤度と散漫対数尤度	221
7 章	実装：Rによる状態空間モデル	223
7-1	この章で使うパッケージ	223

7-2	分析の対象.....	223
7-3	Rで実装するカルマンフィルタ：関数を作る.....	224
7-4	Rで実装するカルマンフィルタ：状態を推定する.....	226
7-5	Rで実装するカルマンフィルタの対数尤度.....	228
7-6	Rで実装する最尤法.....	229
7-7	Rで実装する平滑化：関数を作る.....	231
7-8	Rで実装する平滑化：状態を推定する.....	232
7-9	dmlmによるカルマンフィルタ.....	233
7-10	dmlmによる対数尤度の計算.....	234
7-11	dmlmによる平滑化.....	234
7-12	参考：dmlmの使い方.....	235
7-13	Rで実装する散漫カルマンフィルタ.....	237
7-14	Rで実装する散漫対数尤度.....	239
7-15	KFASによる散漫カルマンフィルタ.....	240
7-16	KFASによる散漫対数尤度の計算.....	240
7-17	dmlmとKFASの比較とKFASの優位性.....	241
8章	実装：KFASの使い方.....	242
8-1	この章で使うパッケージ.....	242
8-2	分析の対象となるデータ.....	242
8-3	KFASによる線形ガウス状態空間モデルの推定.....	242
8-4	推定結果の図示.....	246
8-5	KFASによる状態の推定と信頼・予測区間.....	247
8-6	KFASによる予測.....	248
8-7	補足：ローカルレベルモデルにおける予測.....	250
8-8	補足：補間と予測の関係.....	252
9章	実装：変化するトレンドのモデル化.....	254

9-1	この章で使うパッケージ	254
9-2	トレンドと観測値の関係	254
9-3	シミュレーションデータの作成	256
9-4	KFAS によるローカル線形トレンドモデル	258
9-5	補足：モデルの行列表現*	259
9-6	トレンドの図示	260
9-7	補間と予測	262
9-8	ローカル線形トレンドモデルによる予測の考え方	262
9-9	補間と予測結果の図示	263
9-10	ARIMA による予測結果との比較	264
10 章	応用：広告の効果はどれだけ持続するか	266
10-1	この章で使うパッケージ	266
10-2	シミュレーションデータの作成	266
10-3	KFAS による時変係数モデル	268
10-4	変化する広告効果の図示	269
11 章	応用：周期性のある日単位データの分析	271
11-1	この章で使うパッケージ	271
11-2	データの読み込みと整形	271
11-3	祝日の取り扱い	272
11-4	KFAS による基本構造時系列モデル	274
11-5	推定結果の確認	275
11-6	推定結果の図示	276
11-7	周期成分を取り除く	278
11-8	季節調整のメリット	279
第 6 部	状態空間モデルとベイズ推論	280
1 章	一般化状態空間モデルとベイズ推論	281

1-1	一般化状態空間モデル.....	281
1-2	非ガウシアンな観測データ.....	281
1-3	非線形な状態の更新式.....	281
1-4	複雑なモデルの推定方法.....	282
1-5	補足：HMC法とカルマンフィルタの比較.....	283
2章	パラメタと状態の推定：ベイズ推論とHMC法.....	284
2-1	説明の進め方.....	284
2-2	ベイズの定理と事前確率・事後確率の関係.....	285
2-3	ベイズ更新.....	286
2-4	ベイズの定理.....	287
2-5	事前分布と事後分布.....	288
2-6	補足：確率の基本公式.....	288
2-7	確率密度と確率と積分.....	290
2-8	点推定値としてのEAP推定量.....	291
2-9	統計モデルと階層的な確率分布.....	291
2-10	無情報事前分布.....	292
2-11	事後分布の計算例*.....	293
2-12	積分が困難という問題.....	294
2-13	パラメタ推定と乱数生成アルゴリズムの関係.....	295
2-14	乱数生成で取り組む問題.....	296
2-15	乱数生成：メトロポリス法.....	296
2-16	メトロポリス法の問題点.....	297
2-17	効率の良い乱数生成：HMC法.....	298
2-18	用語：MCMC.....	300
3章	実装：Stanの使い方.....	301
3-1	Stanのインストール.....	301

3-2	この章で使うパッケージ.....	302
3-3	シミュレーションデータの作成.....	302
3-4	stan ファイルの記述.....	304
3-5	data ブロックの指定.....	304
3-6	parameters ブロックの指定.....	305
3-7	model ブロックの指定.....	306
3-8	データ生成過程(DGP)と Stan の関係.....	307
3-9	Stan によるローカルレベルモデルの推定.....	307
3-10	結果の出力と収束の判定.....	310
3-11	収束をよくするための調整.....	311
3-12	ベクトル化による効率的な実装.....	313
3-13	乱数として得られた多数のパラメタの取り扱い.....	314
3-14	推定結果の図示.....	316
4 章	応用：複雑な観測方程式を持つモデル.....	318
4-1	この章で使うパッケージ.....	318
4-2	テーマ①最適な捕獲頭数を求めたい.....	318
4-3	データの特徴.....	319
4-4	モデルの構造を決める.....	320
4-5	ポアソン分布+ランダムエフェクト.....	321
4-6	stan ファイルの記述.....	322
4-7	data ブロックの指定.....	323
4-8	parameters ブロックの指定.....	323
4-9	transformed parameters ブロックの指定.....	324
4-10	model ブロックの指定.....	324
4-11	generated quantities ブロックの指定.....	325
4-12	stan によるモデルの推定.....	326

4-13	推定されたパラメタの確認.....	327
4-14	平滑化された個体数の図示.....	327
4-15	検討事項	329
5章	応用：非線形な状態方程式を持つモデル.....	330
5-1	この章で使うパッケージ	330
5-2	テーマ②密度効果をモデル化する	330
5-3	データの特徴	331
5-4	ロジスティック増殖曲線	332
5-5	弱情報事前分布	333
5-6	<code>data</code> ブロックの指定	334
5-7	<code>parameters</code> ブロックの指定.....	335
5-8	<code>transformed parameters</code> ブロックの指定	335
5-9	<code>model</code> ブロックの指定.....	335
5-10	<code>generated quantities</code> ブロックの指定	336
5-11	<code>stan</code> によるモデルの推定	337
5-12	推定されたパラメタの確認.....	338
5-13	平滑化された個体数の図示.....	339
5-14	<code>Stan</code> で推定できる様々なモデルたち.....	340
	参考文献	342
	パッケージ・R 関数一覧	344
	索引	348