

時系列解析 (Time Series Analysis)

Yugo Nakayama

2026 年 2 月 10 日

目次

第 I 部	時系列解析 (Time Series Analysis)	7
第 1 章	序章：時系列解析とは	9
1.1	時系列データの数学的定義	9
1.2	時系列の代表的な例と（数理的な）解釈	10
1.3	時系列解析の目的（数理的視点）	10
1.4	具体データ例：日本円ドル為替レート	11
第 2 章	基礎確率と統計の復習	15
2.1	確率空間・期待値・分散	15
2.2	正規分布と多変量正規分布	16
2.3	最尤推定（MLE）の基本	17
2.4	仮説検定と p 値の意味整理	18
第 3 章	時系列データの特徴と分解	21
3.1	時系列の構成要素	21
3.2	定義例：分解形式の一般系	22
3.3	移動平均によるトレンド抽出	22
3.4	実データ例：日次売上の分解（具体例）	23
3.5	季節要因と正規化（標準化）	23
3.6	簡単な分解アルゴリズム	23
第 4 章	定常過程と自己相関	25
4.1	弱定常過程（wide-sense stationary）の定義	25
4.2	自己共分散関数・自己相関関数（ACF）	26
4.3	ACF の見方と直感解釈	26
4.4	偏自己相関関数（PACF）の定義	27
4.5	具体例：AR(1) と ACF/PACF	27
4.6	数理的性質と定理	27
第 5 章	AR, MA, ARMA モデル	29
5.1	自己回帰モデル：AR(p)	29
5.2	移動平均モデル：MA(q)	29
5.3	ARMA(p, q) モデル：AR と MA の融合	30
5.4	数理的性質と定理	30

第 6 章	ARIMA と季節調整	33
6.1	差分と非定常性	33
6.2	差分によるトレンド対処	34
6.3	季節性と季節差分 (SARIMA)	34
6.4	季節調整の直感と実務の流れ	34
6.5	差分化と定常性に関する性質	34
第 7 章	状態空間モデルとカルマンフィルタ	37
7.1	状態空間表現の定義	37
7.2	カルマンフィルタの仕組み	37
7.3	具体例: ローカルレベルモデル	38
7.4	状態空間モデルの利点	38
7.5	状態空間モデルの理論的性質	39
第 8 章	多変量時系列 (VAR, VECM)	41
8.1	ベクトル自己回帰 (VAR) の基本	41
8.2	VAR の推論と Granger 因果性	41
8.3	共統合と VECM	42
8.4	共統合の基本定理	42
第 9 章	周波数領域の時系列解析 (スペクトル解析)	45
9.1	フーリエ級数と周期の分解	45
9.2	スペクトル密度の基本定理	45
9.3	スペクトル密度の解釈	46
9.4	スペクトル密度の解釈	46
9.5	実務的なスペクトル密度の推定	46
第 10 章	異常検知・構造変化・レジームスイッチ	47
10.1	異常検知: 概要と定義	47
10.2	予測誤差ベースの異常検知	47
10.3	構造変化とレジームスイッチ	47
第 11 章	機械学習と時系列: RNN・LSTM・Transformer など	49
11.1	時系列に特化した機械学習の考え方	49
11.2	RNN・LSTM: 時刻ごとに更新される内部状態	49
11.3	Transformer と時系列予測	50
11.4	統計モデルとの接続と補完	50
第 12 章	実務ワークフローとケーススタディ	51
12.1	実務ワークフローの一般的構成	51
12.2	ケーススタディ 1: 店舗売上予測 (ARIMA と STL)	51
12.3	ケーススタディ 2: センサー異常検知 (状態空間モデル)	51
付録 A	集中不等式の道具箱	53
A.1	Orlicz ノルムと確率変数クラス	53
A.2	sub-Gaussian の同値性 (mgf・尾・モーメント)	54

A.3	sub-exponential と Bernstein 型評価	54
A.4	最大値評価と union bound の定石	55
A.5	Matrix Bernstein (自己共役行列の集中)	55
A.6	例題	56
A.7	演習問題	56

第 I 部

時系列解析 (Time Series Analysis)

第 1 章

序章：時系列解析とは

本章では、時系列データとは何か、またその解析の目的について数理的な定義と直感的な解釈を交えて概説する。

1.1 時系列データの数学的定義

時系列データは、時間 t を添字とする確率変数列 $\{X_t\}_{t \in T}$ として記述される。ここでは、時間集合 T を次のように導入する。

- $T = \mathbb{Z}$: 無限期間のガウス過程など理論寄りの設定
- $T = \{1, 2, \dots, n\}$: 有限長の観測系列（実務では一般的）

ランダム過程 (stochastic process) $\{X_t\}$ の各関数値 $X_t(\omega)$ は、同じ確率空間上の「時刻 t での観測値」であるを考える。

注意 1.1. 実務では、理論的なランダム過程そのものではなく、一回の実現パス「インスタンス系列」

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

を観測するという視点で話すことが多い。

例：日次株価終値

日次株価終値は次のようにモデリングされることがある：

- 各時刻 t は一営業日を表す（週末・休日はスキップ）。
- 観測系列

$$x_1, x_2, \dots, x_n$$

は、「株価終値」または「対数収益率」 $r_t = \log(x_t/x_{t-1})$ として記録される。

この例では、

- $x_t \in \mathbb{R}_{>0}$ (株価)
- 必要なら $r_t \in \mathbb{R}$ (対数収益率)

となる。

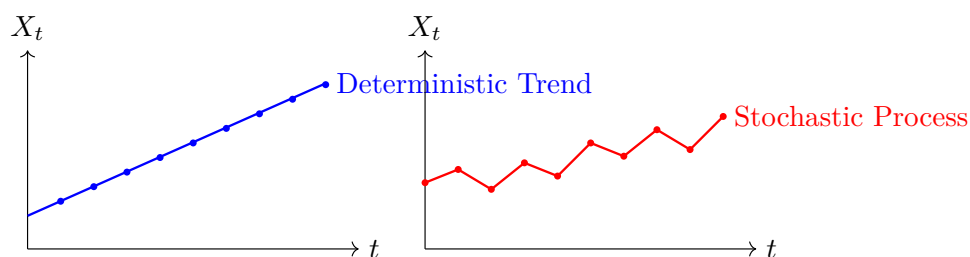


図 1.1 確定的トレンド（左）と確率過程（右）のイメージ

1.2 時系列の代表的な例と（数理的な）解釈

1.2.1 例 1：平均一定＋ノイズ の単純モデル

もっとも簡単な例は、独立同分布ノイズを足したモデル

$$X_t = \mu + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

である。このモデルは、次の特徴を持つ：

- 母平均： $\mathbb{E}[X_t] = \mu$ （全ての t で同じ）
- 自己相関： $\text{Cov}(X_t, X_s) = 0$ ($t \neq s$)（純粋な白ノイズ）

直感：

- 「いつも一定の平均 μ の周りを、ノイズがランダムにふれる」と読める。
- 過去の値 x_{t-1} を見ても、未来の値 x_t を予測できる構造はない（完全な自己相関 0）。

1.2.2 例 2：自己回帰的な単純モデル

次に、自己回帰的な構造を入れたモデル

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

という AR(1) モデルを考える。このモデルでは、

- $|\phi| < 1$ のとき、定常過程となる（後章で詳述）
- 過去の値に「比例して」現時点の値が決まる

直感：

- $\phi > 0$ だと、「前回の値が高いと今も若干高い傾向」
- $\phi < 0$ だと、「値が上下に振動する」ような振る舞い
- $\phi = 0$ に戻れば例 1 と同じ → ノイズだけ

1.3 時系列解析の目的（数理的視点）

時系列解析では、観測された系列 $\mathbf{x} = (x_1, \dots, x_n)$ から次の主要な目的を達成しようとする。

1.3.1 モデル識別（モデル構造の特定）

「どの種類のランダム過程が採用できるか」を見つける。例えば：

- ARMA (p, q) :

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- 自己回帰モデル： $\log(x_t) - \log(x_{t-1}) = r_t = \phi r_{t-1} + \varepsilon_t$

これらのモデルパラメータが定まれば、未来の条件付き分布が記述できる。

1.3.2 予測値と分散の定量化

時点 n までの観測値 $\mathbf{x}_n = (x_1, \dots, x_n)$ を与えたとき、多期間先の予測分布を記述する。例：

- 1 期先予測平均： $\mathbb{E}[X_{n+1} \mid \mathbf{x}_n]$
- 1 期先予測分散： $\text{Var}(X_{n+1} \mid \mathbf{x}_n)$

直感：

- 時系列モデルは「未来の値が完全にわからない」のではなく、「どれくらいの範囲で値が入りそうか」を分散という形で教えてくれる。
- 信頼区間・リスク管理（VaR など）との接続が自然である。

1.3.3 需要予測・異常検知としての応用

時系列解析の「実務的な顔」は

- 売上・需要・トラフィック・電力需要の予測
- スキャンデータ・センサー値からの異常検知

などに現れる。

例：日次の電力需要

- シリーズ： $\{x_t\}_{t=1}^n$ （日ごとの最大需要 or 総消費量）
- 目的：翌日や翌週の最大需要の予測・リスク分散の見積り

その際、モデルを

$$X_t = f_t + Y_t$$

のように分解する考え方がある：

- f_t ：時間とともに変化するトレンド・季節関数
- Y_t ：その時点に残る「ゆらぎ」→ 定常過程のモデル対象

1.4 具体データ例：日本円ドル為替レート

ここでは、日次の日本円／米ドル為替レート（例： $x_t = 1$ 米ドルが何円か）を観測系列とする。

観測された実現値の集合：

$$\{(t, x_t)\}_{t=1}^n$$

と見なせる。

数理的な取り扱い例（対数収益率）

トレンドを除去し、ボラティリティ構造を捉えるために、

$$r_t = \log \left(\frac{y_t}{y_{t-1}} \right)$$

という対数収益率を考えることが多い。

ここで y_t が為替レートの水準価格（円数）であれば：

- r_t ：日次の変動率（exp すると大体パーセント）
- $\mathbb{E}[r_t]$ は「平均リターン」近似
- $\text{Var}(r_t)$ や $\text{Cov}(r_t, r_{t-1})$ は「ボラティリティ構造」の指標

という形で、時系列解析の対象変数が導入される。

練習問題（第 1 章）

演習 1.1（基本的な観測系列の導入）．日次の売上が

$$x_1 = 100, x_2 = 110, x_3 = 115, x_4 = 105$$

であるとする。

1. まず、この系列を

$$\mathbf{x} = (x_1, x_2, x_3, x_4)$$

というベクトルとして書け。

2. 差分系列

$$d_t = x_t - x_{t-1}$$

の各値 (d_2, d_3, d_4) を計算せよ。

3. これらの差分を使って、「売上のトレンド」について直感的にどう解釈できるかの一文を書け。

演習 1.2（単純モデルと期待値）．ランダム過程

$$X_t = 100 + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 25)$$

で観測されたとする。

1. $\mathbb{E}[X_t]$ を求めよ。
2. $\text{Var}(X_t)$ を求めよ。
3. 2 期間先の値 X_{t+2} は、 X_t に比べて「事前にどんな情報量があるか」を直感的に述べよ（数式も含むとよい）。

演習 1.3（自己回帰的な直感）．以下の AR(1) モデルを考えてみる：

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

ここで $\phi \in \{0, 0.3, -0.5\}$ のとき、それぞれ：

1. 母平均 $\mu = \mathbb{E}[X_t]$ を求めよ（一定であればその値を書く）。
2. それぞれの場合で、時系列の振る舞いを「数式＋直感文で」説明せよ（振動するか、なめらかか、ランダム・ノイズ的かなど）。

練習問題の解答（第 1 章）

解答 (問 1). 1. $\mathbf{x} = (100, 110, 115, 105)$

2. 差分 :

$$d_2 = 110 - 100 = 10, \quad d_3 = 115 - 110 = 5, \quad d_4 = 105 - 115 = -10$$

よって $(d_2, d_3, d_4) = (10, 5, -10)$

3. 「最初は増加トレンド (+10→+5) だが、最後に売上が急降下 (-10) しており、トータルとしては増加傾向が弱まった可能性がある」という直感が得られる。

解答 (問 2). 1. $\mathbb{E}[X_t] = 100$

2. $\text{Var}(X_t) = \text{Var}(\varepsilon_t) = 25$

3. X_t と X_{t+2} は互いに独立であり、

$$\text{Cov}(X_t, X_{t+2}) = 0$$

である。したがって、 X_t を観測しても X_{t+2} の予測には「平均、分散、分布形状のどれをとっても」新しい情報はないという意味で、予測能力は持たない。

解答 (問 3). 母平均 :

$$\mu = \mathbb{E}[X_t] = \mathbb{E}[\phi X_{t-1} + \varepsilon_t] = \phi\mu + 0 \implies \mu(1 - \phi) = 0 \implies \mu = 0 \quad (\text{if } |\phi| < 1).$$

よって、定常条件のもとでは、すべてのケースで $\mu = 0$ 。

直感的意味 :

- $\phi = 0$: 独立ノイズだけ $\rightarrow \{X_t\}$ は白ノイズ。
- $\phi = 0.3 > 0$: 前回の値が高いと今も若干高い傾向。なめらかに揺れるが、振動しない形。
- $\phi = -0.5 < 0$: 符号が反対。前回高くて、今日は低くなって、その次はまた高くなる、という振動的な傾向が出るが、分散は一定。

第 2 章

基礎確率と統計の復習

時系列モデルの多くは、正規分布や共分散構造に基づき、また最尤推定などの統計手法を基盤としている。この章では、後続の ARMA モデル・状態空間モデルなどを読むために必要な確率・統計の最低限の道具を復習・整理する。

2.1 確率空間・期待値・分散

2.1.1 確率変数とそのモーメント

まず、確率変数 X を扱うための最小限の設定を持つ。

- 確率空間 (Ω, \mathcal{F}, P)
- 実数値確率変数 $X : \Omega \rightarrow \mathbb{R}$
- 既知の分布（離散 or 連続）

代表的なモーメント量を以下とする。

- 平均（期待値）：

$$\mu = \mathbb{E}[X]$$

- 分散：

$$\sigma^2 = \text{Var}(X) = \mathbb{E}[(X - \mu)^2]$$

- 標準偏差： $\sigma = \sqrt{\text{Var}(X)}$

直感：

- 期待値：「平均的にどこに重心があるか」
- 分散：「分布がどれくらい散らばっているか」

例題 2.1（離散：コイントスで表の確率 0.3 の場合）.

$$X = \begin{cases} 1 & (p = 0.3) \\ -1 & (p = 0.7) \end{cases}$$

このとき

$$\mathbb{E}[X] = 1 \cdot 0.3 + (-1) \cdot 0.7 = -0.4$$

$$\text{Var}(X) = (1 + 0.4)^2 \cdot 0.3 + (-1 + 0.4)^2 \cdot 0.7 \approx 0.84$$

→ 平均は「負寄り」であり、値のバラつきもサイズ 0.84 程度に見えている。

2.1.2 共分散と相関

次に、「2 変数以上」の関係を扱うための量を導入する。2 つの確率変数 X, Y に対して：

- 共分散：

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

- 相関係数：

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

直感：

- $\text{Cov}(X, Y) > 0$: X が大きいとき Y も大きい傾向
- $\text{Cov}(X, Y) < 0$: ひねりがあつて、 X が大きいとき Y は小さい傾向
- $\rho_{XY} \in [-1, 1]$ であり、絶対値が大きいほど「直線的な依存度」が高い

2.2 正規分布と多変量正規分布

時系列解析において、もっとも多用される分布は正規分布（Gaussian）である。

2.2.1 単変量正規分布

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

とは、密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

を持つ分布を意味する。

- $\mathbb{E}[X] = \mu$
- $\text{Var}(X) = \sigma^2$

正規分布の良い性質：

- 線形変換でも正規：もし $X \sim \mathcal{N}(\mu, \sigma^2)$ なら

$$aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$$

- 独立なら和も正規： X, Y 独立で正規であれば

$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

例題 2.2 (ノイズモデル).

$$\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

は、時系列解析における「ベースノイズ」に最もよく使われる仮定である。

2.2.2 多変量正規分布

次は多次元バージョン。 $\mathbf{X} = (X_1, \dots, X_k)^\top$ が多変量正規分布

$$\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

とは、

- $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ (平均ベクトル)
- $\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = (\text{Cov}(X_i, X_j))_{i,j}$ (分散・共分散行列)

であり、密度関数

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)$$

をもつときを言う。

重要なこと：

- 正規分布の部分は、線形変換 $Y = \mathbf{A}\mathbf{X} + \mathbf{b}$ に対しても正規
- 任意の線形結合 $a_1 X_1 + \dots + a_k X_k$ も正規分布に従う

2.3 最尤推定 (MLE) の基本

時系列モデルでは「尤度関数を最大化するパラメータ」を推定する場面が多く、最尤推定のコンセプトを押さえておく必要がある。

2.3.1 尤度関数の定義

観測値が

$$\mathbf{x} = (x_1, \dots, x_n)$$

であり、モデルがパラメータ θ (例: μ, σ^2) に依存するとき、

- 密度関数が独立であれば：

$$f(\mathbf{x} | \theta) = \prod_{t=1}^n f(x_t | \theta)$$

- 尤度関数とはこの右辺を、 θ の関数と見なしたものにすぎない：

$$L(\theta) = f(\mathbf{x} | \theta)$$

- 対数尤度：

$$\ell(\theta) = \log L(\theta) = \sum_{t=1}^n \log f(x_t | \theta)$$

最尤推定 (MLE)：

$$\hat{\theta}_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}} \ell(\theta)$$

は、「観測データが得られる確率 (密度) を一番大きくするようなパラメータ」を指す。

2.3.2 正規分布の例（平均と分散の推定）

$$X_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

のとき、平均 μ と分散 σ^2 を求めよう。

対数尤度：

$$\begin{aligned} \ell(\mu, \sigma^2) &= \sum_{t=1}^n \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x_t - \mu)^2}{2\sigma^2} \right) \right] \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n (x_t - \mu)^2 \end{aligned}$$

極大条件を μ, σ^2 でそれぞれ取ると：

- μ に関する一階条件：

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{t=1}^n (x_t - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{t=1}^n x_t$$

- σ^2 に関する一階条件：

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{t=1}^n (x_t - \mu)^2$$

分母を $(\sigma^2)^2$ として整理し、 μ を $\hat{\mu}$ に置き換えると：

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{\mu})^2$$

直感：

- 標本平均 $\hat{\mu}$ は、「データの重心」
- 分散の MLE $\hat{\sigma}^2$ は、「その重心からの二乗誤差を n で割った値」

2.4 仮説検定と p 値の意味整理

時系列では「これはランダムに見えるか・それ以上の構造か？」という判断をしたい場面が多い。検定と p 値の解釈を整理しておく。

2.4.1 簡単な設定（母平均の検定）

例：

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$$

と仮定する。ただし、分散 σ^2 は既知とする。原仮説 H_0 と対立仮説 H_1 として

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

を考える。検定統計量（標準化平均）：

$$Z = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sigma}$$

を用いると、 H_0 下では $Z \sim \mathcal{N}(0, 1)$ 。5% 有意水準の棄却域： $|Z| > 1.96$ のとき H_0 を棄却。

p 値：

- 実際の標本から計算した z_{obs} に対して

$$p = P_{H_0}(|Z| \geq |z_{\text{obs}}|)$$

と定義される。

p 値の直感的な解釈：

- p 値が小さい（例：0.001）→ 「もし $\mu = \mu_0$ なら、こんな極端な結果が起こる確率がとても小さい」と見せている
- p 値が大きい（0.4 など）→ 「この程度の外れなら十分起こる可能性がある」と読める

練習問題（第 2 章）

演習 2.1（分散と共分散の計算）. 次のデータを考える。

$$(x_1, x_2, x_3) = (1, 4, 3)$$

1. 標本平均 \bar{x} を求めよ。
2. 標本分散の MLE（分母 n ）を求めよ。
3. \bar{x} と「最後のデータ x_3 」の偏差ベクトルを \mathbf{d} とし、共分散のコンセプトを用いて $\text{Cov}(\text{“平均”}, x_3)$ のような直感的な量を解釈せよ。

演習 2.2（多変量正規の直感）. $\mathbf{X} = (X_1, X_2)^\top$ が多変量正規ただし：

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} 4 & 1 \\ 1 & 9 \end{bmatrix}$$

とする。

1. X_1 の平均と分散を求めよ。
2. X_2 の平均と分散を求めよ。
3. X_1, X_2 の共分散と相関係数を求めよ。
4. 相関係数の値から、直感的に「どちらがより強い依存性か」を一文で述べよ。

演習 2.3（最尤推定の自明例）. 独立同分布で

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, 1)$$

とする（分散 $\sigma^2 = 1$ ）。観測された系列は

$$x_1 = 2, x_2 = 3, x_3 = 1$$

であった。

1. 対数尤度 $\ell(\mu)$ を μ だけの関数として式を書け。
2. 上で求めた $\ell(\mu)$ を μ で微分し、 $\hat{\mu}$ を計算せよ。
3. 解釈：なぜ結果は $\hat{\mu} = \frac{1}{3}(2 + 3 + 1)$ と一致すると考えられるか、直感的に説明せよ。

練習問題の解答（第 2 章）

解答（問 1）. 1. 標本平均 $\bar{x} = \frac{8}{3} \approx 2.67$

2. 標本分散 $S^2 = \frac{1}{3}((1 - 2.67)^2 + (4 - 2.67)^2 + (3 - 2.67)^2) \approx 1.56$

解答 (問 2). 1. $\mu_1 = 3, \sigma_1^2 = 4$

2. $\mu_2 = 4, \sigma_2^2 = 9$

3. $\text{Cov} = 1, \rho = \frac{1}{\sqrt{4}\sqrt{9}} = \frac{1}{6}$

第 3 章

時系列データの特徴と分解

時系列データを解析するとき、最初に行うのは「見えるパターンを構成要素ごとに整理する」ことである。トレンド、季節性、周期、残差（ノイズ）に分解することで、予測に必要な構造が見えやすくなり、モデル化の起点が明確になる。この章では、分解の数学的枠組みと、それを導く手計算・簡単な統計モデルを扱う。

3.1 時系列の構成要素

実用的には、観測された時系列 $\{x_t\}_{t=1}^n$ を次のような成分の組み合わせとして考えることが多い。

$$x_t = f_t + s_t + y_t$$

ただし：

- f_t ：トレンド（長期的な傾向）
- s_t ：季節成分（周期的な繰り返し）
- y_t ：残差成分（ノイズやモデル化しない成分）

3.1.1 加法モデルと乗法モデル

具体的な組み合わせ方には、以下のような形が広く用いられる。

■加法モデル（背景は線形）

$$x_t = f_t + s_t + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

このモデルでは、「平均からどれだけずれるか」を各成分の線形和として扱う。

- 絶対スケールが一定
- 季節の振幅もオーダー一定（例：売上がいくら増えても、週末増加分の絶対値は変わらない）

■乗法モデル（背景は非線形）

$$x_t = f_t \cdot s_t \cdot (1 + \varepsilon'_t)$$

ここで ε'_t は小ノイズ。対数変換すると

$$\log x_t = \log f_t + \log s_t + \log(1 + \varepsilon'_t) \approx \log f_t + \log s_t + \varepsilon_t$$

となるため、実際には「対数取った系列を加法分解する」と考えられる。乗法モデルは「売上の絶対値が大きくなるほど、季節の差も大きくなる」ような場合に相性がよい。

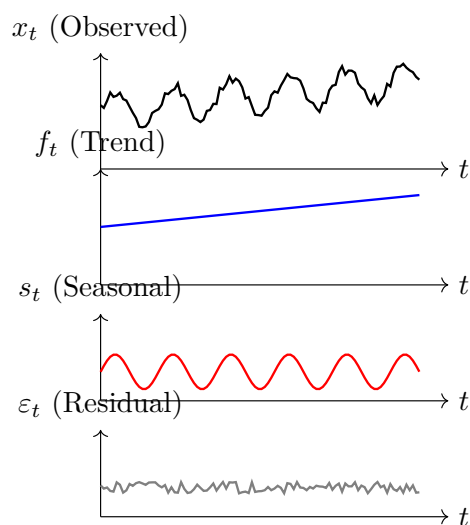


図 3.1 時系列分解のイメージ（上から：原系列、トレンド、季節成分、残差）

3.2 定義例：分解形式の一般系

より一般的に書くと、

$$x_t = g_t + z_t + \varepsilon_t$$

ここで：

- g_t ：長期トレンド＋周期性など
- z_t ：季節性など
- ε_t ：均質なノイズ（定常変動）

というパラメトリック／非パラメトリックな「構造成分」を明示的に仮定する。尤も単純な例が：

$$x_t = a + bt + c_t + \varepsilon_t$$

- $a + bt$ ：線形トレンド
- c_t ：周期成分（例：週周期）
- ε_t ：残差

3.3 移動平均によるトレンド抽出

非パラメトリックかつ直感的なトレンド抽出手段として、移動平均（moving average）が広く用いられる。

3.3.1 定義

奇数 $2k + 1$ に対して、時点 t の移動平均：

$$\bar{x}_t^{(k)} = \frac{1}{2k + 1} \sum_{j=-k}^k x_{t+j}$$

直感：

- 長さ $2k + 1$ の「窓」でだけ平均を取る

- 窓が大きいほどトレンドに近くなり、細かい揺らぎが削ぎ取られる
- 窓が小さいほど、もとの系列に近いがノイズも多い

例：日次売上 $\{x_t\}$

- 季節（週周期）が ≈ 7 日サイクルなら、
- $k = 1 : 3$ 日平均 \rightarrow トレンド抽出は弱く、近視的な平滑化
- $k = 3 : 7$ 日平均 \rightarrow 週周期がかなり滑らかになる
- $k = 10 : 21$ 日平均 \rightarrow 長期トレンド・推移を拾いやすくなる

3.4 実データ例：日次売上の分解（具体例）

3.4.1 データ設定

仮想的な日次売上データを次のように生成する。

$$x_t = 100 + 0.5t + 5 \sin\left(\frac{2\pi t}{7}\right) + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2^2)$$

ここで：

- 線形トレンド： $100 + 0.5t$
- 週周期： $\sin(2\pi t/7)$ （1 週間周期）
- ノイズ： $\varepsilon_t \sim \mathcal{N}(0, 4)$

数学的解釈と観察：時点 t で観測される値 x_t は、「線形トレンド（長期）＋周期（短期）＋ノイズ（瞬間）」の重ね合わせである。

3.5 季節要因と正規化（標準化）

分解の「季節成分」を取り出すときは、「平均からの偏差」として考えるのが近代的な処理の基本になる。

3.5.1 週周期の標準偏差偏差表記

週ごとのデータを次のように思う：

- 各曜日ごとの「偏差」

$$d_j = (\text{曜日 } j \text{ の平均}) - \text{全体の平均}$$

このとき、

- $d_j > 0$ ：その曜日は「全体より高い」
- $d_j < 0$ ：「全体より低い」

3.6 簡単な分解アルゴリズム

最後に、「実務で行う最小限の分解プロセス」を数理的に整理しておく。

1. トレンド近似 \hat{f}_t ：移動平均などで推定。
2. トレンド除去系列： $d_t^{(1)} = x_t - \hat{f}_t$ 。

3. 季節成分の抽出: 曜日ごとに平均を取るなどして $\hat{s}_t = \bar{d}_{\text{曜日}(t)}$ とする。

4. 残差: $\hat{\varepsilon}_t = x_t - \hat{f}_t - \hat{s}_t$ 。

このとき各成分が満たすべき条件:

- $\frac{1}{7} \sum_{j=1}^7 \hat{s}_j = 0$: 季節成分の平均が 0
- $\hat{\varepsilon}_t$ は近似的に「均質なノイズ」とみなせること

練習問題 (第 3 章)

演習 3.1 (移動平均の計算). ある商品の日次売上が以下であるとする。

$$x_1 = 100, x_2 = 110, x_3 = 105, x_4 = 120, x_5 = 115, x_6 = 100$$

1. 3 点移動平均 ($k = 1$) の、時点 2: $\bar{x}_2^{(1)}$ 、時点 4: $\bar{x}_4^{(1)}$ を求めよ。
2. 5 点移動平均 ($k = 2$) の、時点 3: $\bar{x}_3^{(2)}$ を求めよ。
3. どのウィンドウの方が「トレンドに近づきやすいか」「ノイズに弱いかな」を一文で論じよ。

演習 3.2 (季節成分と直感). 以下は、ある店舗の 1 週間の日別売上平均である (月～日)。

$$c_{\text{月}} = -3, c_{\text{火}} = -2, c_{\text{水}} = -1, c_{\text{木}} = 0, c_{\text{金}} = 1, c_{\text{土}} = 5, c_{\text{日}} = 4$$

1. 季節成分の平均 \bar{c} を求めよ。
2. この平均が 0 でないと、分解モデル $x_t = f_t + s_t + \varepsilon_t$ で何が不便なのか、直感的に一文で説明せよ。

演習 3.3 (パターンの数理解的解釈). 次のモデルを考える:

$$x_t = 50 + 0.3t + 2 \sin\left(\frac{2\pi t}{4}\right) + \varepsilon_t$$

ただし $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ 。

1. トレンド成分 f_t を答えよ。
2. 周期成分の周期は何か?
3. 正弦波成分の振幅 2 は、売上価格を直感的に「どれくらい振動させる効果を持つ」と読めるか、一文で述べよ。

練習問題の解答 (第 3 章)

解答 (問 1). 1. $\bar{x}_2^{(1)} = 105, \bar{x}_4^{(1)} \approx 113.33$

2. $\bar{x}_3^{(2)} = 110$

3. 3 点の方がノイズに弱くトレンド追従性が低い一方、5 点の方が平滑化が強くトレンドに近い。

解答 (問 2). 1. $\bar{c} \approx 0.57$

2. 平均が 0 でないと、トレンド成分と季節成分の区別 (定数項の帰属) が曖昧になる。

解答 (問 3). 1. $f_t = 50 + 0.3t$

2. 周期 4

3. 平均売上を上下に合計約 4 単位変動させる効果を持つ。

第 4 章

定常過程と自己相関

時系列解析では、「定常性」という性質を仮定することで、モデルの構成・推論が劇的に扱いやすくなる。また、その性質を実際に見るための道具が、自己相関関数（ACF）と偏自己相関関数（PACF）である。この章では、弱定常過程の定義、ACF/PACF の定義・計算・直感解釈、そして実データ観察の方法を復習する。

4.1 弱定常過程（wide-sense stationary）の定義

4.1.1 定義（確率的）

無限ランダム過程 $\{X_t\}$ が弱定常であるとは、次の条件を満たすこととする。

1. 平均が時間によらず一定：

$$\mathbb{E}[X_t] = \mu \quad (\text{全ての } t)$$

2. 分散が有界で一定：

$$\text{Var}(X_t) = \gamma_0 < \infty$$

3. 任意の時間差 h に対して、共分散が時間差のみに依存する：

$$\text{Cov}(X_t, X_{t+h}) = \gamma(h)$$

このとき $\gamma(h)$ を自己共分散関数と呼び、これを使って自己相関関数（ACF）を

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\text{Cov}(X_t, X_{t+h})}{\text{Var}(X_t)}$$

と定義する。

よくある誤解と直感

- 「定常＝静止している」ではない → むしろ、平均・分散・共分散が時間的な“構造”として一定で、そのうえでランダムに揺れる。
- 例：トレンド（大まかに上昇）がある系列は弱定常でない。 → 差分を取ったり、トレンド部分を除去して「残った揺らぎ」を定常と考えることが多い。

4.2 自己共分散関数・自己相関関数 (ACF)

4.2.1 推定版 (標本 ACF)

実際には有限系列 (x_1, \dots, x_n) しか観測できない。このとき、標本自己共分散は通常次のように推定される：

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x})$$

ただし、数理的には割る数が n か $n - |h|$ かで流派が分かれるが、大まかな振る舞いは同じ。

標本自己相関：

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

4.3 ACF の見方と直感解釈

4.3.1 例 1：白ノイズ ($\gamma(h) = 0$ ($h \neq 0$))

$$X_t = \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

このとき：

- $\gamma(h) = 0$ ($h \neq 0$)
- $\rho(h) = 0$ ($h \neq 0$)

→ ACF プロットでは $h = 0$ のみが 1 で、それ以外は 0 のような振る舞い（実際には短命なゆらぎは観測されるが）。直感：「過去の値を見ても、未来の値を予測できる構造はない」。

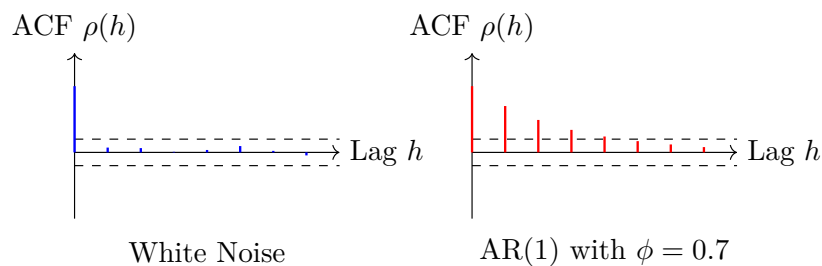


図 4.1 自己相関関数 (ACF) の例：白ノイズ (左) と AR(1) 過程 (右)

4.3.2 例 2：AR(1) (ϕ あり)

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \quad (|\phi| < 1)$$

ACF 理論式：

$$\rho(h) \propto \phi^{|h|}$$

$\phi > 0$ なら高いラグでも正の相関が残り、 ϕ が大きければ「過去に近い値からの影響が長く残る」。

4.4 偏自己相関関数 (PACF) の定義

自己相関 $\rho(h)$ には「中間ラグの影響」が混ざっているため、「現在値との直接的な関係」だけを取り出したければ、偏自己相関関数 (PACF) を使う。

ラグ h の偏自己相関 $\alpha(h)$ は、現在 X_t と X_{t-h} の間に $X_{t-1}, \dots, X_{t-h+1}$ を“条件付け”たうえでの「残りの相関」である。

統計的には、

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_h X_{t-h} + \varepsilon_t$$

という $\text{AR}(h)$ モデルを仮定したとき、 α_h がラグ h の偏自己相関である。

4.5 具体例 : $\text{AR}(1)$ と ACF/PACF

$$X_t = 0.6X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, 1)$$

- $\text{ACF} : \rho(h) \approx 0.6^{|h|}$
- $\text{PACF} : \text{ラグ } 1 \text{ で } 0.6, \text{ 以降は } 0$

4.6 数理的性質と定理

本章で導入した定常過程、特に AR モデルについて、その存在と性質を保証する重要な定理をここで整理する。

定理 4.1 (弱定常 $\text{AR}(p)$ の存在定理). $\tilde{\phi}(\mathbf{B})X_t = \varepsilon_t$ で表される $\text{AR}(p)$ モデルにおいて、以下の条件を仮定する：

1. ノイズは $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ である。
2. 特性多項式

$$\tilde{\phi}(z) = 1 - \sum_{j=1}^p \phi_j z^j$$

のすべての根 z が単位円外にある (すなわち $|z| > 1$)。

このとき、この $\text{AR}(p)$ モデルは弱定常なガウス過程として一意に存在する。さらに、その解は因果的 (Causal) な $\text{MA}(\infty)$ 表現

$$X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}, \quad \sum_{k=0}^{\infty} |\psi_k| < \infty$$

を持つ。

証明. この証明は、特性多項式の逆数展開による $\text{MA}(\infty)$ 表現の構成と、その級数の収束・定常性の確認によって行われる。

Step 1 : 特性多項式と $\text{MA}(\infty)$ 表現

$\text{AR}(p)$ モデル $\tilde{\phi}(\mathbf{B})X_t = \varepsilon_t$ を形式的に書き直すと、

$$X_t = \frac{1}{\tilde{\phi}(\mathbf{B})} \varepsilon_t$$

となる。ここで複素関数 $f(z) = 1/\tilde{\phi}(z)$ を考える。仮定より $\tilde{\phi}(z) \neq 0$ for $|z| \leq 1$ であるため、 $f(z)$ は単位円を含む領域で正則である。したがって、テイラー展開（あるいはローラン展開）により

$$\frac{1}{\tilde{\phi}(z)} = \sum_{k=0}^{\infty} \psi_k z^k$$

とべき級数展開でき、その係数列 $\{\psi_k\}$ は一意に定まる。さらに収束半径が 1 より大きいため、係数は指数的に減衰し、絶対収束条件 $\sum_{k=0}^{\infty} |\psi_k| < \infty$ を満たす。

Step 2 : 弱定常性の確認

構成した $X_t = \sum_{k=0}^{\infty} \psi_k \varepsilon_{t-k}$ について、

1. 平均 : $\mathbb{E}[X_t] = \sum_{k=0}^{\infty} \psi_k \mathbb{E}[\varepsilon_{t-k}] = 0$ (一定)
2. 共分散 : $\gamma(h) = \text{Cov}(X_t, X_{t+h}) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+h}$

共分散 $\gamma(h)$ は t に依存せずラグ h のみの関数となり、かつ級数の絶対収束性より有限値を持つ。以上より、この表現は弱定常過程の定義を満たす。 \square

定理 4.2 (AR(p) の指数的自己共分散減衰). AR(p) モデルが上記の定常性条件（特性根がすべて単位円外）を満たすとき、その自己共分散関数 $\gamma(h)$ は $h \rightarrow \infty$ で指数的に減衰する。すなわち、ある定数 $C > 0$ と $0 < r < 1$ が存在して

$$|\gamma(h)| \leq Cr^{|h|}$$

が成り立つ。

練習問題（第 4 章）

演習 4.1 (定常性と非定常性の判断). 系列 $\{X_t\}$ が次のように定義される：

1. $X_t = 10 + 2t + \varepsilon_t$
2. $X_t = \mu + \varepsilon_t$

各モデルに対して、平均 $\mathbb{E}[X_t]$ が時間 t に依存するか否かを述べよ。

演習 4.2 (ACF の計算と直感). 次の 5 点の系列がある： $x = (1, 3, 2, 4, 3)$ 。

1. 標本平均 \bar{x} を求めよ。
2. 標本自己相関 $\hat{\rho}(1)$ を求めよ。

練習問題の解答（第 4 章）

解答 (問 1). 1. $X_t = 10 + 2t$ は平均が t に依存するため非定常。
2. $X_t = \mu$ は平均・分散が一定で、時間は無関係なので弱定常。

解答 (問 2). 1. $\bar{x} = 2.6$
2. $\hat{\gamma}(0) \approx 1.04, \hat{\gamma}(1) \approx 0.296 \implies \hat{\rho}(1) \approx 0.285$

第 5 章

AR, MA, ARMA モデル

時系列解析の古典的フレームワークの中心は、自己回帰（AR）、移動平均（MA）、そして両者の混合である ARMA モデルである。本章では、それぞれのモデルの定義・特徴・自己共分散構造・安定性・可逆性、および ACF/PACF との関係を数理的に整理し、直感的な解釈例を示す。

5.1 自己回帰モデル：AR(p)

5.1.1 AR(1) の定義

もっともシンプルな自己回帰モデルは AR(1) である：

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

- ϕ は自己回帰係数（ラグ係数）
- ε_t は i.i.d. 白ノイズ

自己回帰の意味：「現在の値 X_t は、過去の値 X_{t-1} の線形再帰＋ノイズ」という構造を持っている。

5.1.2 AR(p) の一般形

AR(p) モデルは次のように定義される：

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

演算子記号 \mathbf{B} を用いて、

$$\tilde{\phi}(\mathbf{B})X_t = \varepsilon_t, \quad \tilde{\phi}(\mathbf{B}) = 1 - \sum_{i=1}^p \phi_i \mathbf{B}^i$$

と書ける。安定条件は特性方程式の根が単位円外にあることである。

5.2 移動平均モデル：MA(q)

MA モデルは、（潜在的な）過去のノイズの線型和として現在値をモデル化する。

5.2.1 MA(1) の定義

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$$

- θ は移動平均係数

- 平均は常に $\mathbb{E}[X_t] = 0$
- 自己共分散は有限範囲しか値をもたない（カットオフ現象）

MA(q) では $\rho(h) = 0$ ($h > q$) となる。

5.3 ARMA(p, q) モデル：AR と MA の融合

AR(p) と MA(q) を組み合わせて

$$\tilde{\phi}(\mathbf{B})X_t = \tilde{\theta}(\mathbf{B})\varepsilon_t$$

と定義されるモデルを ARMA(p, q) という。AR は記憶を長く保持し、MA は短期的なショックを記述する。これらを組み合わせることで、柔軟なモデリングが可能になる。

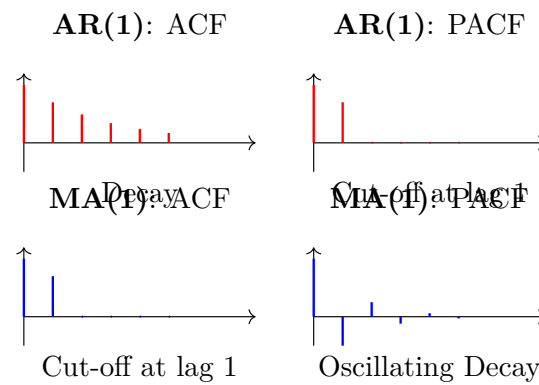


図 5.1 AR(1) と MA(1) の ACF/PACF の特徴比較

5.3.1 モデル選択：AIC/BIC

次数 p, q の決定には情報量基準が用いられる。

$$\text{AIC}(p, q) = -2\ell(\hat{\theta}) + 2k$$

$$\text{BIC}(p, q) = -2\ell(\hat{\theta}) + k \log n$$

AIC/BIC が最小となるモデルが候補として選ばれる。

5.4 数理的性質と定理

ARMA モデルにおける可逆性や表現定理など、モデルの正当性を支える数学的性質を整理する。

定理 5.1 (MA(q) 過程の可逆性 (Invertibility)). MA(q) モデル

$$X_t = \tilde{\theta}(\mathbf{B})\varepsilon_t = \left(1 + \sum_{j=1}^q \theta_j \mathbf{B}^j\right) \varepsilon_t$$

において、多項式 $\tilde{\theta}(z) = 1 + \sum_{j=1}^q \theta_j z^j$ のすべての根 z が単位円外 ($|z| > 1$) にあるならば、この直交過程は可逆である。すなわち、 ε_t を過去の X_t, X_{t-1}, \dots の無限線形和として一意に表現できる：

$$\varepsilon_t = \sum_{k=0}^{\infty} \pi_k X_{t-k}, \quad \sum_{k=0}^{\infty} |\pi_k| < \infty$$

定理 5.2 (MA(q) の有限記憶性). MA(q) 過程の自己共分散関数 $\gamma(h)$ は、ラグ q を超えると完全に 0 になる (Cut-off 性質)。

$$\gamma(h) = 0 \quad (\text{for } |h| > q)$$

証明. MA(q) の定義 $X_t = \sum_{j=0}^q \theta_j \varepsilon_{t-j}$ (ただし $\theta_0 = 1$) より、

$$\gamma(h) = \mathbb{E} \left[\left(\sum_{j=0}^q \theta_j \varepsilon_{t-j} \right) \left(\sum_{k=0}^q \theta_k \varepsilon_{t+h-k} \right) \right]$$

ε_t の独立性から、添字が一致する項のみが残り、それ以外は消える。 $|h| > q$ のとき、共通する時間のノイズ項が存在しないため、期待値は 0 となる。□

定理 5.3 (Wold の分解定理 (Wold Decomposition)). 任意の平均 0 の弱定常過程 $\{X_t\}$ は、確定的成分 V_t と非決定論的成分 (MA(∞)) の和として一意に分解できる。

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} + V_t$$

ここで、 $\psi_0 = 1, \sum \psi_j^2 < \infty$ 、 ε_t は白色雑音 (予測誤差系列)、 V_t は過去の X から完全に予測可能な成分である。

練習問題 (第 5 章)

演習 5.1 (AR(1) と MA(1) の比較). 次の 2 つのモデルを考える。

1. AR(1) : $X_t = 0.7X_{t-1} + \varepsilon_t$
2. MA(1) : $X_t = \varepsilon_t + 0.7\varepsilon_{t-1}$

それぞれの理論 ACF $\rho(h)$ を $h = 0, 1, 2$ について求めよ。

演習 5.2 (ARMA(1,1) の自己共分散). モデル $X_t = 0.6X_{t-1} + \varepsilon_t + 0.3\varepsilon_{t-1}$ を考える。分散 $\gamma(0)$ を理論的に表せ。

練習問題の解答 (第 5 章)

解答 (問 1). • AR(1): $\rho(0) = 1, \rho(1) = 0.7, \rho(2) = 0.49$

- MA(1): $\rho(0) = 1, \rho(1) = \frac{0.7}{1+0.7^2} \approx 0.47, \rho(2) = 0$

第 6 章

ARIMA と季節調整

多くの実務的な時系列は「定常ではない」＝平均や分散が時間とともに変化する傾向がある。こうした非定常系列に対しては、差分操作によって定常な系列を導いたうえで、第 5 章で扱った ARMA モデルを適用するという方針が有効になる。本章では $ARIMA(p, d, q)$ モデルと季節 ARIMA (SARIMA) の構造・統計的意味・実データへの当てはめ方を、数学的に明瞭に整理する。

6.1 差分と非定常性

6.1.1 一階差分の定義

時系列 $\{y_t\}$ が、例えば上昇トレンドを持っている場合、

$$y_t \approx \mu + \beta t + \varepsilon_t$$

といった非定常構造を考えることが多い。このとき、直接 ARMA を当てはめると推定が不安定になりやすい。そこで「一階差分」を導入する：

$$x_t = y_t - y_{t-1} = \Delta y_t$$

ここで $\Delta = 1 - \mathbf{B}$ と演算子で書けば

$$x_t = (1 - \mathbf{B})y_t$$

差分系列 x_t はトレンドが除去されており、しばしば定常あるいは弱定常に近い振る舞いを見せる。

6.1.2 d 階差分・ $ARIMA(p, d, q)$ の定義

$$\tilde{\phi}(\mathbf{B})\Delta^d y_t = \tilde{\theta}(\mathbf{B})\varepsilon_t$$

ただし、

- d : 差分階数 (トレンド除去に必要とされる次数)
- $\tilde{\phi}(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \cdots - \phi_p \mathbf{B}^p$
- $\tilde{\theta}(\mathbf{B}) = 1 + \theta_1 \mathbf{B} + \cdots + \theta_q \mathbf{B}^q$

ARIMA は「原系列の差分系列に ARMA を当てはめる統一フレームワーク」である。

6.2 差分によるトレンド対処

6.2.1 線形トレンドと一階差分

原系列が次のように線形トレンドを持つとしよう：

$$y_t = a + bt + \varepsilon_t$$

一階差分：

$$x_t = b + (\varepsilon_t - \varepsilon_{t-1})$$

このとき、平均は b 、分散は一定となる。実務では、散布図を見て、まっすぐ底上げなら一階差分 ($d = 1$)、加速度的増加なら二階差分 ($d = 2$) を選択する。

6.3 季節性と季節差分 (SARIMA)

6.3.1 季節差分の導入

多くの実データは「季節周期」を持つ（例：週周期、年周期）。これを統計的に扱うとき、季節差分を使う。

例：7 日周期（週周期）があるとき、

$$x_t^{(s)} = \Delta y_t, \quad x_t^{(s)} = y_t - y_{t-7}$$

と定義される。 $\Delta_s = 1 - \mathbf{B}^7$ とすれば $x_t^{(s)} = (1 - \mathbf{B}^7)y_t$ となる。

6.3.2 SARIMA : 季節 ARIMA モデル

季節成分を取り入れた一般化モデルとして $\text{SARIMA}(p, d, q \times P, D, Q)$ がよく用いられる。

$$(1 - \phi\mathbf{B})(1 - \Phi\mathbf{B}^s)\Delta^d\Delta_s^D y_t = (1 + \theta\mathbf{B})(1 + \Theta\mathbf{B}^s)\varepsilon_t$$

演算子多項式解釈：

- $(1 - \phi\mathbf{B})$: 通常の AR(1) 部分
- $(1 - \Phi\mathbf{B}^s)$: s ステップごとの季節 AR 項
- $\Delta^d\Delta_s^D$: トrend・季節性の同時除去

6.4 季節調整の直感と実務の流れ

1. 原系列 $\{y_t\}$ を描画（トレンド・季節・外れ）を確認
2. ADF や目視で「差分例: $d = 0$ か $d = 1$ か」を決める
3. GSTL / 移動平均・別途分解で「季節を取り出し」
4. 残された定常系列 $\{x_t\}$ に対して ARMA / ARIMA パラメータ推定
5. 予測を行う（差分系列の累積和）

6.5 差分化と定常性に関する性質

差分操作 (Differencing) がなぜ非定常系列を定常化できるのか、その数学的な保証を与える定理を整理する。

定理 6.1 (差分による多項式トレンドの除去). 時系列 y_t が k 次の多項式トレンド m_t と定常な誤差項 z_t の和で表されるとする:

$$y_t = m_t + z_t, \quad m_t = \sum_{j=0}^k \beta_j t^j$$

ここで z_t は弱定常過程とする。このとき、 k 階差分系列

$$x_t = \Delta^k y_t = (1 - \mathbf{B})^k y_t$$

は弱定常過程となる (平均が一定、分散が有界)。

証明. 階差演算子 Δ の性質として、 $\Delta^j t^j$ は $(j-1)$ 次の多項式となる (微分と同様)。したがって、 $\Delta^k m_t$ は定数 (0 次多項式) となる。一方、定常過程 z_t の線形変換である $\Delta^k z_t$ もまた定常性を保つ (定理)。よって、 $\Delta^k y_t = \text{定数} + \Delta^k z_t$ は平均一定の定常過程となる。 \square

定理 6.2 (過剰差分 (Over-differencing) の問題). もし y_t が既に定常であるにもかかわらず差分をとった場合 ($x_t = \Delta y_t$)、その可逆性が失われるか、あるいは人工的な単位根が MA 部分に導入される。具体的には、

$$y_t = \varepsilon_t \implies \Delta y_t = \varepsilon_t - \varepsilon_{t-1}$$

となり、これは MA(1) 過程 $x_t = (1 - \mathbf{B})\varepsilon_t$ となるが、MA 多項式 $\theta(z) = 1 - z$ は単位円上に根 $z = 1$ を持つため、可逆ではない。これによりパラメータ推定が不安定になる。

練習問題 (第 6 章)

演習 6.1 (差分操作とトレンド). 原系列が $y_t = 3t + 5 + \varepsilon_t$ であるとする。

1. 一階差分系列 x_t を具体的な式で書け。
2. x_t の期待値 $\mathbb{E}[x_t]$ を求めよ。
3. x_t は定常か否か、理由を一文で述べよ。

演習 6.2 (ARIMA(p, d, q) の直感). モデル $\Delta y_t = 0.5\Delta y_{t-1} + \varepsilon_t + 0.3\varepsilon_{t-1}$ を考える。

1. 差分系列 Δy_t について、どのタイプの ARMA モデルか。
2. 原系列 y_t にどんな「トレンド的特徴」があると言えるか。
3. ARIMA パラメータ (p, d, q) を答えよ。

演習 6.3 (季節差分と SARIMA (概念問題)). ある店舗の日次売上 y_t が、月～金は一定、週末は高い、全体は年 5% 増とする。

1. 季節差分 $x_t^{(s)} = y_t - y_{t-7}$ を記述せよ。
2. 上の $x_t^{(s)}$ はどのような振る舞いをするか。
3. SARIMA を使う利点を一文で述べよ。

練習問題の解答 (第 6 章)

解答 (問 1). 1. $x_t = 3 + (\varepsilon_t - \varepsilon_{t-1})$

2. $\mathbb{E}[x_t] = 3$

3. 定常である (平均一定、分散一定、共分散はラグ依存のみ)。

- 解答 (問 2). 1. ARMA(1,1)
2. 増加トレンドかつ短期的記憶を持つ。
3. ARIMA(1,1,1)

- 解答 (問 3). 1. $x_t^{(s)} = y_t - y_{t-7}$
2. トレンドやレベル差が相殺され、週末効果のみが強く現れる。
3. 季節的な周期変動とトレンドを同時にモデル化できる。

第 7 章

状態空間モデルとカルマンフィルタ

これまでの ARIMA モデルは観測された系列のみを数理的に扱ったが、状態空間モデルは「観測できない真の状態 (State)」と「誤差を含んだ観測値 (Observation)」を分けてモデル化する。このアプローチは、構造変化や欠損値への対応が柔軟で、物理的な動態 (ロボティクス) から経済構造の推定まで幅広く利用される。

7.1 状態空間表現の定義

状態空間モデルは、次の 2 つの基本的な方程式で記述される (線形ガウスモデルの場合)。

7.1.1 状態方程式 (システム方程式)

$$\alpha_t = \mathbf{T}_t \alpha_{t-1} + \mathbf{R}_t \eta_t, \quad \eta_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$$

- α_t : 時刻 t における不可視の状態ベクトル。
- \mathbf{T}_t : 状態の遷移行列 (State Transition Matrix)。
- η_t : システムノイズ (工程ノイズ)。プロセスの不確実性を表す。

7.1.2 観測方程式

$$y_t = \mathbf{Z}_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, H_t)$$

- y_t : 実際に観測される値 (スカラまたはベクトル)。
- \mathbf{Z}_t : 観測行列。状態がどのように観測値に反映されるかを示す。
- ε_t : 観測ノイズ。測定誤差や一時的な揺らぎを表す。

数理的直感: 「真の価値 (状態)」は刻一刻と変化しているが、私たちはそれを「ノイズの乗った不完全なセンサー (観測)」を通してしか見ることができない。

7.2 カルマンフィルタの仕組み

カルマンフィルタは、新しい観測値 y_t を得るたびに、状態の推定値を再帰的に更新するアルゴリズムである。

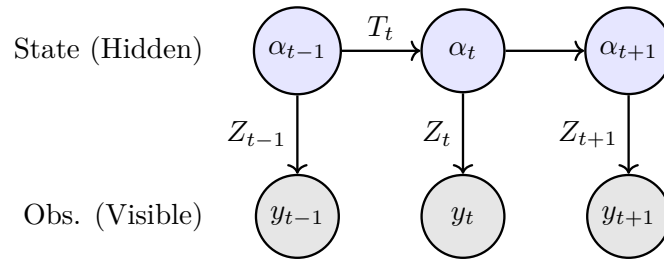


図 7.1 状態空間モデルのグラフィカルモデル表現

7.2.1 予測 (Prediction) ステップ

前時点までの情報に基づき、現時点の状態を予測する。

- 状態予測 : $\mathbf{a}_{t|t-1} = \mathbf{T}_t \mathbf{a}_{t-1|t-1}$
- 分散予測 : $\mathbf{P}_{t|t-1} = \mathbf{T}_t \mathbf{P}_{t-1|t-1} \mathbf{T}_t^\top + \mathbf{Q}_t$

7.2.2 更新 (Filtering) ステップ

予測値と実際の観測値の「ズレ (イノベーション)」を用いて、推定値を修正する。

- イノベーション : $v_t = y_t - \mathbf{Z}_t \mathbf{a}_{t|t-1}$
- カルマンゲイン : $\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top (\mathbf{Z}_t \mathbf{P}_{t|t-1} \mathbf{Z}_t^\top + H_t)^{-1}$
- 状態の修正 : $\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{K}_t v_t$

数理的直感 : カルマンゲイン \mathbf{K}_t は、「予測」と「観測」のどちらをどれだけ信じるかを決める重みである。観測ノイズ H_t が非常に大きい (観測が信じられない) 場合、ゲインは小さくなり、予測値に頼った修正が行われる。

7.3 具体例 : ローカルレベルモデル

最も単純な状態空間モデルとして、トレンドがランダムウォークする「ローカルレベルモデル」を考える。

$$\begin{cases} \alpha_t = \alpha_{t-1} + \eta_t, & \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \\ y_t = \alpha_t + \varepsilon_t, & \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{cases}$$

- σ_η^2 が大きい : トレンドが激しく動く。追従性は高いがノイズに敏感。
- σ_ε^2 が大きい : 観測誤差が激しい。推定されるトレンドは滑らか (鈍感) になる。

7.4 状態空間モデルの利点

- 欠損値への強さ : 観測値 y_t が欠けていても、予測ステップ (状態方程式) のみを繰り返すことで状態を推定し続けられる。
- 時変パラメータ : ARIMA では係数が一定であることを仮定するが、状態空間モデルでは回帰係数そのものを「状態」と見なすことで、時間とともに変化する関係性をモデル化できる。

7.5 状態空間モデルの理論的性質

定理 7.1 (カルマンフィルタの最適性 (Optimality of Kalman Filter)). 線形ガウス状態空間モデル (各ノイズが正規分布に従う) において、カルマンフィルタによって計算される状態推定値 $\mathbf{a}_{t|t} = \mathbb{E}[\alpha_t | Y_t]$ (ただし $Y_t = \{y_1, \dots, y_t\}$) は、平均二乗誤差

$$\mathbb{E}[||\alpha_t - \hat{\alpha}_t||^2]$$

を最小にするという意味で、最小分散不偏推定量 (MVUE) を与える。また、ノイズがガウス分布に従わない場合でも、線形推定量のクラスの中では最小分散推定 (BLUE) を与える。

定理 7.2 (ローカルレベルモデルの単位根性質). ローカルレベルモデル

$$y_t = \mu_t + \epsilon_t, \quad \mu_t = \mu_{t-1} + \eta_t$$

において、観測値 y_t の一階差分 Δy_t は、MA(1) 過程に従う (ARIMA(0,1,1) と等価)。

$$\Delta y_t = (1 + \theta \mathbf{B})\xi_t$$

ここで ξ_t は新たなホワイトノイズであり、 θ は信号対雑音比 $q = \sigma_\eta^2 / \sigma_\epsilon^2$ に依存して決まる (q が大きいほど θ は 0 に近づく)。これは、状態空間モデルと ARIMA モデルの理論的な架け橋となる重要な性質である。

練習問題 (第 7 章)

演習 7.1 (カルマンゲインの計算). 以下のローカルレベルモデルを考える。予測誤差の分散 $P_{t|t-1} = 10$ 、観測ノイズの分散 $H_t = 30$ 、観測行列 $Z_t = 1$ とする。

1. カルマンゲイン K_t を計算せよ。
2. 観測値 y_t が予測値 \hat{y}_t より 4 大きかった場合、状態の推定値はどう修正されるか。
3. 観測ノイズ H_t が 1000 に激増した場合、修正量はどう変化するか。

演習 7.2 (モデルの構造理解). 次の状態方程式が与えられている。

$$\begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_{\mu,t} \\ \eta_{\beta,t} \end{bmatrix}$$

1. μ_t を式で表せ。
2. このモデルにおける「傾き β_t 」はどのような挙動を示すか。
3. ARIMA の文脈ではどの「 d (差分)」に相当するか。

練習問題の解答 (第 7 章)

解答 (問 1). 1. $K_t = \frac{10}{10+30} = 0.25$

2. 修正量は $0.25 \times 4 = 1.0$ 。

3. $K_t \approx 0.01$ となり、修正量は激減する (観測値を無視)。

解答 (問 2). 1. $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_{\mu,t}$

2. 傾き β_t 自身がランダムウォークする「ローカルトレンドモデル」。

3. 二階差分 ($d = 2$) に相当する。

第 8 章

多変量時系列 (VAR, VECM)

ここまで扱ってきた時系列は「1 変数だけ」の UNIVARIATE モデルだったが、実務では複数変数が相互に影響し合う MULTIVARIATE 時系列が多く見られる。この章では、もっとも基本的な多変量モデルであるベクトル自己回帰 (VAR) と、非定常多変量データ向けの誤差修正モデル (VECM) の数学的構造を整理し、直感的な解釈と具体例を示す。

8.1 ベクトル自己回帰 (VAR) の基本

多変量時系列では、複数の変数を「ベクトル」で同時に扱う。

$$\mathbf{y}_t = (y_{1,t}, y_{2,t}, \dots, y_{k,t})^\top$$

8.1.1 VAR(p) モデルの定義

VAR(p) モデルは、

$$\mathbf{y}_t = \mathbf{c} + \mathbf{A}_1 \mathbf{y}_{t-1} + \dots + \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t$$

と書かれる。ここに：

- \mathbf{y}_t : k 次元時系列ベクトル
- \mathbf{A}_i : $k \times k$ 行列 (ラグ係数行列)
- $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$: 多変量ノイズ

直感：

- 各変数は「過去の全変数」の線形結合 + 共通ノイズで説明される。
- \mathbf{A}_1 の第 (i, j) 成分は、「変数 j の 1 期前にどの程度、変数 i に影響しているか」を表す。

8.2 VAR の推論と Granger 因果性

8.2.1 Granger 因果性

VAR(1) $\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$ において、「変数 j が変数 i に Granger 因果性を持つ」とは、 \mathbf{A}_1 の第 (i, j) 成分が 0 でないこと (有意であること) を意味する。これは予測精度の向上に寄与するかどうか検定する。

8.3 共統合と VECM

実務では多変量がそれぞれ非定常であっても、「足し合わせると定常な関係」が存在する可能性があり、これを共統合 (cointegration) という。

8.3.1 誤差修正モデル (VECM)

共統合関係があれば、VAR を「短期の変動」＋「長期の均衡探索」に分けることができる。

$$\Delta \mathbf{y}_t = \alpha \beta^\top \mathbf{y}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \mathbf{y}_{t-i} + \varepsilon_t$$

- $\beta^\top \mathbf{y}_t$: 長期均衡残差
- α : 調整係数行列 (均衡からの修正速度)

直感 : 「ばねに似た再均衡メカニズム」。

8.4 共統合の基本定理

共統合関係と誤差修正モデル (VECM) の等価性を保証する重要な定理をここで整理する。

定理 8.1 (Engle-Granger の共統合定理). 2 次元時系列ベクトル $\mathbf{y}_t = (y_{1,t}, y_{2,t})^\top$ の各成分が $I(1)$ (単位根過程) であり、その階差系列 $\Delta \mathbf{y}_t$ が弱定常であるとする。さらに、ある非ゼロベクトル $\boldsymbol{\theta} = (1, -\theta)^\top$ に対して、線形結合

$$z_t = \boldsymbol{\theta}^\top \mathbf{y}_t = y_{1,t} - \theta y_{2,t}$$

が弱定常過程 ($I(0)$) となるとき、 \mathbf{y}_t は共統合しているという。このとき、 \mathbf{y}_t に対して以下の誤差修正表現 (VECM) が存在する：

$$\Delta \mathbf{y}_t = \alpha \beta^\top \mathbf{y}_{t-1} + \mathbf{u}_t$$

ここで $\beta = (1, -\theta)^\top$ は共統合ベクトル、 α は調整係数ベクトル、 \mathbf{u}_t は弱定常な残差項である。

証明. 証明の概略を示す。

1. 定常性の確認 : 仮定より $z_t = \boldsymbol{\theta}^\top \mathbf{y}_t$ および $\Delta \mathbf{y}_t$ は弱定常である。
2. モデルの再構成 : $\Delta \mathbf{y}_t$ を過去の「均衡誤差」 z_{t-1} と「短期変動」のラグ項で回帰するモデルを考える。
Engle と Granger は、もし z_t が定常ならば、 $\Delta \mathbf{y}_t$ の生成過程において z_{t-1} が説明変数として有効 (係数 $\alpha \neq 0$) でなければならないことを示した。
3. VECM の導出 : 線形代数的な変形により、VAR モデル $\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \varepsilon_t$ を階差形式に書き換えると、右辺にレベル項 $\Pi \mathbf{y}_{t-1}$ が残る。 \mathbf{y}_t が $I(1)$ かつ共統合している場合、行列 Π はランク落ちしており、 $\Pi = \alpha \beta^\top$ と分解できる (Granger の表現定理)。この $\beta^\top \mathbf{y}_{t-1}$ が定常な均衡誤差 z_{t-1} に他ならない。

したがって、共統合関係が存在すれば、必ず誤差修正項を含むモデル (VECM) で表現できる。 \square

練習問題（第 8 章）

演習 8.1 (VAR 方程式の解釈). 次の VAR(1) を考える。

$$\mathbf{y}_t = \begin{bmatrix} 0.7 & 0.1 \\ 0.2 & 0.5 \end{bmatrix} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

1. \mathbf{A}_1 の第 (1,2) 成分 0.1 が y_1 に与える影響を説明せよ。
2. $y_{2,t-1}$ が +1 ノイズで跳ねた場合、期待される $y_{1,t}$ の変化量はいくつか。

演習 8.2 (Granger 因果性の直感). モデル $y_{1,t} = 0.6y_{1,t-1} + 0.4y_{2,t-1} + \varepsilon_{1,t}$, $y_{2,t} = 0.8y_{2,t-1} + \varepsilon_{2,t}$ を考える。

1. y_2 が y_1 に Granger 因果性を持つか。
2. y_1 が y_2 に Granger 因果性を持つか。

演習 8.3 (VECM と均衡の直感). 均衡式 $z_t = y_{1,t} - 0.8y_{2,t}$ に対する VECM を考える。調整項が $-0.1z_{t-1}$ である。

1. $\alpha = -0.1$ はどのような修正速度を意味するか。
2. $z_{t-1} > 0$ のとき、モデルは $\Delta y_{1,t}$ をどう動かそうとするか。

練習問題の解答（第 8 章）

解答 (問 1). 1. 昨日の y_2 が 1 単位増えると、今日の y_1 を平均的に 0.1 単位増やす。
2. +0.1 単位。

解答 (問 2). 1. 持つ（係数 0.4 がゼロでないなら）。
2. 持たない（ y_2 の式に y_1 が含まれない）。

解答 (問 3). 1. 弱い負の修正速度（ゆっくり戻る）。
2. 負の方向に動かそうとする（均衡に戻すため）。

第 9 章

周波数領域の時系列解析（スペクトル解析）

第 1 章～第 8 章までは「時間領域の表現」に焦点を当ててきたが、時系列解析のもう一つの巨大な流派が「周波数領域」＝スペクトル解析である。ここでは時系列を「正弦波の重ね合わせ」と見なして、どの周波数にどれくらいのパワーがあるか、周期的な構造を抽出する方法を数理的に整理する。

9.1 フーリエ級数と周期の分解

9.1.1 離散時間：確率過程への拡張

無限弱定常確率過程 $\{X_t\}$ の場合、自己共分散関数 $\gamma(h)$ とスペクトル密度関数 $f(\omega)$ の間に双対性がある。

9.2 スペクトル密度の基本定理

自己共分散とスペクトル密度の関係を数学的に厳密に保証するのが、以下の Wiener-Khinchin 定理である。

定理 9.1 (Wiener-Khinchin 定理). $\{X_t\}$ を弱定常過程とし、その自己共分散関数 $\gamma(h) = \text{Cov}(X_t, X_{t+h})$ が絶対可総和 ($\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$) であるとする。このとき、非負の関数 $f(\omega) \geq 0$ (スペクトル密度) が一意に存在し、以下のフーリエ変換対が成立する：

$$\gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} f(\omega) d\omega \quad (\text{逆変換}) \quad (9.1)$$

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h} \quad (\text{順変換}) \quad (9.2)$$

すなわち、自己共分散関数とスペクトル密度は互いにフーリエ変換の関係にある。

証明. 証明の概略を示す。

1. **正定値性**：自己共分散関数 $\gamma(h)$ は正定値数列である（任意の数列 a_t に対して分散 $\text{Var}(\sum a_t X_t) \geq 0$ となることから示される）。
2. **Bochner の定理**：Bochner の定理（または Herglotz の定理）より、任意の正定値数列 $\gamma(h)$ は、ある有界な非減少関数 $F(\omega)$ （スペクトル分布関数）を用いて $\gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} dF(\omega)$ と表せる。
3. **スペクトル密度の存在**：仮定 $\sum |\gamma(h)| < \infty$ により、スペクトル分布 $F(\omega)$ は絶対連続であり、その導関数としてスペクトル密度 $f(\omega) = F'(\omega)$ が存在する。このとき積分は $\gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} f(\omega) d\omega$ と書ける。
4. **反転公式**：フーリエ解析の一般論より、係数 $f(\omega)$ は $\gamma(h)$ のフーリエ級数として $f(\omega) = \frac{1}{2\pi} \sum \gamma(h) e^{-i\omega h}$ で与えられる。

以上より、時間領域の「相関構造」と周波数領域の「パワー分布」が等価であることが示された。 \square

9.3 スペクトル密度の解釈

9.4 スペクトル密度の解釈

9.4.1 周期性とピーク

スペクトル密度 $f(\omega)$ のグラフを描くと、周期 T に対応する周波数で山ができる。例： $\omega = 2\pi/7$ で大きなピーク → 「7 単位時間の周期」が強く存在。

9.4.2 白ノイズのスペクトル

白ノイズ $X_t = \varepsilon_t$ の場合、

$$f(\omega) = \frac{\sigma^2}{2\pi}$$

となり、全ての ω で一定（平坦）。直感的に「周期性がない」。

9.5 実務的なスペクトル密度の推定

自己共分散 $\gamma(h)$ を標本推定 $\hat{\gamma}(h)$ に置き換え、ウィンドウ関数 w_h をかけた推定量が使われる。

$$\hat{f}(\omega; M) = \frac{1}{2\pi} \sum_{h=-M}^M w_h \hat{\gamma}(h) e^{-i\omega h}$$

練習問題（第 9 章）

演習 9.1 (白ノイズとスペクトル). $X_t \sim \text{WN}(0, 4)$ のスペクトル密度 $f(\omega)$ を求めよ。

演習 9.2 (周期成分の解釈). $X_t = 3 \sin(2\pi t/5) + \varepsilon_t$ の周期 T と、スペクトルピークが期待される周波数 ω を求めよ。

練習問題の解答（第 9 章）

解答 (問 1). $f(\omega) = \frac{4}{2\pi} = \frac{2}{\pi}$ (一定)。

解答 (問 2). $T = 5$, $\omega = \frac{2\pi}{5}$ 。

第 10 章

異常検知・構造変化・レジームスイッチ

ここまで扱ってきたモデルは「時間を通して『構造は一定』である」ことを前提としてきたが、実際にはニューモデルに移行する変化点や不正利用などの異常シフトが気付かれる。この章では、時系列における異常検知、変化点検出（Change Point Detection）、レジームスイッチの数学的枠組みを整理する。

10.1 異常検知：概要と定義

10.1.1 異常の 2 タイプ

- 漸変型異常（Drift）：平均値や共分散が「ゆっくり」変化
- 点異常（Outliers）：一瞬だけ外れ値的に値が跳ねる

10.1.2 統計的モデルによる異常検知

既定モデルからの逸脱確率 $P(X_t = x_t \mid \hat{\theta})$ が小さすぎる場合、異常と判定する。

10.2 予測誤差ベースの異常検知

10.2.1 予測誤差スコア

$$\text{Anomaly Score} = \frac{|y_t - \hat{y}_t|}{\sqrt{\text{Var}(\varepsilon_t)}}$$

このスコアが閾値（例: 3）を超えたら異常とする。

10.3 構造変化とレジームスイッチ

10.3.1 変化点検出

分布パラメータ θ が時刻 t で不連続に変化することを見つける。

10.3.2 レジームスイッチングモデル

複数の状態（レジーム）間をマルコフ遷移確率で移動するモデル。

$$\mathbf{y}_t = \mathbf{c}_m + \mathbf{A}_m \mathbf{y}_{t-1} + \varepsilon_{m,t}, \quad m \in \{1, \dots, M\}$$

練習問題（第 10 章）

演習 10.1 (異常検知). 正常モデル $\mathcal{N}(0, 4)$ に対して観測値 $y_t = 4$ が得られた。異常スコアを計算し、閾値 2 で判定せよ。

練習問題の解答（第 10 章）

解答 (問 1). 予測値 0、標準偏差 2。スコア $= |4 - 0|/2 = 2$ 。閾値 2 以上なので異常と判定される（境界線）。

第 11 章

機械学習と時系列：RNN・LSTM・Transformer など

これまで扱ってきた ARIMA・状態空間モデル・VAR などは、統計的仮定を前提としたモデルである。一方、ディープラーニングや機械学習手法は、明示的な仮定を少なくし、データから直接構造を学習する。この章では、時系列に有効な機械学習の枠組み、特に RNN・LSTM・Transformer ベースのモデル構造と、統計的視点からの解釈を整理する。

11.1 時系列に特化した機械学習の考え方

11.1.1 ラグ特徴量による「古典 ML」

単純なやり方は、ラグを用いた回帰学習である。

$$\hat{x}_t = f(x_{t-1}, x_{t-2}, \dots, x_{t-p})$$

f として線形回帰、ランダムフォレスト、GBDT (LightGBM/XGBoost) などを使用する。

- ARMA より非線形構造を扱える。
- 解釈性はモデルに依存する (Tree 系は Feature Importance で寄与度を確認可)。

11.2 RNN・LSTM：時刻ごとに更新される内部状態

11.2.1 RNN の基本構造

$$\mathbf{h}_t = \tanh(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{b}_h)$$

過去のコンテキストを \mathbf{h}_t に圧縮するが、勾配消失問題により長期記憶は苦手。

11.2.2 LSTM：長期記憶のために

$$\begin{aligned} \mathbf{i}_t &= \sigma(\dots) && \text{(Input Gate)} \\ \mathbf{f}_t &= \sigma(\dots) && \text{(Forget Gate)} \\ \mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\dots) && \text{(Cell State)} \end{aligned}$$

細胞状態 \mathbf{c}_t を通じて長期依存を保持する。

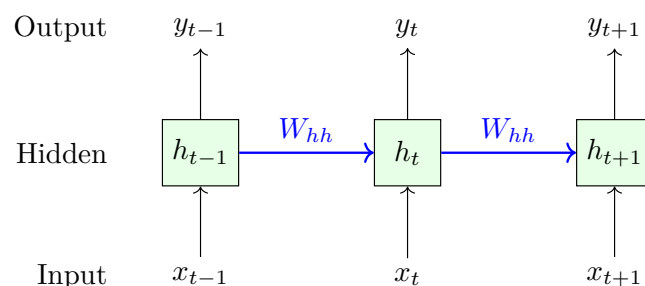


図 11.1 時間展開された RNN の構造

11.3 Transformer と時系列予測

11.3.1 Self-Attention の数学的イメージ

過去の系列全体にアクセスし、重要度（Attention Weight）を動的に計算して予測に用いる。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d}}\right)V$$

統計的には「動的な重みを持つ AR モデル」のような解釈も可能。

11.4 統計モデルとの接続と補完

- **ARIMA:** トレンド・季節の解釈性が高く、小データに強い。
- **ML/DL:** 非線形性・大量データ・複雑なパターンに強い。

実務では、ARIMA でトレンド除去した残差を ML で予測するハイブリッド手法も有効。

練習問題（第 11 章）

演習 11.1 (ML ラグ特徴量の理解). $\hat{y}_t = \beta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2}$ はどの統計モデルに相当するか。また $\alpha_1 = 0.6, \alpha_2 = 0.2$ の直感的な解釈を述べよ。

演習 11.2 (RNN vs LSTM). 長い依存構造を持つデータで RNN が不利な理由を勾配消失の観点から述べよ。

演習 11.3 (Attention と AR). Attention Weight α_i が動的に変わることは、AR モデルと比較してどのような利点があるか。

練習問題の解答（第 11 章）

解答 (問 1). AR(2) モデルに相当。1 期前の影響が最も強く、2 期前も寄与する（指数減衰的）。

解答 (問 2). RNN は再帰計算により勾配が連鎖的に小さくなり（または爆発し）、過去の情報が消失するため。

解答 (問 3). 時点ごとに「どの過去時点（ラグ）」を重視するかを適応的に変えられるため、突発的なイベントや変化に柔軟に対応できる。

第 12 章

実務ワークフローとケーススタディ

実務で時系列解析を用いるときは、単独自動回帰（ARIMA など）や現代機械学習（RNN / LSTM）のどちらを選ぶかよりも、問題の設定・データの構造・モデルの複雑さに合わせて「分析・設計・運用」の一貫したワークフローを工夫することが重要である。本章では、実務シーンでよく現れるケーススタディを通じて、数理的に整理しつつ実務運用フローを示す。

12.1 実務ワークフローの一般的構成

1. 目的・問いの明確化：何を予測したいか、許容誤差はどの程度か。
2. データ理解と前処理：欠損処理、外れ値除去、トレンド・季節分解。
3. モデル候補の設計：ARIMA / 状態空間 / ML モデルの選定。
4. 訓練・評価指標の設定：RMSE / MAE / MAPE などの選定と CV。
5. 実務運用から監視まで：推論パイプラインの構築とドリフト監視。

12.2 ケーススタディ 1：店舗売上予測（ARIMA と STL）

12.2.1 問題の設定

日次売上データ y_t から翌週の需要を予測し、在庫最適化を行う。

12.2.2 アプローチ

- 分解：STL 分解により週周期とトレンドを分離。
- 定常化：残差成分に対して定常性を確認。
- 予測：トレンドは外挿、季節成分は繰り返し、残差は ARIMA で予測して合算。

12.3 ケーススタディ 2：センサー異常検知（状態空間モデル）

12.3.1 問題の設定

工場のセンサーデータから、機器の故障予兆（異常な振る舞い）を検知する。

12.3.2 アプローチ

- モデル化：正常時の動作をローカルレベルモデル＋季節成分（稼働サイクル）で学習。

- スコアリング：カルマンフィルタによる予測残差（イノベーション）の標準化誤差を異常スコアとする。
- 判定：スコアが閾値（例: 3σ ）を超えた場合にアラートを発報。

まとめ

時系列解析は、過去のデータに潜む構造（トレンド・周期・自己相関）を数理的に抽出し、未来の不確実性を定量化する強力な道具である。基礎的な確率統計から古典的 ARIMA、柔軟な状態空間モデル、そして最新の機械学習手法まで、それぞれの特性を理解し、適切な場面で組み合わせることが実務での成功の鍵となる。

付録 A

集中不等式の道具箱

(Orlicz ノルム・同値な定義・matrix Bernstein)

A.1 Orlicz ノルムと確率変数クラス

定義 A.1 (ψ_2 Orlicz ノルム : sub-Gaussian). 確率変数 X に対し

$$\|X\|_{\psi_2} := \inf \left\{ K > 0 : \mathbb{E} \exp(X^2/K^2) \leq 2 \right\}$$

と定める。 $\|X\|_{\psi_2} < \infty$ のとき X を sub-Gaussian と呼ぶ。

定義 A.2 (ψ_1 Orlicz ノルム : sub-exponential).

$$\|X\|_{\psi_1} := \inf \left\{ K > 0 : \mathbb{E} \exp(|X|/K) \leq 2 \right\}$$

と定める。 $\|X\|_{\psi_1} < \infty$ のとき X を sub-exponential と呼ぶ。

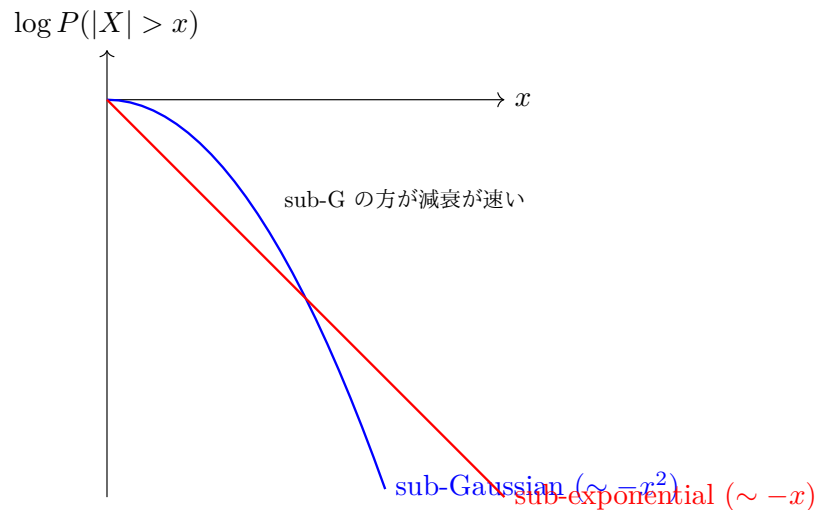


図 A.1 Orlicz ノルムと尾確率の減衰 : ψ_2 (sub-Gaussian) は正規分布のように二乗で減衰し、 ψ_1 (sub-exponential) は指数分布のように線形で減衰する。Matrix Bernstein ではこの二つの混合形が現れる。

補題 A.3 (基本性質 : スケーリング・三角不等式型). 任意の $a \in \mathbb{R}$ について $\|aX\|_{\psi_\alpha} = |a|\|X\|_{\psi_\alpha}$ ($\alpha = 1, 2$). またある普遍定数 $C > 0$ が存在して

$$\|X + Y\|_{\psi_2} \leq C(\|X\|_{\psi_2} + \|Y\|_{\psi_2}), \quad \|X + Y\|_{\psi_1} \leq C(\|X\|_{\psi_1} + \|Y\|_{\psi_1})$$

が成り立つ (厳密な「ノルム」ではなく準ノルムであることに注意)。

証明. スケーリングは定義に $X \mapsto aX$ を代入して直ちに従う。加法については $\exp((x+y)^2) \leq \exp(2x^2)\exp(2y^2)$ と Hölder を組み合わせ、定義の「 ≤ 2 」を満たす係数を調整すればよい (ψ_1 も同様に $\exp(|x+y|) \leq \exp(|x|)\exp(|y|)$ を使う)。

A.2 sub-Gaussian の同値性 (mgf・尾・モーメント)

定理 A.4 (sub-Gaussian の同値な特徴付け). 平均 0 の確率変数 X について、以下は互いに同値であり、定数は普遍定数倍で行き来できる。

1. (Orlicz) $\|X\|_{\psi_2} \leq K$.
2. (尾確率) ある $c > 0$ が存在して任意の $t \geq 0$ で

$$\mathbb{P}(|X| \geq t) \leq 2 \exp\left(-c \frac{t^2}{K^2}\right).$$

3. (モーメント増大) ある $C > 0$ が存在して任意の $q \geq 1$ で

$$(\mathbb{E}|X|^q)^{1/q} \leq CK\sqrt{q}.$$

4. (mgf) ある $C > 0$ が存在して任意の $s \in \mathbb{R}$ で

$$\mathbb{E} \exp(sX) \leq \exp(CK^2 s^2).$$

証明. (1) \Rightarrow (2): 任意の $t \geq 0$ に対し Markov より

$$\mathbb{P}(|X| \geq t) = \mathbb{P}\left(\exp(X^2/K^2) \geq e^{t^2/K^2}\right) \leq e^{-t^2/K^2} \mathbb{E} e^{X^2/K^2} \leq 2e^{-t^2/K^2}.$$

よって (2) が $c = 1$ で成立。

(2) \Rightarrow (3): 分布関数の積分表示 (tail integration) を使う:

$$\mathbb{E}|X|^q = \int_0^\infty qt^{q-1}\mathbb{P}(|X| \geq t) dt \leq 2q \int_0^\infty t^{q-1} \exp\left(-c \frac{t^2}{K^2}\right) dt.$$

変数変換 $u = ct^2/K^2$ により右辺は $\lesssim K^q q(q/2 - 1)!$ 型になり、Stirling を使って $(\mathbb{E}|X|^q)^{1/q} \lesssim K\sqrt{q}$ を得る。

(3) \Rightarrow (4): 級数で mgf を評価する:

$$\mathbb{E} e^{sX} = 1 + \sum_{m \geq 2} \frac{s^m \mathbb{E} X^m}{m!} \leq 1 + \sum_{m \geq 2} \frac{|s|^m \mathbb{E}|X|^m}{m!}.$$

(3) より $\mathbb{E}|X|^m \leq (CK\sqrt{m})^m$ 。これを代入すると

$$\frac{|s|^m \mathbb{E}|X|^m}{m!} \leq \frac{(|s|CK)^m m^{m/2}}{m!}$$

であり、Stirling $m! \asymp (m/e)^m \sqrt{m}$ から級数は $\exp(C'K^2 s^2)$ で抑えられる。

(4) \Rightarrow (1): (4) から X のガウス型尾確率 (Chernoff) を得て (2) を経由してもよいし、直接 $\mathbb{E} e^{X^2/K^2}$ をモーメントで展開して収束半径を示してもよい。いずれも定数調整で $\|X\|_{\psi_2} \lesssim K$ が従う。

A.3 sub-exponential と Bernstein 型評価

定理 A.5 (sub-exponential の同値性: ψ_1). 確率変数 X について $\|X\|_{\psi_1} \leq K$ であることは、ある $c > 0$ が存在して

$$\mathbb{P}(|X| \geq t) \leq 2 \exp\left(-c \frac{t}{K}\right), \quad t \geq 0$$

が成り立つことと同値である（定数は普遍定数倍で行き来）。

証明. ψ_2 と同様に Markov で (Orlicz) \Rightarrow (尾) を示し、逆は tail integration により $\mathbb{E} \exp(|X|/CK) \leq 2$ を構成すればよい。

定理 A.6 (独立 sub-exponential 和の Bernstein). 独立で平均 0 の X_1, \dots, X_n が $\|X_i\|_{\psi_1} \leq K$ を満たすとき、ある普遍定数 $c, C > 0$ が存在して任意の $t \geq 0$ で

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_i \right| \geq t \right) \leq 2 \exp \left[-cn \min \left(\frac{t^2}{K^2}, \frac{t}{K} \right) \right]$$

が成り立つ。

証明. mgf の局所上界 ($|s| \leq c/K$ で $\mathbb{E} e^{sX_i} \leq \exp(Cs^2K^2)$) を sub-exponential の同値性から導く。独立性で積に分解し、Chernoff により $\exp(-st + Cs^2K^2)$ の最小化を「二次領域 (t 小) / 一次領域 (t 大)」に分けて行くと結論が出る。

A.4 最大値評価と union bound の定石

補題 A.7 (sub-Gaussian 最大値の上界). X_1, \dots, X_p が (独立でなくてもよいが) 同一の尾上界 $\mathbb{P}(|X_j| \geq t) \leq 2 \exp(-t^2/(2K^2))$ を満たすとき

$$\mathbb{P} \left(\max_{1 \leq j \leq p} |X_j| \geq K \sqrt{2 \log(2p/\delta)} \right) \leq \delta.$$

証明. union bound より

$$\mathbb{P} \left(\max_j |X_j| \geq t \right) \leq \sum_{j=1}^p \mathbb{P}(|X_j| \geq t) \leq 2p \exp(-t^2/(2K^2)).$$

右辺を $\leq \delta$ とおくように t を解けばよい。

(Lasso の $\|(1/n)X^\top \varepsilon\|_\infty$ 抑え込みが、まさにこの型で出てきます。)

A.5 Matrix Bernstein (自己共役行列の集中)

ここからは「非可換性 (行列の積が可換でない)」が障害になるが、行列版のラプラス変換法で同型の結果が得られる。

補題 A.8 (行列版 Markov : ラプラス変換法). 自己共役行列 Y と $t > 0$ に対し

$$\mathbb{P}(\lambda_{\max}(Y) \geq u) = \mathbb{P}(e^{t\lambda_{\max}(Y)} \geq e^{tu}) \leq e^{-tu} \operatorname{tr} \mathbb{E}[e^{tY}]$$

が成り立つ。

証明. $\lambda_{\max}(Y) \geq u \Rightarrow \operatorname{tr}(e^{tY}) \geq e^{tu}$ (最大固有値の寄与で trace が下から抑えられる) を用い、Markov を $\operatorname{tr}(e^{tY})$ に適用する。

定理 A.9 (Matrix Bernstein : 簡潔版). 独立な自己共役ランダム行列 $X_1, \dots, X_n \in \mathbb{R}^{d \times d}$ が

$$\mathbb{E}[X_k] = 0, \quad \|X_k\|_{\text{op}} \leq R \text{ a.s.}$$

を満たすとする。分散パラメータを

$$\sigma^2 := \left\| \sum_{k=1}^n \mathbb{E}[X_k^2] \right\|_{\text{op}}$$

とおくと、任意の $u \geq 0$ で

$$\mathbb{P}\left(\left\|\sum_{k=1}^n X_k\right\|_{\text{op}} \geq u\right) \leq 2d \cdot \exp\left(-\frac{u^2}{2\sigma^2 + \frac{2Ru}{3}}\right). \quad (\text{A.1})$$

証明 (主要ステップを明示). (1) まず片側を抑える。 $Y = \sum X_k$ とし、補題 A.6 より

$$\mathbb{P}(\lambda_{\max}(Y) \geq u) \leq e^{-tu} \text{tr} \mathbb{E} e^{tY}. \quad (\text{A.2})$$

(2) 次に mgf を逐次条件付き期待値で処理する。独立性により

$$\mathbb{E} e^{t \sum_{k=1}^n X_k} = \mathbb{E} \left[\mathbb{E} \left[e^{t(\sum_{k=1}^{n-1} X_k)} e^{tX_n} \mid X_1, \dots, X_{n-1} \right] \right] = \mathbb{E} \left[e^{t(\sum_{k=1}^{n-1} X_k)} \mathbb{E} e^{tX_n} \right].$$

非可換性のため単純に積の形にはならないが、「trace と指数関数」に関する標準不等式 (Lieb 型の凸性) により、最終的に

$$\text{tr} \mathbb{E} e^{t \sum_{k=1}^n X_k} \leq \text{tr} \exp \left(\sum_{k=1}^n \log \mathbb{E} e^{tX_k} \right) \quad (\text{A.3})$$

が得られる (ここが行列版の核心補題)。

(3) 各 X_k の mgf をスカラーの Bernstein と同じく二次で抑える。条件 $\|X_k\|_{\text{op}} \leq R$ と $\mathbb{E} X_k = 0$ から、 $|t| < 3/R$ の範囲で

$$\mathbb{E} e^{tX_k} \preceq \exp \left(\frac{t^2}{2(1 - \frac{Rt}{3})} \mathbb{E}[X_k^2] \right) \quad (\text{A.4})$$

が示せる (テイラー展開と $X_k^m \preceq R^{m-2} X_k^2$ 型の順序評価で作る)。

(4) (A.3)(A.4) をまとめると

$$\text{tr} \mathbb{E} e^{tY} \leq \text{tr} \exp \left(\frac{t^2}{2(1 - \frac{Rt}{3})} \sum_{k=1}^n \mathbb{E}[X_k^2] \right) \leq d \cdot \exp \left(\frac{t^2}{2(1 - \frac{Rt}{3})} \sigma^2 \right). \quad (\text{A.5})$$

最後の不等式は $\text{tr}(e^A) \leq d e^{\lambda_{\max}(A)}$ を用いた。

(5) (A.2)(A.5) より

$$\mathbb{P}(\lambda_{\max}(Y) \geq u) \leq d \cdot \exp \left(-tu + \frac{t^2}{2(1 - \frac{Rt}{3})} \sigma^2 \right), \quad 0 < t < 3/R.$$

右辺を t で最小化すると、(A.1) の片側版が出る。もう一方は $\lambda_{\min}(Y) \leq -u$ に同様に適用し、和をとって係数 $2d$ を得る。

A.6 例題

例題 A.10 (ガウスベクトルの線形汎関数). $g \sim N(0, I_d)$, 固定 $u \in \mathbb{R}^d$ に対し $X = u^\top g$ は $N(0, \|u\|_2^2)$ 。したがって $\|X\|_{\psi_2} \asymp \|u\|_2$ を確認せよ (定理 A.2 のいずれかの特徴付けを使う)。

例題 A.11 (中心化 χ^2 は sub-exponential). $Z \sim N(0, 1)$ とし $X = Z^2 - 1$ 。mgf $\mathbb{E} e^{t(Z^2-1)}$ を $|t| < 1/2$ で評価し、 $\|X\|_{\psi_1} < \infty$ を示せ。

A.7 演習問題

演習 A.1. 1. 定理 A.2 の (2) \Rightarrow (3) を、変数変換まで含めて定数を追って示せ。

2. sub-Gaussian の同値性から「 X sub-Gaussian $\Rightarrow X^2 - \mathbb{E}X^2$ sub-exponential」を証明せよ。
3. 補題 A.5 を用いて、独立 sub-Gaussian X_1, \dots, X_p に対し $\mathbb{E} \max_j |X_j| \lesssim K \sqrt{\log p}$ を示せ（積分表示 $\mathbb{E}U = \int_0^\infty \mathbb{P}(U \geq t) dt$ を使う）。
4. 定理 A.7 のステップ (3)（行列 mgf の二次上界）を、テイラー展開と $\|X_k\|_{\text{op}} \leq R$ から自力で埋めよ（ヒント：偶数次と奇数次を分け、作用素順序で抑える）。
5. 応用：第 8 章の標本共分散の評価を、(A.1) を使って「 ϵ -net を使わない」方針で組み直せ（どこで d が現れるかを比較せよ）。