

時系列解析

基本的なモデル

村田 昇

2020.12.15

今週の内容

- 第 1 日: 時系列のモデル
- 第 2 日: モデルの推定と予測

時系列解析の概要

時系列解析とは

- 時系列データ
 - 時間軸に沿って観測されたデータ
 - 観測の順序に意味がある
 - 異なる時点間での観測データの従属関係が重要
 - 独立性にもとづく解析は行えない
(そのままでは大数の法則や中心極限定理は使えない)
- 時系列解析の目的
 - 時系列データの特徴を効果的に記述すること
 - 時系列モデルの推定と評価

時系列データ

- 統計学・確率論における表現: **確率過程**
- 時間を添え字として持つ確率変数列:

$$X_t, t = 1, 2, \dots, T \quad (\text{あるいは } t = 0, 1, \dots, T)$$

- 時系列解析で利用される代表的な確率過程
 - ホワイトノイズ
 - ランダムウォーク
 - 自己回帰モデル (AR モデル)
 - 移動平均モデル (MA モデル)
 - 自己回帰移動平均モデル (ARMA モデル)

基本的なモデル

ホワイトノイズ

- 定義

P を平均 0, 分散 σ^2 の確率変数の確率分布とする

$$X_t = \epsilon_t, \quad \epsilon_t \stackrel{i.i.d.}{\sim} P$$

- 記号 $\text{WN}(0, \sigma^2)$ で表記することが多い

$$X_t \sim \text{WN}(0, \sigma^2)$$

- 独立であるため系列としての予測は不可能

トレンドのあるホワイトノイズ

- 定義

μ, α を定数として

$$X_t = \mu + \alpha t + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

で与えられる確率過程

- $\mu + \alpha t$: **トレンド**
- **平均** が時間とともに変動する時系列モデルの 1 つ
- **トレンド項** はより一般化されることもある
 - t の 1 次式 (上記の基本的な場合)
 - 高次の多項式
 - 非線形関数 (指数関数, 三角関数など)

ランダムウォーク

- 定義

X_0 を定数もしくは確率変数として

$$X_t = X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

で帰納的に定義される確率過程

- **分散** が時間とともに増加する時系列モデルの 1 つ
- 最も単純な **記憶** のあるモデル

人工データによる例

- 同じモデルに従うパス (系列) を複数観測してみる

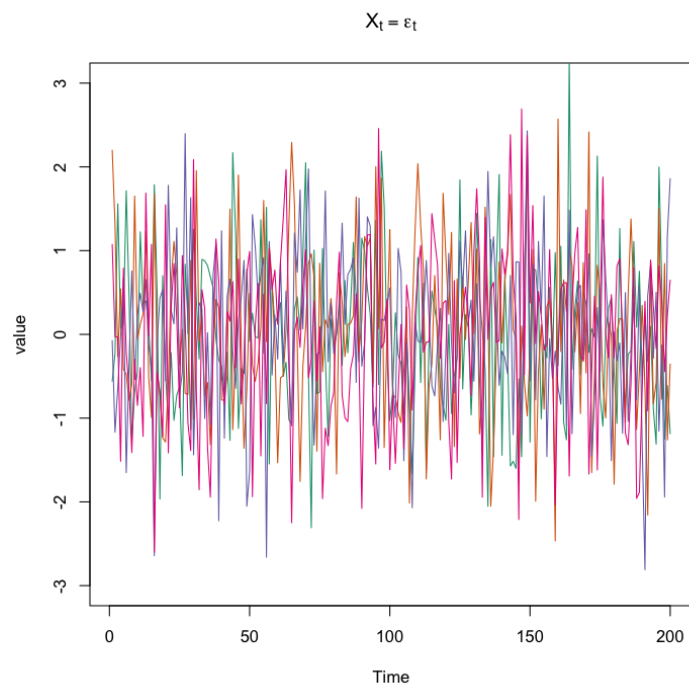


図 1: ホワイトノイズ

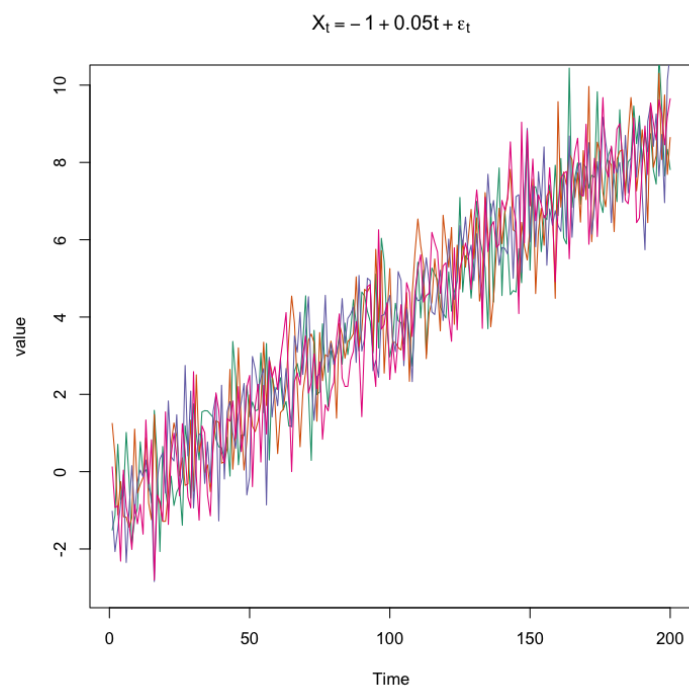


図 2: トレンドのあるホワイトノイズ

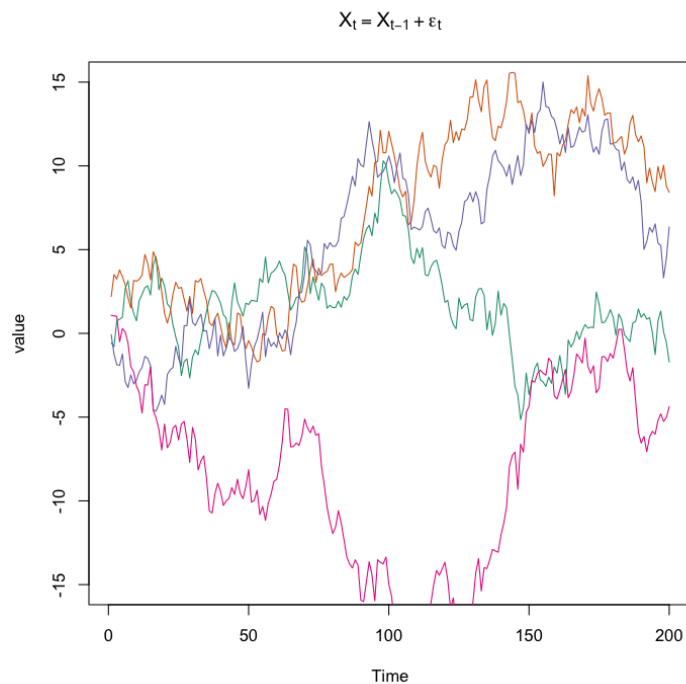


図 3: ランダムウォーク

演習

問題

- 以下の問に答えなさい。
 - トレンドのあるホワイトノイズ X_t の
 - * 平均 $E[X_t]$
 - * 分散 $\text{Var}(X_t)$
 を求めなさい
 - ランダムウォークの平均と分散を求めなさい

より一般的なモデル

自己回帰過程

- 定義 (次数 p ; AR(p), auto regressive の略)
 a_1, \dots, a_p を定数とし, X_1, \dots, X_p が初期値として与えられたとき,

$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

で帰納的に定義される確率過程

- ランダムウォークの一般化
 - * $p = 1, a_1 = 1$ かつ ϵ_t が独立同分布ならランダムウォーク
- 忘却しながら記憶するモデル ($|a_i| < 1$ などの条件が必要)

移動平均過程

- 定義 (次数 q ; MA(q), moving average の略)

b_1, \dots, b_q を定数とし, X_1, \dots, X_q が初期値として与えられたとき

$$X_t = b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$

で定義される確率過程

- 記憶のあるホワイトノイズ (構成する部品を記憶)

自己回帰平均移動過程

- 定義 (次数 (p, q) ; ARMA(p, q))

$a_1, \dots, a_p, b_1, \dots, b_q$ を定数とし, $X_1, \dots, X_{\max\{p, q\}}$ が初期値として与えられたとき

$$\begin{aligned} X_t &= a_1 X_{t-1} + \dots + a_p X_{t-p} \\ &\quad + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} + \epsilon_t, \\ \epsilon_t &\sim \text{WN}(0, \sigma^2) \end{aligned}$$

で帰納的に定まる確率過程

- AR(p) モデルは ARMA($p, 0$), MA(q) モデルは ARMA($0, q$)
- 単純な形ながら異なる時点間の従属構造を柔軟に記述
- 基本的な時系列モデルとして広く利用されている

人工データによる例

- 同じモデルに従うパス (系列) を複数観測してみる

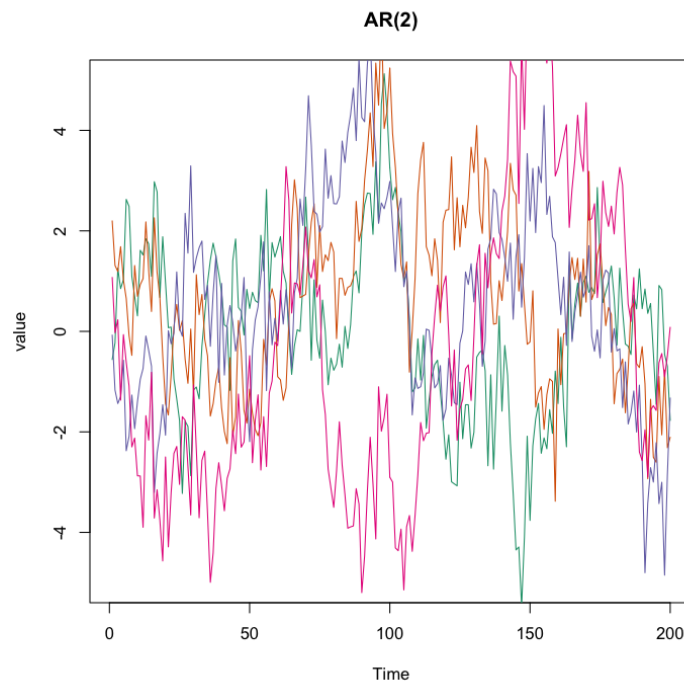


図 4: AR 過程

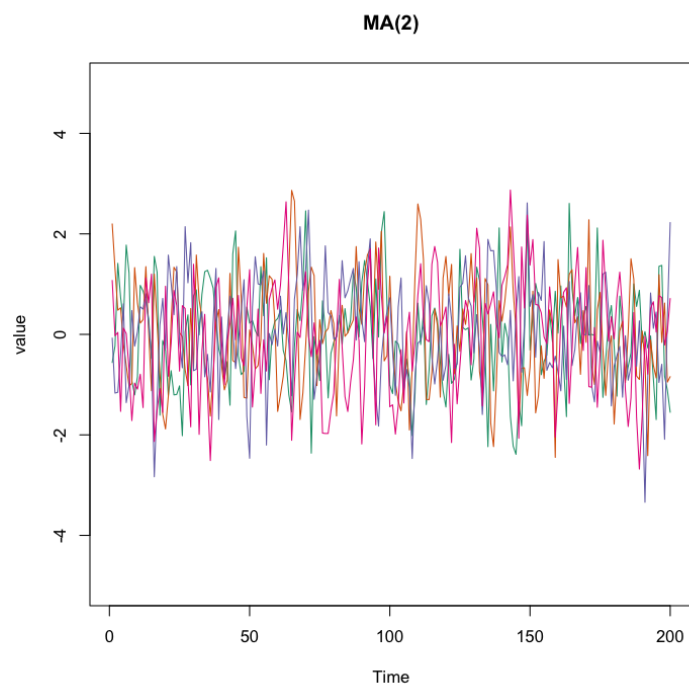


图 5: MA 過程

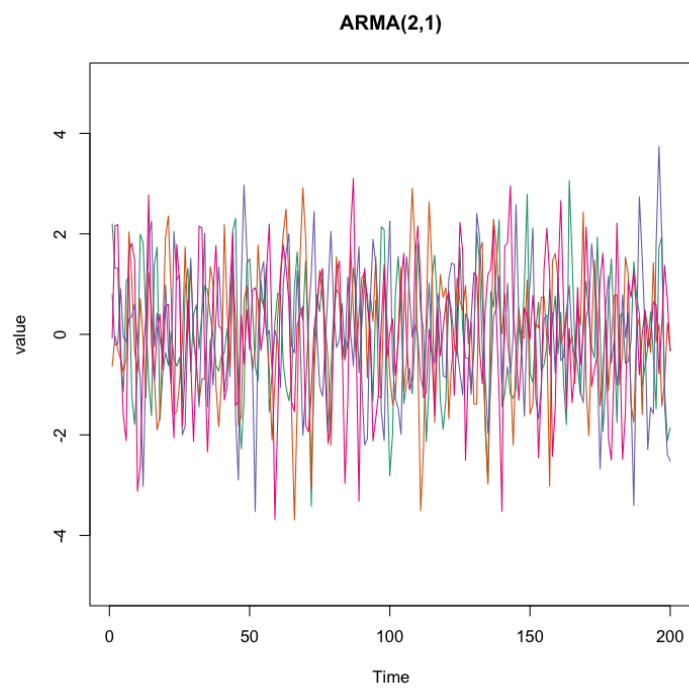


图 6: ARMA 過程

演習

問題

- 以下の問に答えなさい.
 - AR(1) の平均と分散を求めなさい
 - MA(1) の平均と分散を求めなさい

定常過程と非定常過程

弱定常性

- 確率過程 $X_t, t = 1, \dots, T$ が次の性質をもつ:
 - X_t の平均は時点 t によらない

$$\mathbb{E}[X_t] = \mu \quad (\text{時間の添字を持たない})$$

- X_t と X_{t+h} の共分散は時点 t によらず時差 h のみで定まる

$$\text{Cov}(X_t, X_{t+h}) = \gamma(h) \quad (\text{時間の添字を持たない})$$

- 特に X_t の分散は時点 t によらない ($h = 0$ の場合)

$$\text{Var}(X_t) = \gamma(0)$$

定常性と非定常性

- 定常でない確率過程は **非定常** であるという
- いろいろな確率過程の定常性
 - 定常: ホワイトノイズ, MA
 - 非定常: トレンドのあるホワイトノイズ, ランダムウォーク
 - 定常にも非定常にもなりうる: AR, ARMA

非定常過程の難しさ

- 特徴付ける特徴量が不在
 - 平均や分散などの基本的な統計量が時間によって変動する
 - 1つの時系列からこれらの統計量の推測はできない
- 擬相関
 - 独立な時系列にも関わらず見掛けの相関が現れることがある
 - <https://tylervigen.com/spurious-correlations>

非定常過程の取り扱い

- 定常過程とみなせるように変換したあと分析を実行

- 階差をとる変換

ランダムウォークは階差をとればホワイトノイズ (定常過程) となる

$$X_t = X_{t-1} + \epsilon_t \Rightarrow Y_t = X_t - X_{t-1} = \epsilon_t$$

- 対数変換

対数変換と階差で微小な比率の変動を取り出すことができる

$$X_t = (1 + \epsilon_t)X_{t-1} \Rightarrow Y_t = \log(X_t) - \log(X_{t-1}) = \log(1 + \epsilon_t) \simeq \epsilon_t$$

- トレンド成分+季節成分+変動成分への分解

適当な仮説のもとに取り扱いやすい成分の和に分解する

自己共分散・自己相関

自己共分散・自己相関

- 確率過程 X_t が定常過程の場合

- X_t と X_{t+h} の共分散は時点 t によらずラグ h のみで定まる

自己共分散 (定常過程の性質よりラグは $h \geq 0$ を考えればよい)

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

- X_t と X_{t+h} の相関も t によらずラグ h のみで定まる

自己相関

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

- 異なる時点間での観測データの従属関係を要約するための最も基本的な統計量

標本自己共分散・標本自己相関

- 観測データ X_1, \dots, X_T からの推定

- ラグ h の自己共分散の推定: 標本自己共分散

$$\frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X})(X_{t+h} - \bar{X})$$

$\bar{X} = \frac{1}{T} \sum_{t=1}^T X_t$ は標本平均

- ラグ h での自己相関の推定: 標本自己相関

$$\frac{\sum_{t=1}^{T-h} (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

人工データによる例

- 同じモデルに従うパス (系列) の自己相関を比較してみる

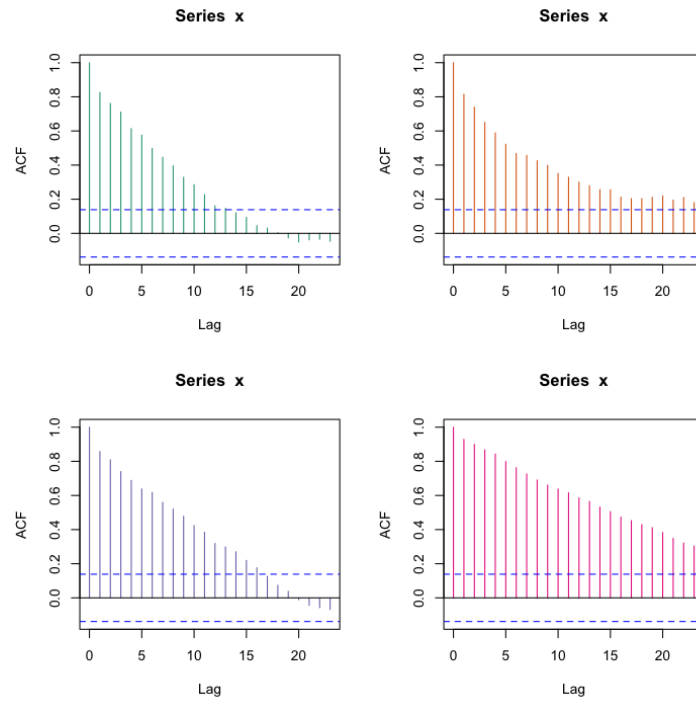


図 7: AR 過程の自己相関

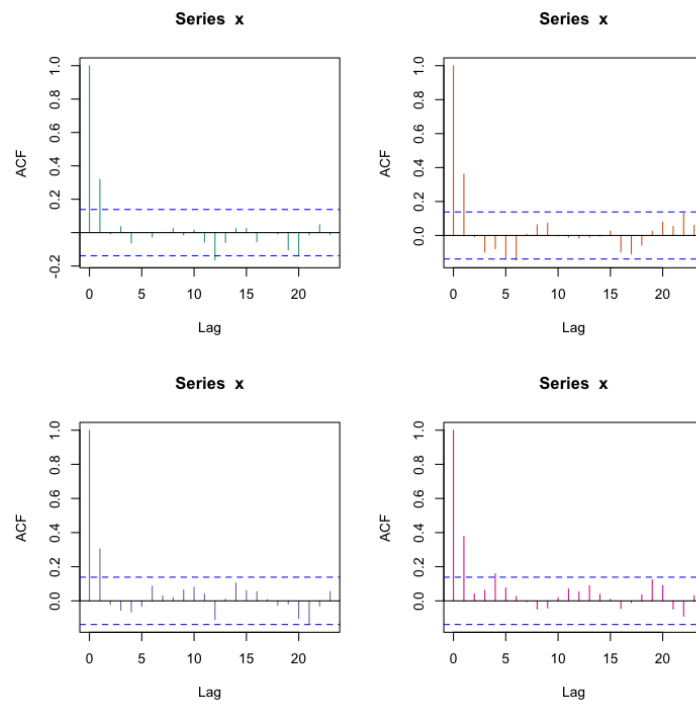


図 8: MA 過程の自己相関

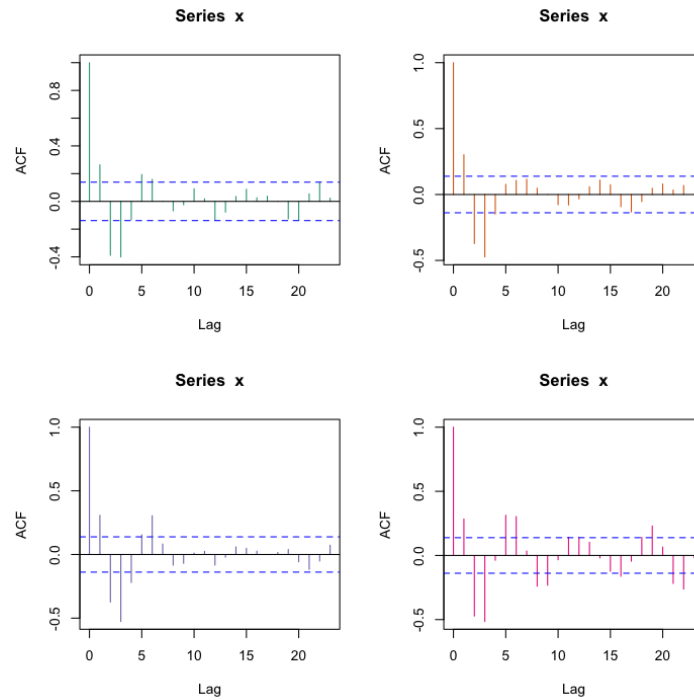


図 9: ARMA 過程の自己相関

演習

問題

- 以下の問に答えなさい.
 - $AR(p)$ の係数と自己共分散の間に成り立つ関係を考えなさい

次週の内容

- 第 1 日: 時系列のモデル
- 第 2 日: モデルの推定と予測