時系列解析

時系列の基本モデル

村田 昇

講義の内容

- **第 1 日: 時系列の基本モデル**
- 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

  平均 0, 分散 σ^2 である確率変数の確率分布 P からの独立かつ同分布な確率変数列

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

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

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

- 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 を定数として以下で定義される確率過程

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

- **トレンド** $\mu + \alpha t$ 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)
- **平均** が時間とともに変動する時系列モデルの 1 つ

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以下で帰納的に定義される確率過程

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

- **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

数値例

- 同じモデルに従うパス (系列) を複数観測してみる
  - ホワイトノイズ
  - トレンドのあるホワイトノイズ
  - 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 $\mathbb{E}[X_t]$
    * 分散 $\mathrm{Var}(X_t)$

    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

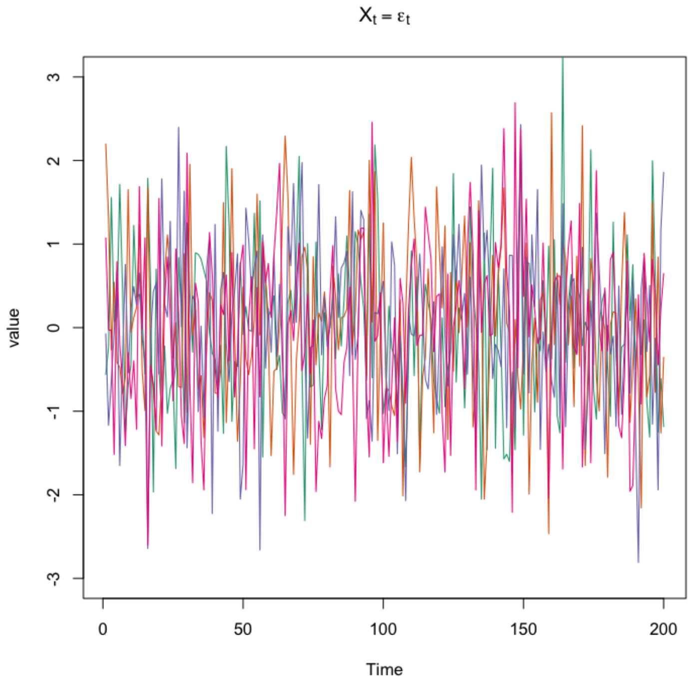

図 1: ホワイトノイズ

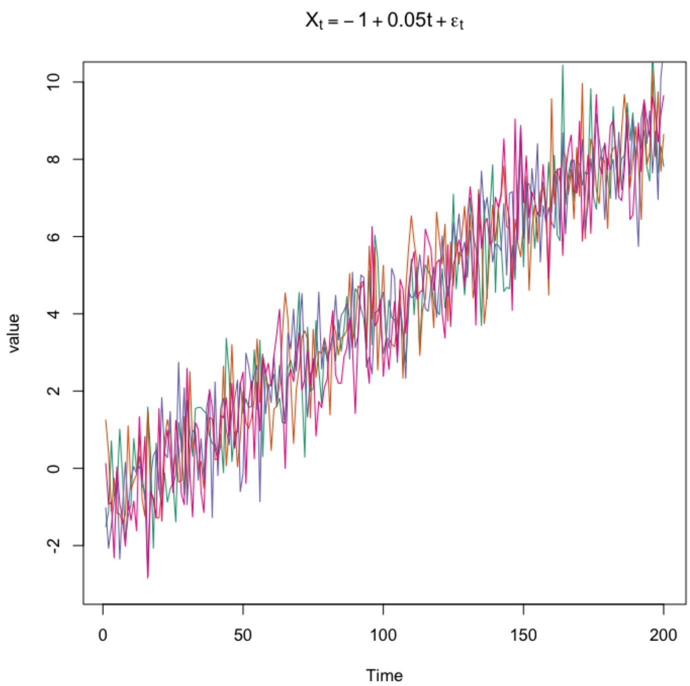

図 2: トレンドのあるホワイトノイズ

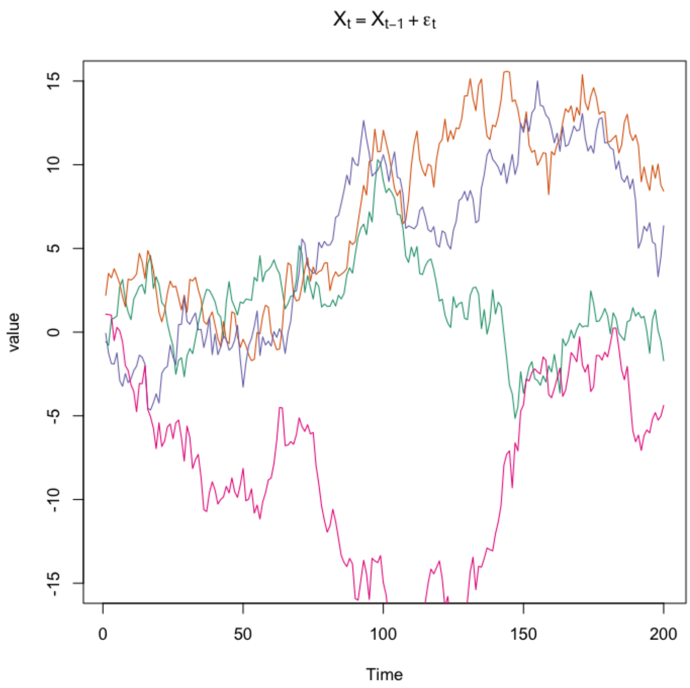

図 3: ランダムウォーク

解答例

- 定義に従い計算する

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[\mu + \alpha t + \epsilon_t] \\
&= \mathbb{E}[\mu] + \mathbb{E}[\alpha t] + \mathbb{E}[\epsilon_t] \\
&= \mu + \alpha t \\
\mathrm{Var}(X_t) &= \mathrm{Var}(\mu + \alpha t + \epsilon_t) \\
&= \mathrm{Var}(\mu) + \mathrm{Var}(\alpha t) + \mathrm{Var}(\epsilon_t) \\
&= 0 + 0 + \sigma^2 \\
&= \sigma^2
\end{aligned}
$$

- 定義に従い帰納的に計算する

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[X_{t-1} + \epsilon_t] \\
&= \mathbb{E}[X_{t-1}] + \mathbb{E}[\epsilon_t] \\
&= \mathbb{E}[X_1] \\
\mathrm{Var}(X_t) &= \mathrm{Var}(X_{t-1} + \epsilon_t) \\
&= \mathrm{Var}(X_{t-1}) + \mathrm{Var}(\epsilon_t) \\
&= \mathrm{Var}(X_1) + (t-1) \cdot \sigma^2
\end{aligned}
$$

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)

$a_1, \ldots, a_p$ を定数とし $X_1, \ldots, X_p$ が初期値として与えられたとき以下で帰納的に定義される確率過程

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

- 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)

  $b_1, \ldots, b_q$ を定数とし，$X_1, \ldots, X_q$ が初期値として与えられたとき以下で帰納的に定義される確率過程

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

  - 記憶のあるホワイトノイズ (構成する部品を記憶)

自己回帰移動平均過程

- 定義 (次数 (p, q); ARMA(p, q))

  $a_1, \ldots, a_p, b_1, \ldots, b_q$ を定数とし $X_1, \ldots, X_{\max\{p,q\}}$ が初期値として与えられたとき以下で帰納的に定まる確率過程

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

  - 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

数値例

- 同じモデルに従うパス (系列) を複数観測してみる
  - 自己回帰過程 (AR 過程)
  - 移動平均過程 (MA 過程)
  - 自己回帰移動平均過程 (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

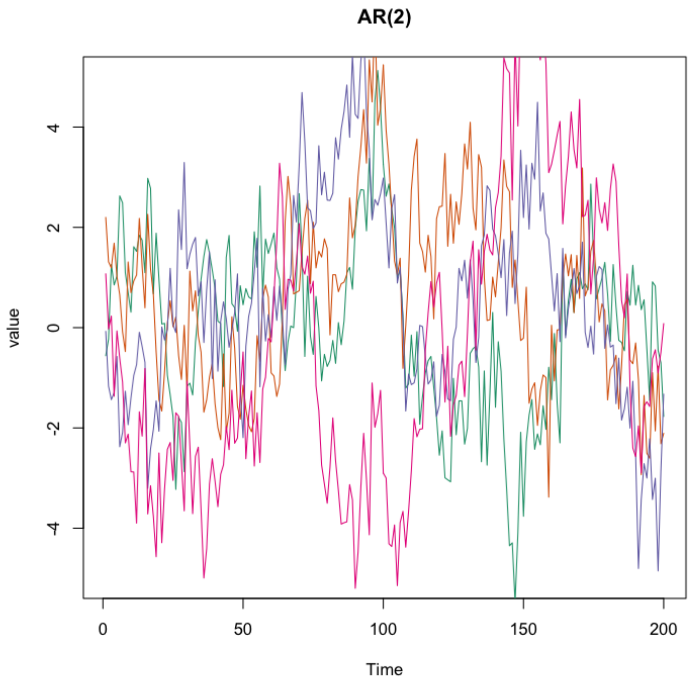

図 4: AR 過程

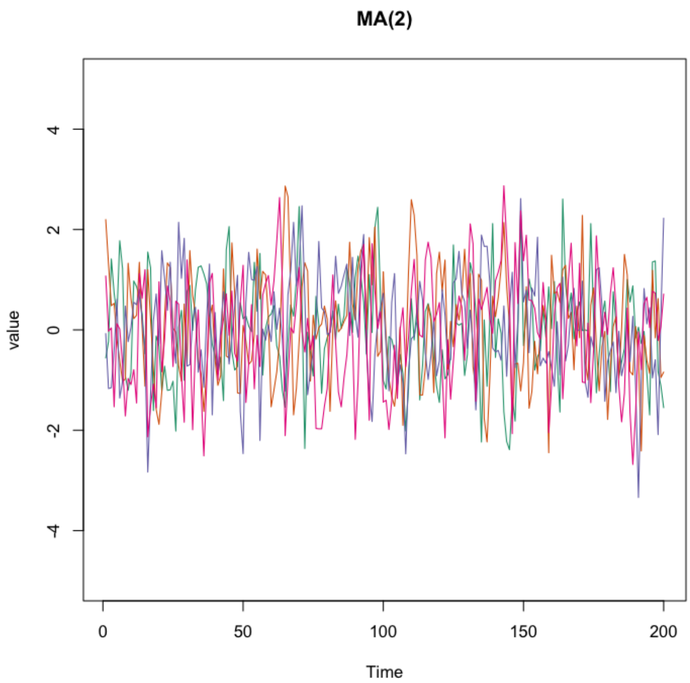

図 5: MA 過程

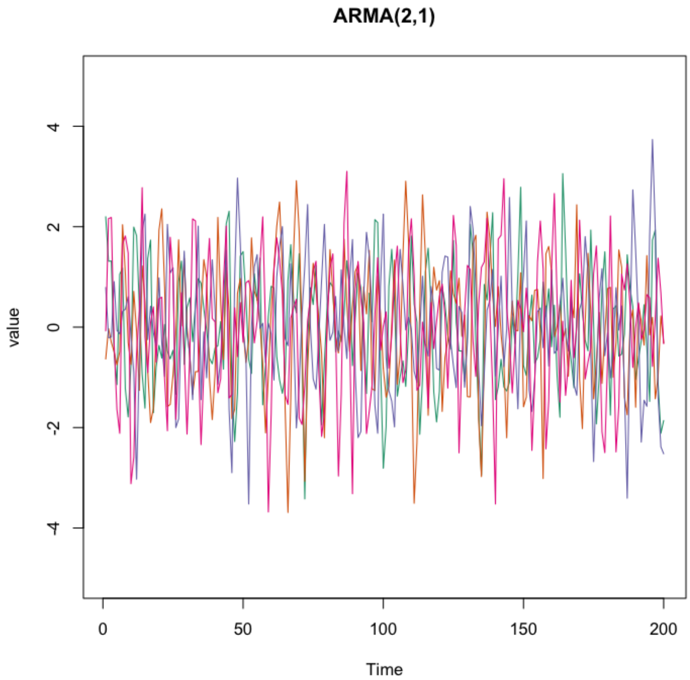

図 6: ARMA 過程

解答例

- 定義に従い帰納的に計算する

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[a_1 X_{t-1} + \epsilon_t] \\
&= a_1 \mathbb{E}[X_{t-1}] + \mathbb{E}[\epsilon_t] \\
&= a_1^{t-1} \mathbb{E}[X_1] \\
\mathrm{Var}(X_t) &= \mathrm{Var}(a_1 X_{t-1} + \epsilon_t) \\
&= a_1^2 \mathrm{Var}(X_{t-1}) + \mathrm{Var}(\epsilon_t) \\
&= a_1^{2(t-1)} \mathrm{Var}(X_1) + \frac{1 - a_1^{2(t-1)}}{1 - a_1^2} \cdot \sigma^2
\end{aligned}
$$

- 定義に従い帰納的に計算する

$$
\begin{aligned}
\mathbb{E}[X_t] &= \mathbb{E}[b_1 \epsilon_{t-1} + \epsilon_t] \\
&= b_1 \mathbb{E}[\epsilon_{t-1}] + \mathbb{E}[\epsilon_t] \\
&= 0 \\
\mathrm{Var}(X_t) &= \mathrm{Var}(b_1 \epsilon_{t-1} + \epsilon_t) \\
&= b_1^2 \mathrm{Var}(\epsilon_{t-1}) + \mathrm{Var}(\epsilon_t) \\
&= (b_1^2 + 1) \cdot \sigma^2
\end{aligned}
$$

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

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

– 特に X_t の分散は時点 t によらない ($h = 0$ の場合)

$$\mathrm{Var}(X_t) = \gamma(0), \quad (X_t \text{ は二乗可積分であることを仮定})$$

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

- 性質を特徴付ける統計量が観測値から得られない
  - 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 つの時系列から記述統計量の推測は一般にできない
- 擬似相関の問題
  - 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
  - 2 つの独立なランダムウォークは高い確率で “相関” を持つ
  - 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て分析を実行
  - 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

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

  - 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

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

  - 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)
    $$\mathrm{Cov}(X_t, X_{t+h}) = \gamma(h)$$
  - 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
    $$\mathrm{Cov}(X_t, X_{t+h})/\mathrm{Var}(X_t) = \gamma(h)/\gamma(0)$$
- 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
    $$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \bar{X})(X_{t+h} - \bar{X})$$
    $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$ は標本平均
  - ラグ h での自己相関の推定: 標本自己相関
    $$\hat{\gamma}(h)/\hat{\gamma}(0) = \frac{\sum_{t=1}^{T-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

数値例

- 同じモデルに従うパス (系列) の自己相関を比較してみる
  - 自己回帰過程 (AR 過程)
  - 移動平均過程 (MA 過程)
  - 自己回帰移動平均過程 (ARMA 過程)

演習

問題

- 以下の問に答えなさい
  - 定常な AR(p) 過程を考える．$\mathbb{E}[X_t] = 0$ であるとき，AR 過程の係数と自己共分散の間に成り立つ関係を考えなさい

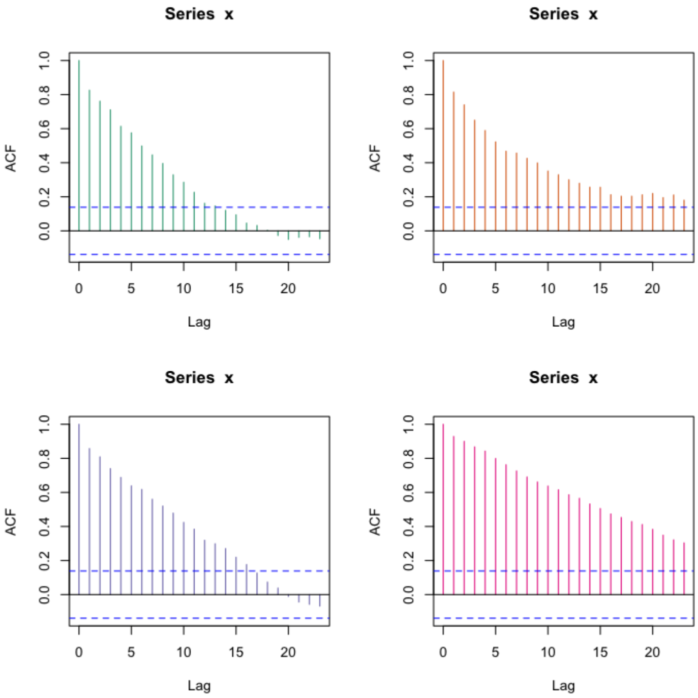

図 7: AR 過程の自己相関

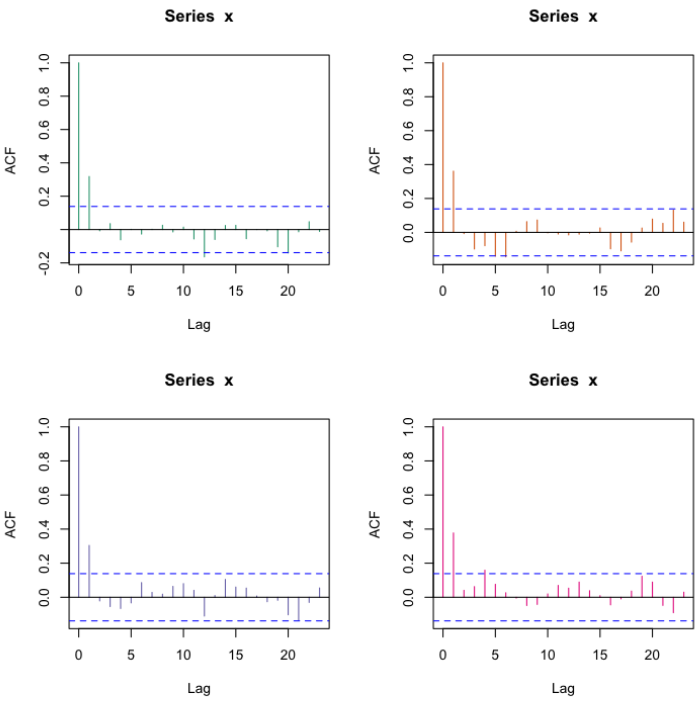

図 8: MA 過程の自己相関

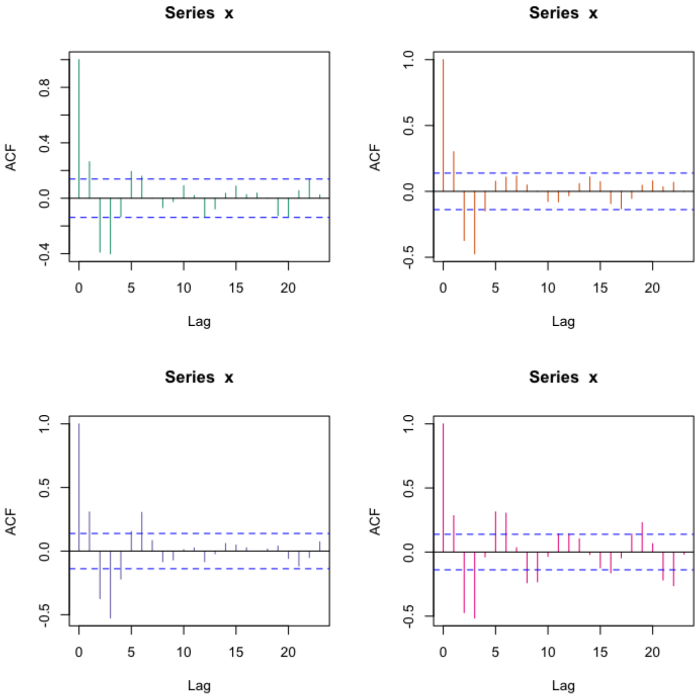

図 9: ARMA 過程の自己相関

解答例

- ラグ $h > 0$ の自己共分散を考える

$$
\begin{aligned}
\gamma(h) &= \mathbb{E}[X_t X_{t+h}] \\
&= \mathbb{E}[X_t(a_1 X_{t+h-1} + \cdots + a_p X_{t+h-p} + \epsilon_{t+h})] \\
&= a_1 \mathbb{E}[X_t X_{t+h-1}] + \cdots + a_p \mathbb{E}[X_t X_{t+h-p}] + \mathbb{E}[X_t \epsilon_{t+h}] \\
&= a_1 \gamma(h-1) + \cdots + a_p \gamma(h-p)
\end{aligned}
$$

- $1 \leq h \leq p$ を考えると以下の関係が成り立つ

$$
\begin{pmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{pmatrix} = \begin{pmatrix} \gamma(0) & \gamma(-1) & \dots & \gamma(-p+1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix}
$$

  Yule-Walker 方程式という

- Yule-Walker 方程式の性質
  - 行列は Toeplitz 行列と呼ばれる
  - $\gamma(h) = \gamma(-h)$ より行列は対称行列
  - 共分散の性質から行列が正定値 (非負定値)
  - 行列が正則ならば AR の係数は一意に決まる
  - 特殊な形を利用した高速な解法としては Levinson − Durbin アルゴリズムが知られている

次週の内容

- 第 1 日: 時系列の基本モデル
- **第 2 日: モデルの推定と予測**