

時系列解析

基本的なモデル

村田 昇

講義概要

- 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

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

- 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 を定数として以下で定義される確率過程

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

- **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

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

- **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
 - ティレンドのあるホワイトノイズ
 - ランダムウォーク

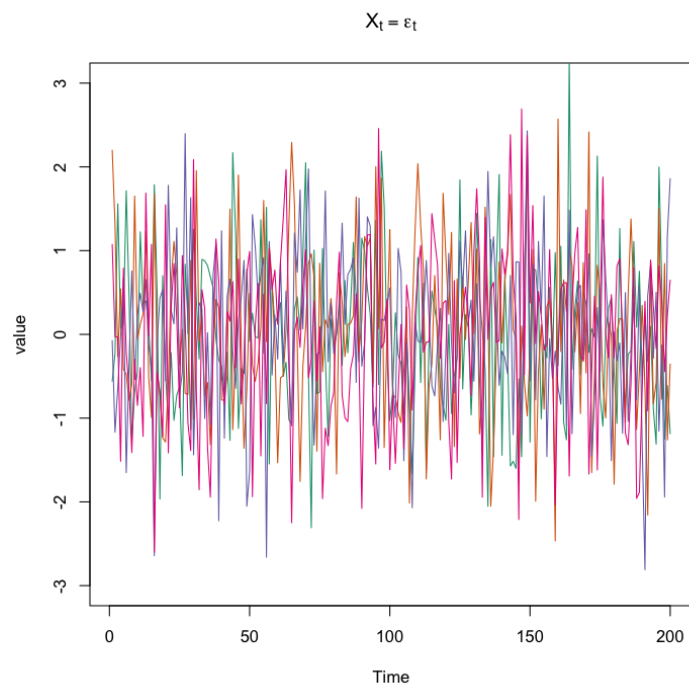


図 1: ホワイトノイズ

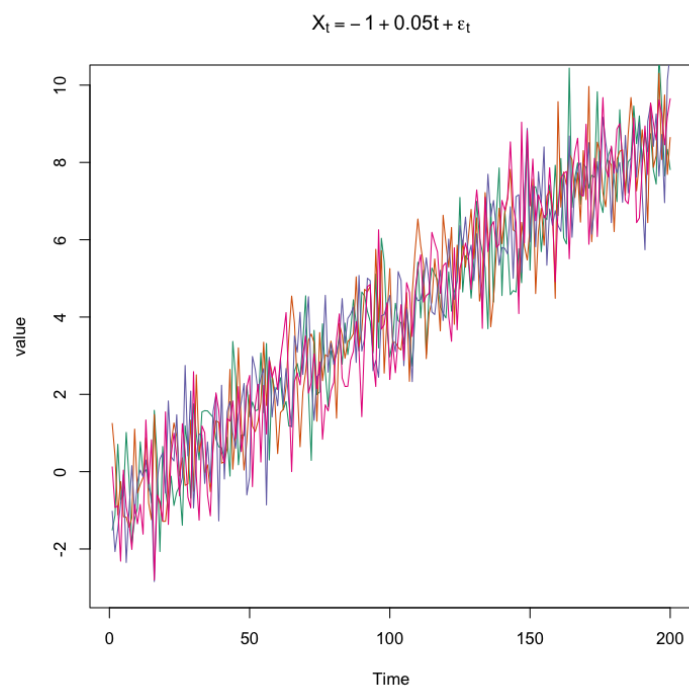


図 2: トレンドのあるホワイトノイズ

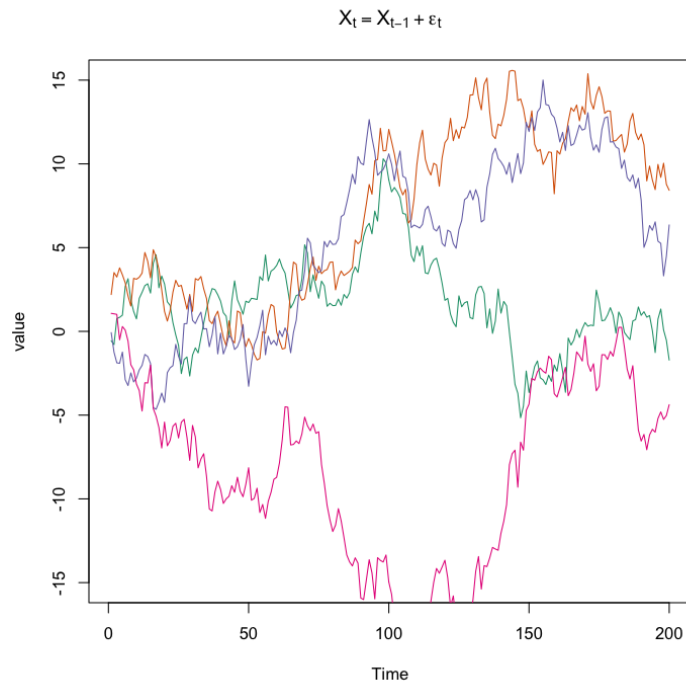


図 3: ランダムウォーク

演習

R: 時系列データの扱い

- 時系列のためのクラス：ts クラス
 - ベクトル (1 次元) データに時間に関する情報が付加される
 - 関数 `ts()` を用いてベクトルから変換
 - 多次元時系列も扱うことは可能
- 関数 `plot()` などの挙動はベクトルと異なる
 - プロットが既定値で折れ線
 - x 軸に時間の情報が表示
 - 通常は時間情報を利用して適切に処理してくれる
- より複雑な時系列を記述するためには `zoo` や `xts` などのパッケージがある

R: 関数 `ts()`

- 時系列クラス `ts` を作成する関数

```
ts(data = NA, start = 1, end = numeric(), frequency = 1)
## data: ベクトル, または行列 (データフレーム)
## start: 開始時刻
## end: 終了時刻
## frequency: 単位時間あたりの観測回数
ts(data = x) # t=1,2,... を添字とする時系列
ts(data = x, start = c(2020,1), frequency =12) # 2020 年 1 月からの月ごと
ts(data = x, start = c(2020,3), frequency =4) # 四半期ごと
```

- その他の詳細は `?ts` を参照

R: 関数 plot()

- 時系列クラスの描画

```
## 基本的な plot の使い方
x <- rnorm(240) # 正規分布のホワイトノイズ
plot(ts(x, start=c(2000,1), frequency=12)) # 2000 年からの毎月のデータを想定
## 複数の系列を表示する場合
y <- rt(240,df=4) # t-分布のホワイトノイズ
z <- ts(data.frame(x,y),
        start=c(2000,1), frequency=12)
plot(z, col="red") # 指定しなければ個別にグラフを描画
plot(z, plot.type="single", col=c("red", "blue"))
```

- その他の詳細は ?plot.ts を参照
- 表示が異なる関数 ts.plot() もあるので調べてみよう

練習問題

- 指定された確率過程を生成して図示しなさい
 - 平均 0, 分散 4 の正規分布に従うホワイトノイズ
 - 上記のホワイトノイズに初期値-1 で単位時刻あたり 1/20 で増加するトレンドを持つ確率過程
 - 上記のホワイトノイズから生成されるランダムウォーク

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以下で帰納的に定義される確率過程
$$X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$
 - 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, \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以下で定義される確率過程
$$X_t = b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} + \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma^2)$$
 - 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))
 $a_1, \dots, a_p, b_1, \dots, b_q$ 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以下で帰納的に定まる確率過程
$$\begin{aligned} X_t = & a_1 X_{t-1} + \dots + a_p X_{t-p} \\ & + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} + \epsilon_t, \\ & \epsilon_t \sim \text{WN}(0, \sigma^2) \end{aligned}$$

- $AR(p)$ 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)

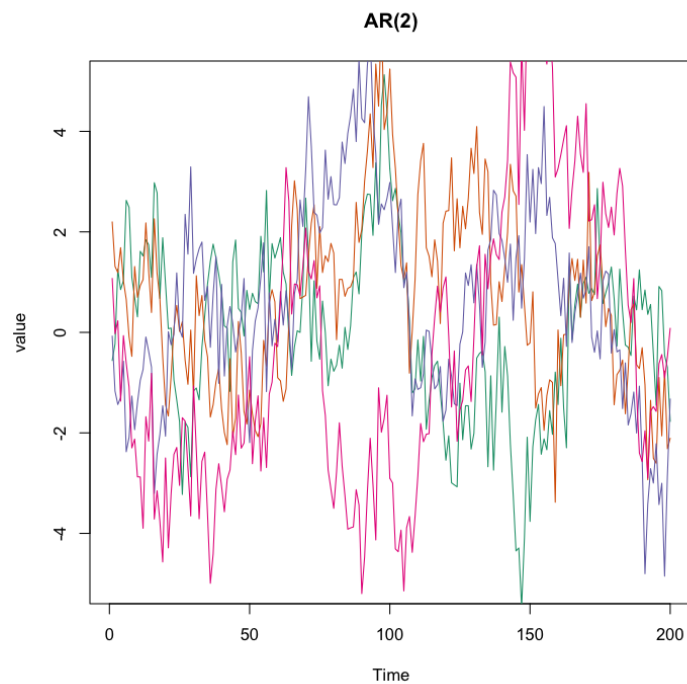


図 4: AR 過程

演習

練習問題

- 平均 0, 分散 1 のホワイトノイズを用いて, 以下の指定された確率過程を生成し, 図示を行いなさい
 - 係数 $a_1 = 0.67, a_2 = 0.26$ を持つ $AR(2)$ 過程
 - 係数 $b_1 = 0.44, b_2 = 0.08$ を持つ $MA(2)$ 過程
 - 係数 $a_1 = 0.8, a_2 = -0.64, b_1 = -0.5$ を持つ $ARMA(2,1)$ 過程

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

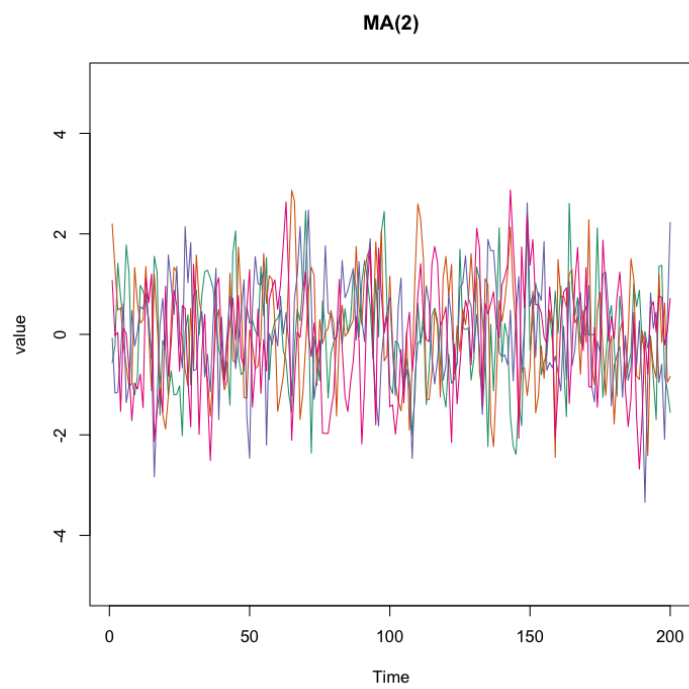


图 5: MA 过程

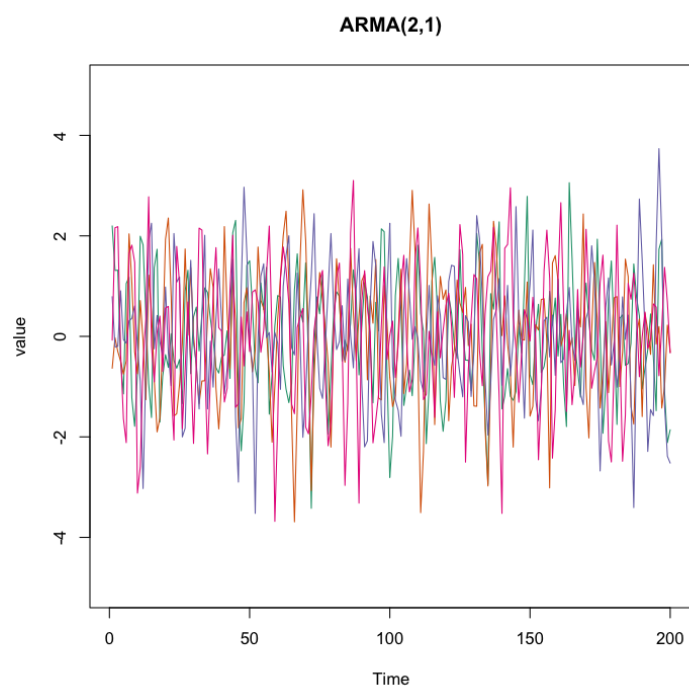


图 6: ARMA 过程

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

- 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)

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

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)

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

- 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

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

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

- 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

R: 関数 acf()

- 自己相関・自己共分散の計算

```
acf(x, lag.max = NULL,
    type = c("correlation", "covariance", "partial"),
    plot = TRUE, na.action = na.fail, demean = TRUE, ...)
## x: 時系列データ
## lag.max: 計算するラグの最大値
## type: 標準は相関, 共分散と偏相関を選ぶこともできる
## plot: 描画するか否か
## na.action: 欠損値の処理, 標準は欠損を含むと計算しない
## demean: 共分散の計算において平均を引くか否か
```

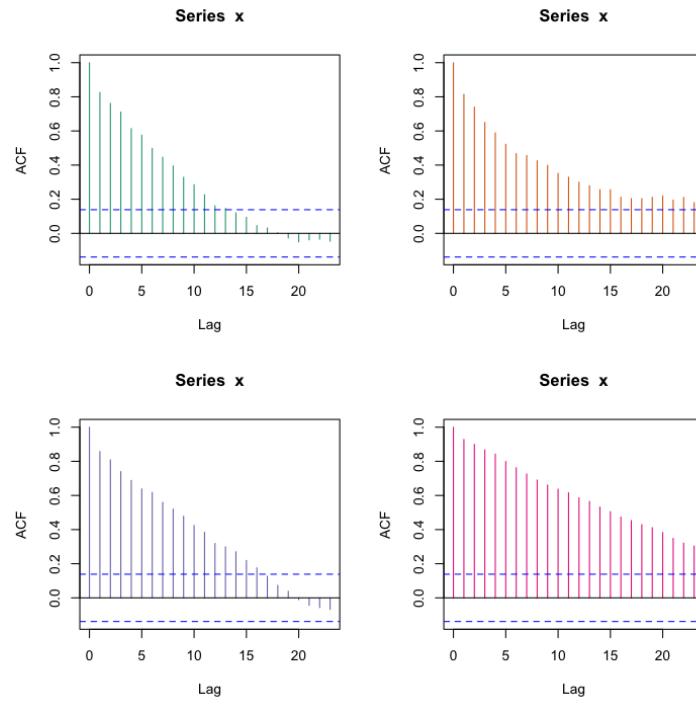


図 7: AR 過程の自己相関

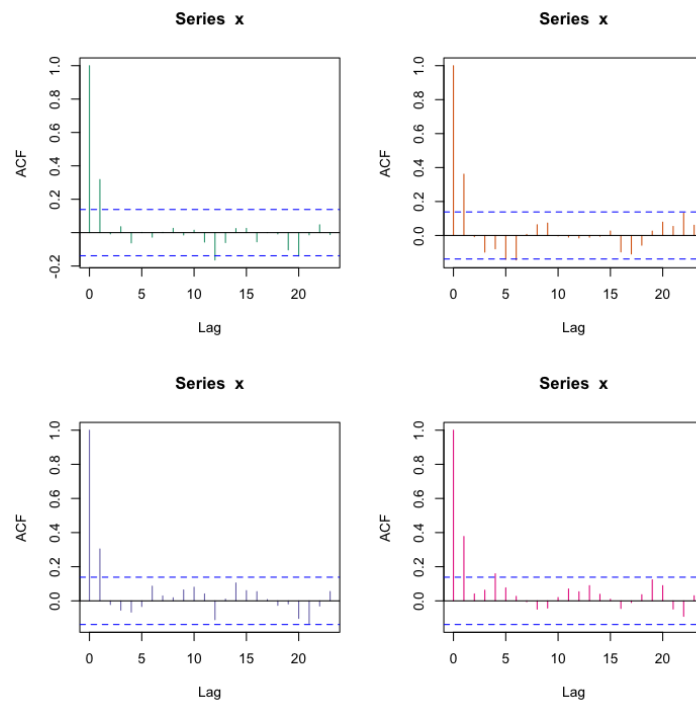


図 8: MA 過程の自己相関

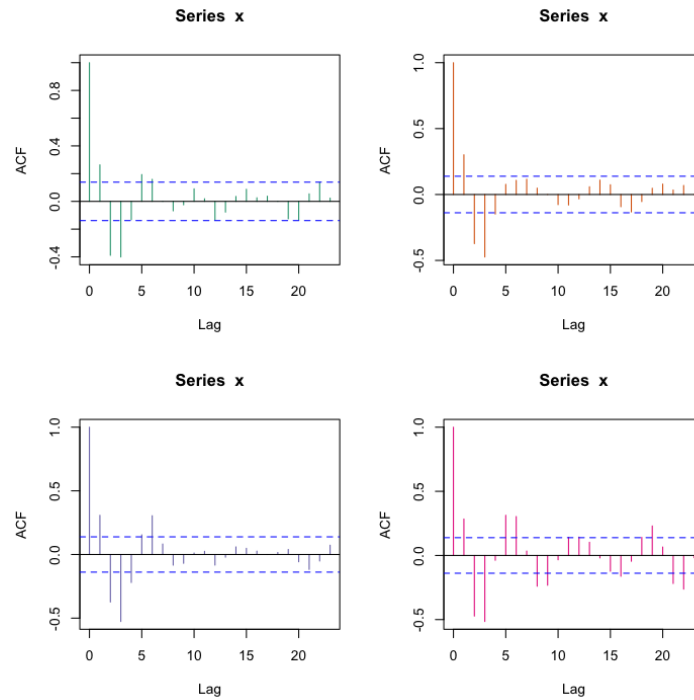


図 9: ARMA 過程の自己相関

- 詳細は ?acf を参照

練習問題

- 以下の問に答えなさい
 - 同じ AR 過程のモデルから生成した時系列の自己相関を比較しなさい (前の練習問題を利用すればよい)
 - MA 過程についても同様な比較を行いなさい
 - ARMA 過程についても同様な比較を行いなさい

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