時系列解析

モデルの推定と予測

村田 昇

講義の内容

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

時系列解析の復習

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

時系列モデルと定常性

- **確率過程**

  時間を添え字として持つ確率変数列

  $$X_t,\ t = 1, \dots, T$$

- **弱定常過程** : 以下の性質をもつ確率過程 X_t
  - X_t の平均は時点 t によらない
  - 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
  - 特に X_t の分散は時点 t によらない ($h = 0$ の場合)
- 多くの場合，弱定常性を考えれば十分なので単に **定常** ということが多い
- 定常でない確率過程は **非定常** であるという

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
  - **定常** な確率過程

トレンドのあるホワイトノイズ

- 定義

  μ, α を定数として

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

  で定義される確率過程

  - **非定常** な確率過程
  - トレンド項 (平均値の変化) は現象に応じて一般化される

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

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

  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・記憶のあるモデル
  - **非定常** な確率過程

自己回帰過程

- 定義 (次数 p の AR モデル)

  $a_1, \dots, a_p$ を定数とし，$X_1, \dots, X_p$ が初期値として与えられたとき,

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

  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
  - 無限長の記憶のある (忘却しながら記憶する) モデル
  - **定常にも非定常にもなる**

移動平均過程

- 定義 (次数 q の MA モデル)

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

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

  で定義される確率過程

  - 有限長の記憶のあるモデル
  - **定常** な確率過程

自己回帰移動平均過程

- 定義 (次数 (p, q) の ARMA モデル)

  $a_1, \dots, a_p, b_1, \dots, b_q$ 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

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

  で帰納的に定まる確率過程

  - AR・MA モデルの一般化・基本的な時系列モデル
  - **定常にも非定常にもなる**

自己共分散・自己相関

- 弱定常な確率過程 : $X_t,\ t = 1, \ldots, T$
  - 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)
    $$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h})$$
  - 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
    $$\rho(h) = \gamma(h)/\gamma(0) = \mathrm{Cov}(X_t, X_{t+h})/\mathrm{Var}(X_t)$$
- 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 $X_1, \ldots, X_T$ 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
    $$\hat{\gamma}(h)/\hat{\gamma}(0) = \frac{\sum_{t=1}^{T-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$$

AR モデルの推定

自己共分散・自己相関

- 平均 0 の弱定常な確率過程 : $X_t,\ t = 1, \ldots, T$
  - 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
    $$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) = \mathbb{E}[X_t X_{t+h}]$$
  - X_t と X_{t+h} の相関も t によらずラグ h のみで定まる
    自己相関係数
    $$\rho(h) = \mathrm{Cov}(X_t, X_{t+h})/\mathrm{Var}(X_t) = \gamma(h)/\gamma(0)$$

自己共分散と AR モデル

- AR(p) モデル :
  $$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + \epsilon_t$$

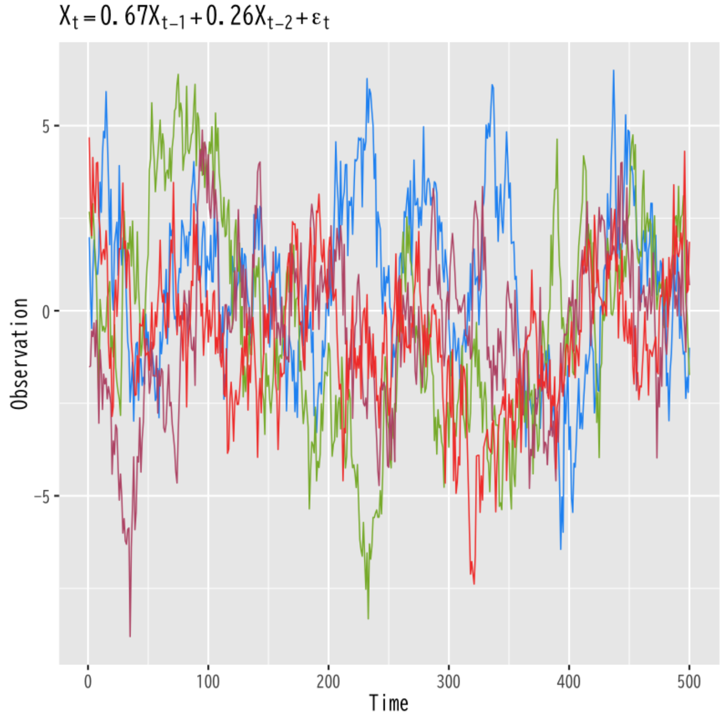

Figure 1: 同じモデルに従う AR 過程の例

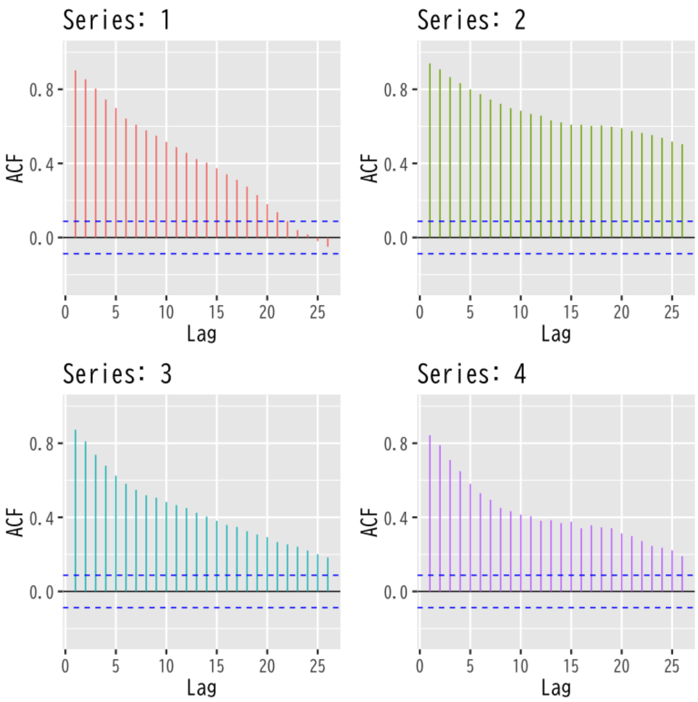

Figure 2: AR 過程の自己相関

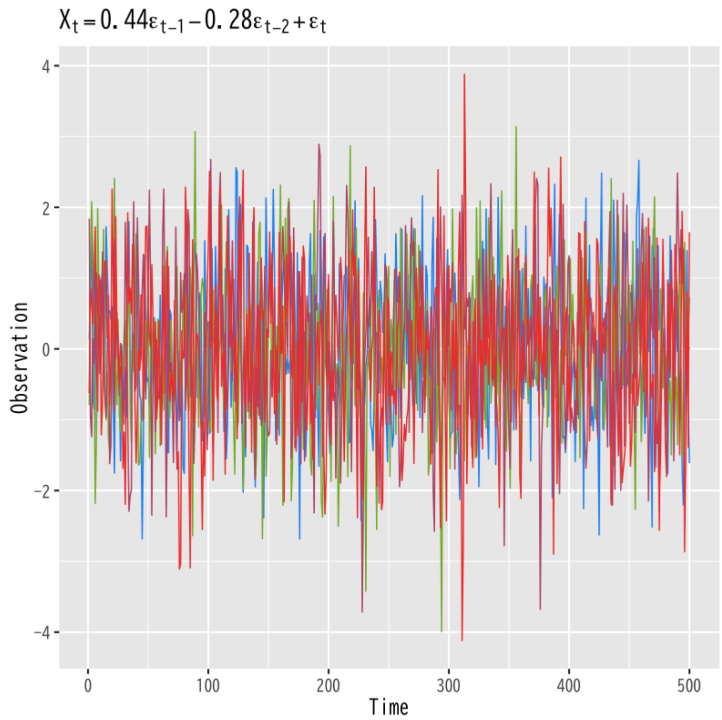

Figure 3: 同じモデルに従う MA 過程の例

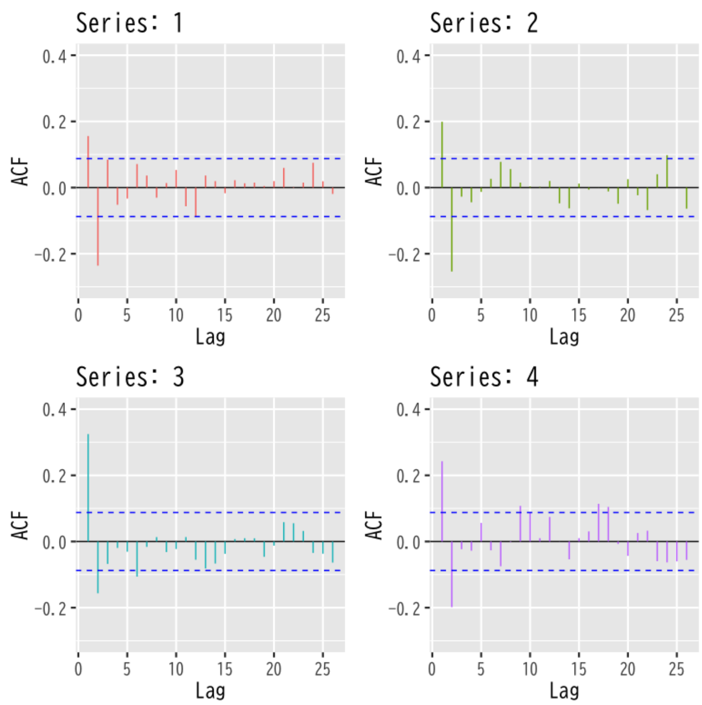

Figure 4: MA 過程の自己相関

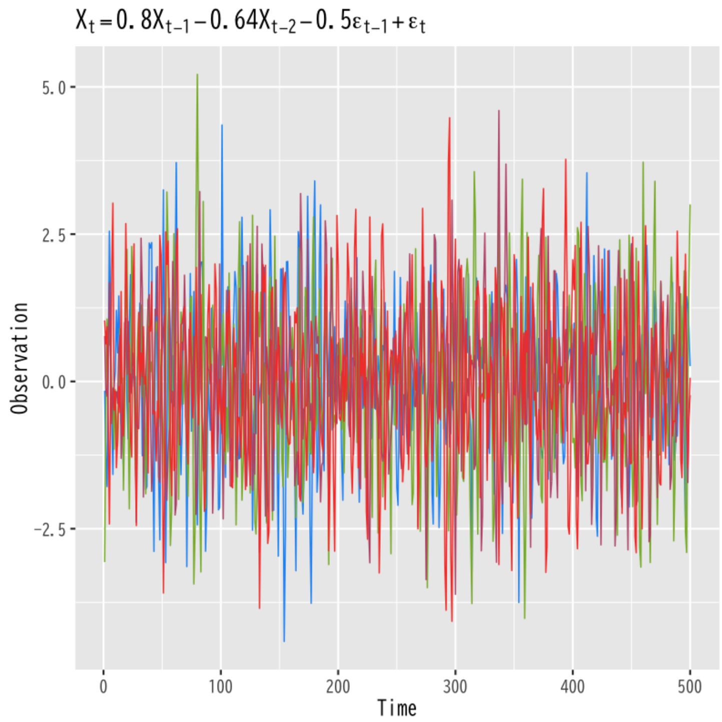

Figure 5: 同じモデルに従う ARMA 過程の例

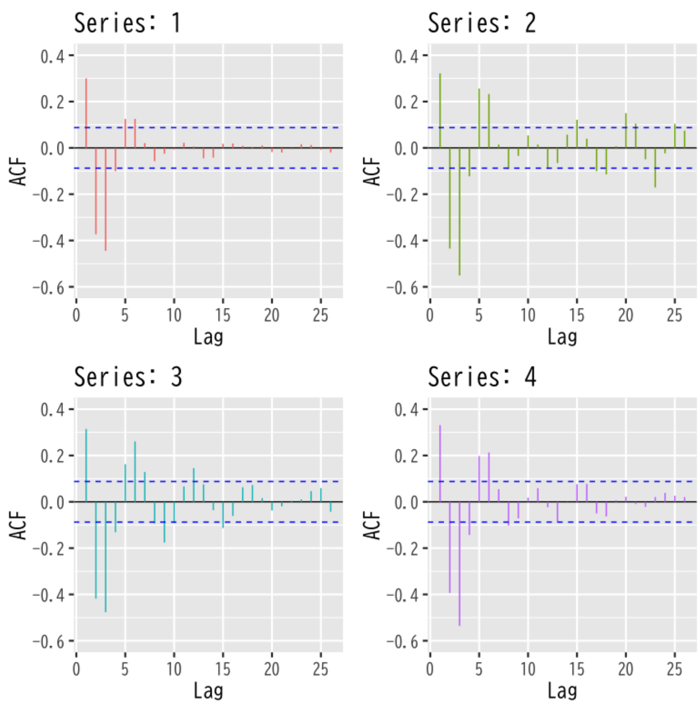

Figure 6: ARMA 過程の自己相関

- 係数と自己共分散の関係

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

Yule-Walker 方程式

- $1 \le h \le p$ を考えると以下の関係が成り立つ

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

  - 行列は Toeplitz 行列と呼ばれる
  - 行列が正則ならば AR の係数は一意に求まる

偏自己相関

- AR(p) モデル

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \cdots + a_p X_{t-p} + \epsilon_t$$

  - ラグ p の **偏自己相関係数**
    AR(p) モデルを仮定したときの a_p の推定値 (Yule-Walker 方程式の解)
  - ラグ p の特別な **自己相関係数**
    $a_1 = a_2 = \cdots = a_{p-1} = 0$ のときの $\rho(p)$ (**特殊なモデルにおける解釈**)

$$\mathbb{E}[X_t X_{t+p}] = a_p\mathbb{E}[X_t X_t] \ \Rightarrow\ \gamma(p) = a_p\gamma(0) \ \Rightarrow\ \rho(p) = a_p$$

モデルの推定に関する補足

- ARMA モデルの推定方法は主に以下の 3 つ
  - Yule-Walker 方程式 (AR 過程)
  - 最小二乗
    * 予測誤差の平方和の最小化
    * 回帰と同じだが，従属系列のため多重共線性に注意
  - 最尤推定
    * WN の分布を仮定して同時尤度関数を設定
    * 非線形最適化を行う
- 一般にモデルは近似なので，どの推定が良いかは問題による

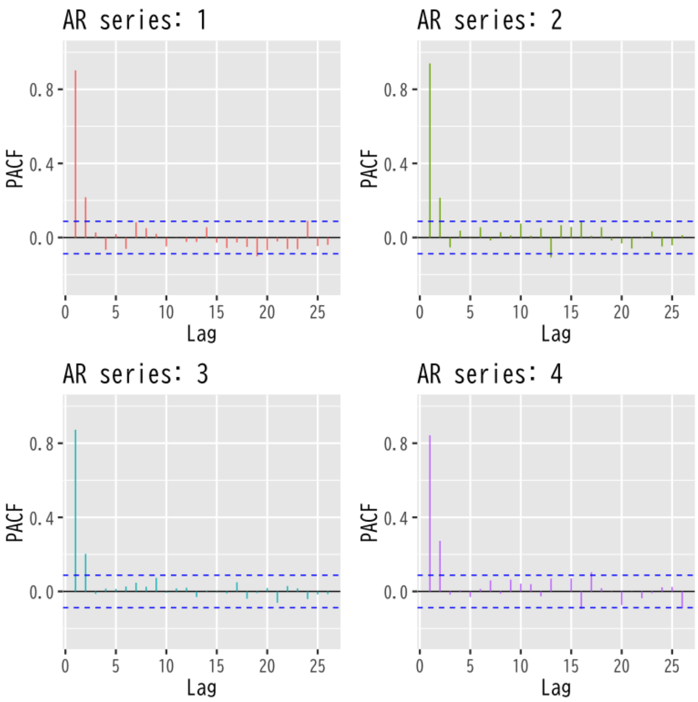

Figure 7: AR 過程の偏自己相関

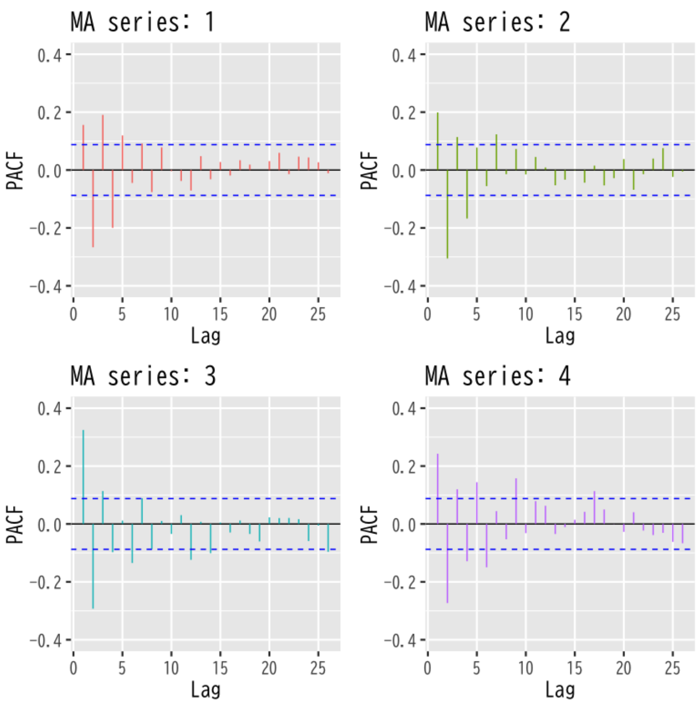

Figure 8: MA 過程の偏自己相関

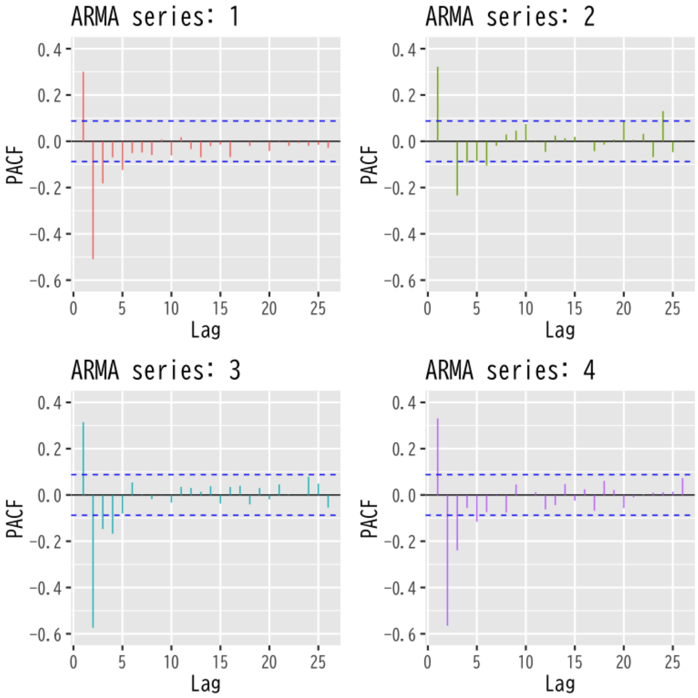

Figure 9: ARMA 過程の偏自己相関

非定常過程の変換

- 定常過程とみなせるように変換して分析
  - 階差の利用

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

    * ランダムウォーク : 階差をとるとホワイトノイズ (定常過程)
    * **ARIMA 過程** : 階差をとると ARMA 過程になる確率過程
  - 対数変換の利用

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

    * 対数変換と階差で微小な比率の変動を抽出

演習

問題

- 以下で定義される MA(1) について問に答えなさい

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

  - ラグ 2 までの自己共分散係数を求めなさい
  - 自己相関係数とパラメタ b_1 が満すべき方程式を求めなさい

解答例

- 平均 0 であることに注意して定義通り計算する

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

- ラグ 3 以降も自己共分散は 0 となることに注意する

$$
\begin{aligned}
\gamma(0) &= (b_1^2 + 1)\sigma^2 \\
\gamma(1) &= b_1\sigma^2
\end{aligned}
$$

  σ^2 を消去して以下が得られる

$$
\gamma(1)/\gamma(0) = \frac{b_1}{b_1^2 + 1} = \rho(1)
$$

$$
\rho(1)b_1^2 - b_1 + \rho(1) = 0
$$

  $\rho(1)$ の値によっては解が求められない場合もある

モデルによる予測

ARMA モデルによる予測

- 推定したモデルを用いて n 期先を予測
  - AR モデル : 観測時点までの観測値を用いて回帰
  - MA モデル : 観測時点までのホワイトノイズで回帰
  - ARMA モデル : 上記の複合
- いずれも n が大きいと不確定性が増大
- 階差による変換は累積 (階差の逆変換) により推定

分解モデルによる予測

- トレンド成分+季節成分+ランダム成分への分解

$$
X_t = T_t + S_t + R_t
$$

  あるいは

$$
X_t = T_t \times S_t \times R_t \qquad (\log X_t = \log T_t + \log S_t + \log R_t)
$$

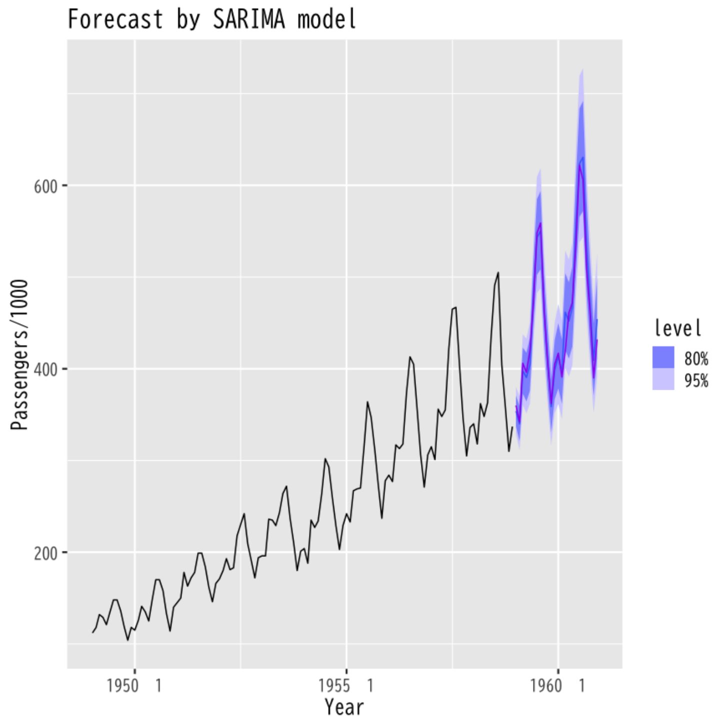

Figure 10: ARIMA モデル (階差あり ARMA) による予測

- トレンド成分 : 時間の関数やランダムウォークなどを想定
- 季節成分 : 周期的な関数を想定
- ランダム成分 : ARMA モデルなどを想定

- 分解の考え方
  - ランダム成分 : 適切な幅の移動平均が 0
  - 季節成分 : 1 周期の平均が 0

解析事例

COVID-19 の感染者数の分析

- 厚生労働省の COVID-19 のデータ
  - 陽性者数 (新規・累積)
  - 重症者数 (推移・性別・年齢別)
  - 死者数 (推移・性別・年齢別・累積)
  - 入院治療等を要する者等推移
  - 集団感染等発生状況
- 以下の解析で用いるデータ
  - 日毎の全国・各都道府県の新規陽性者数 (感染者数) `https://covid19.mhlw.go.jp/public/opendata/newly_confirmed_cases_daily.csv`

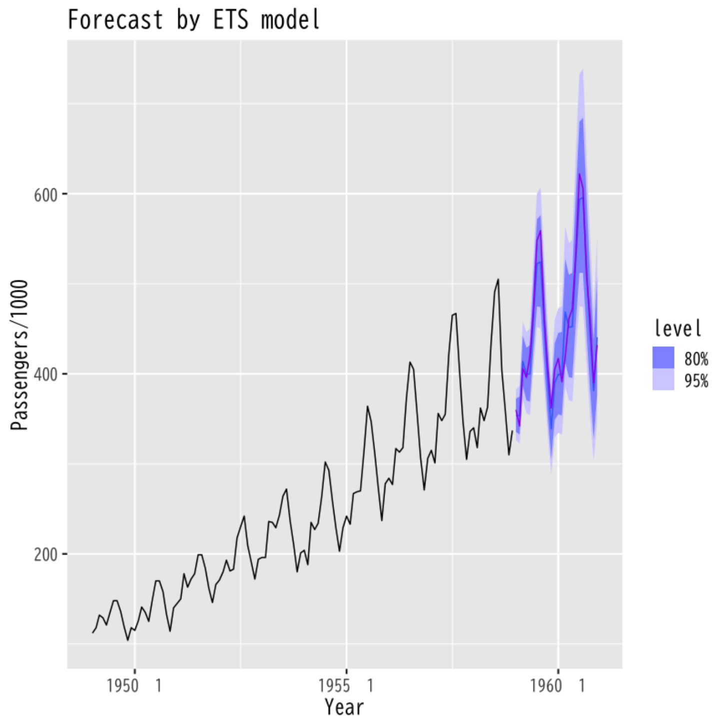

Figure 11: ETS モデル (expornential smoothing) による予測

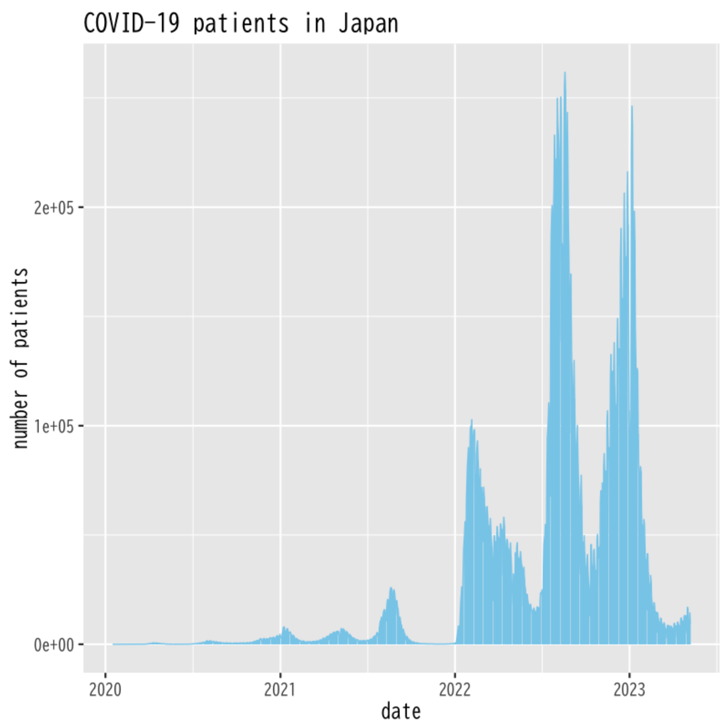

Figure 12: 全国の感染者数

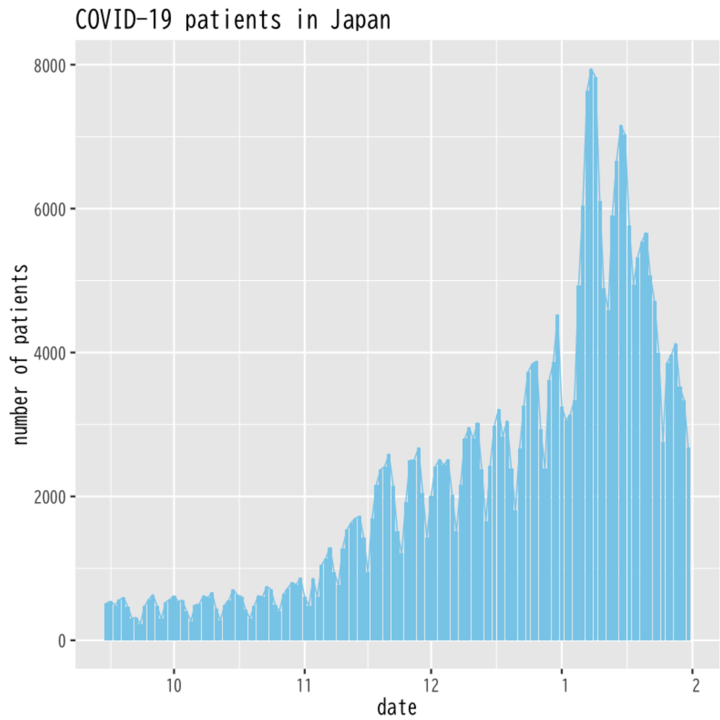

Figure 13: 第 3 波の感染者数

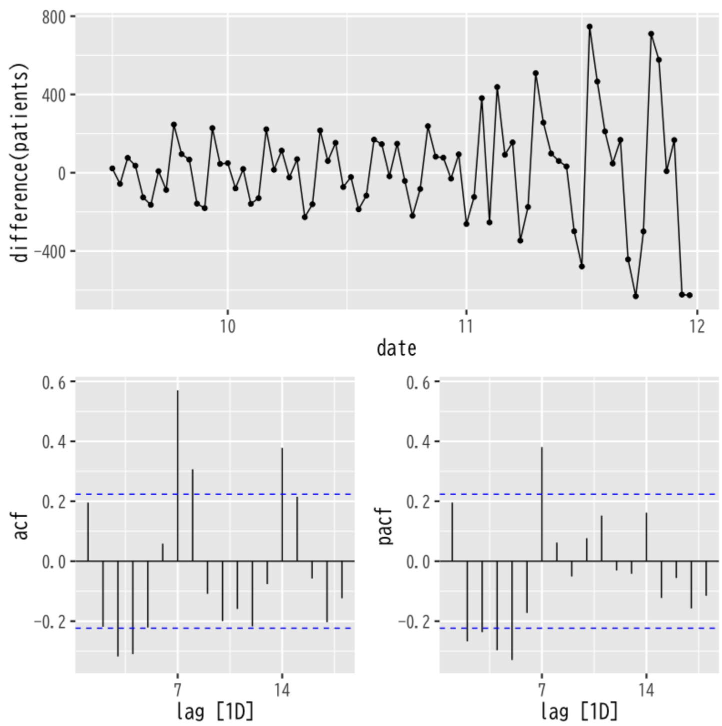

Figure 14: 時系列 (階差)

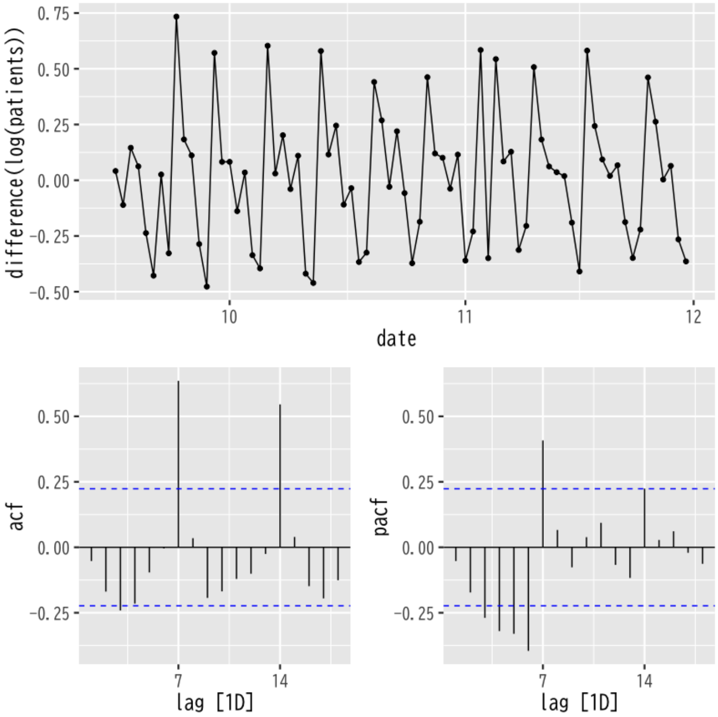

Figure 15: 時系列 (対数変換+階差)

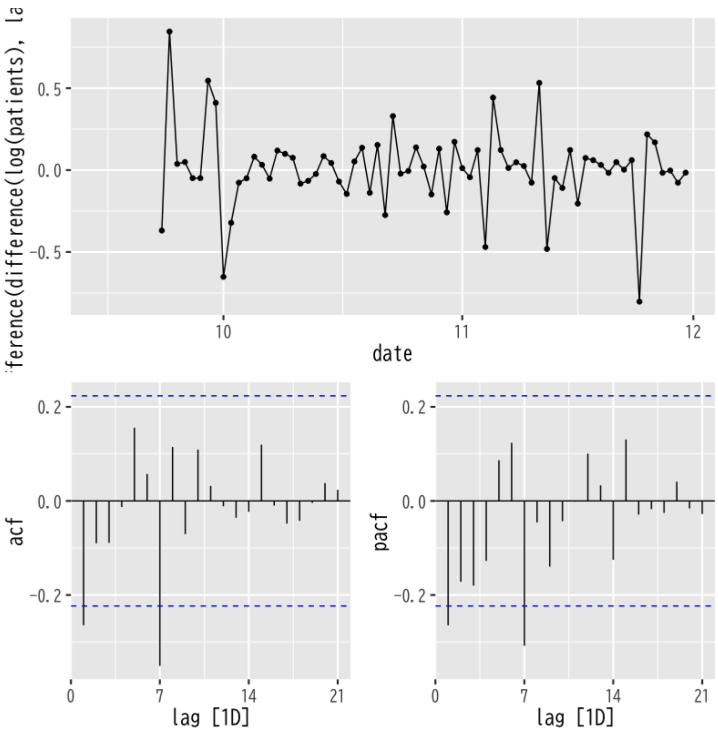

Figure 16: 時系列 (対数変換+階差+7 日階差)

感染者数の推移

第 3 波における感染者数の推移

基礎分析 (分析対象：2020/9/15-11/30)

ARIMA モデルによる推定

- 推定された ARIMA モデル

```
Series: patients
Model: ARIMA(1,1,1)(2,0,0)[7]
Transformation: log(patients)

Coefficients:
         ar1      ma1    sar1    sar2
      0.4493  -0.8309  0.3709  0.4232
s.e.  0.1635   0.0981  0.1212  0.1353

sigma^2 estimated as 0.03811:  log likelihood=15.04
AIC=-20.07   AICc=-19.21   BIC=-8.42
```

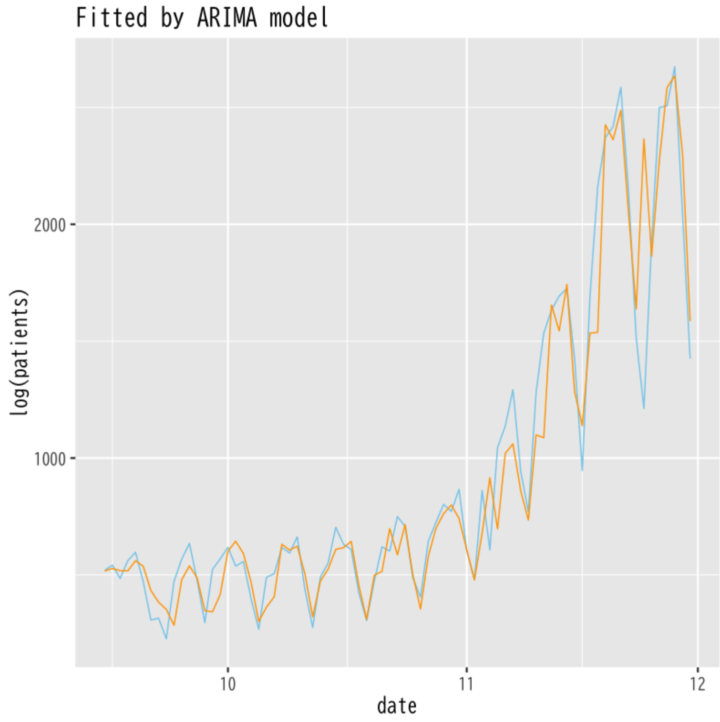

Figure 17: あてはめ値

まとめ

- 感染者数の推移は非定常なデータ
- 構造が時不変と考えられる区間を捉えれば
  - 時系列の適切な変換 (指数的な増大のため対数変換)
  - 基本的な ARMA モデル (階差系列に ARMA モデルを適用)

  の組み合わせである程度の分析は可能

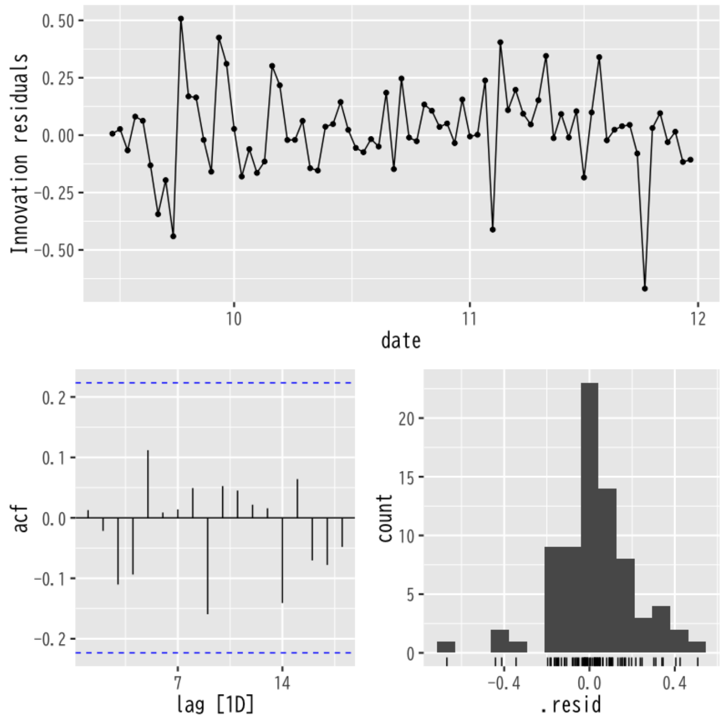

Figure 18: 診断プロット

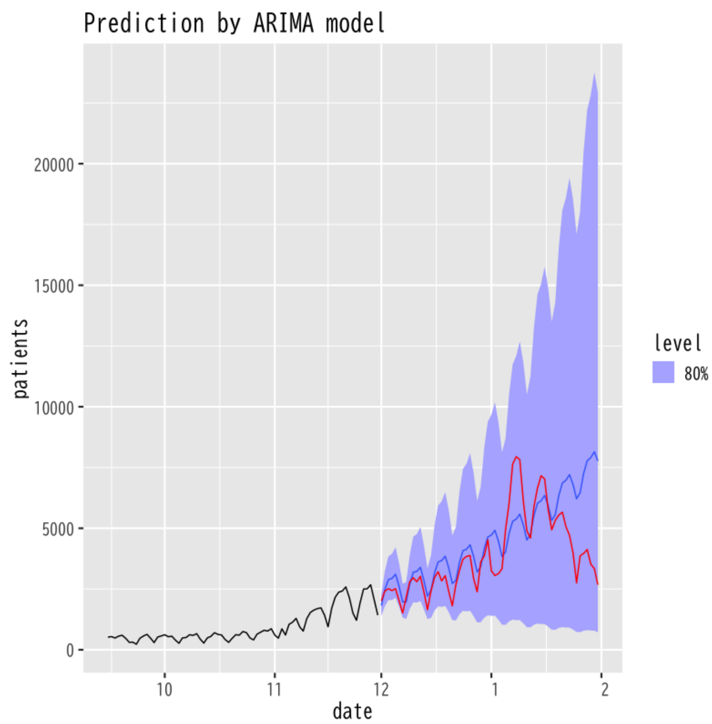

Figure 19: 予測値 (60 日分，80%信頼区間)