統計データ解析 (II) 第 12 回

小池祐太

2018 年 7 月 5 日

時系列解析

- **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ことや，異なる時点間での観測データの間の従属関係が重要であることが特徴
- 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

時系列解析

- 統計学では, 時系列データは**確率過程** (時間を添え字として持つ確率変数列: X_t, $t = 0, 1, \ldots, T$（あるいは $t = 1, ..., T$）) によってモデル化する
- 以下に時系列分析で利用されるいくつかの代表的な確率過程およびその R でのシミュレーション法について述べる
- なお, R には時系列データを表すためのクラス `ts` が用意されており, 例えば関数 `plot()` を適用した際の挙動が通常のベクトルと異なる (`ts` クラスのデータのプロットはデフォルトで折れ線プロットとなる)
  - 通常のベクトル `x` は関数 `ts()` を適用することで `ts` クラスに変換できる

ホワイトノイズ

- 平均 0，分散 σ^2 で互いに無相関な確率変数列からなる確率過程を**ホワイトノイズ** $\mathrm{WN}(0, \sigma^2)$ と呼ぶ
- 平均 0 で有限の分散を持つ同一の分布に従う独立な確率変数列がホワイトノイズの典型的な例
- このことから, ホワイトノイズのシミュレーションには乱数発生器 (`rnorm()` や適当な自由度の `rt()` など) で行うことが多い
- 実行例 `wn.r`

トレンドのあるホワイトノイズ

- 確率過程 ϵ_t $(t = 1, \ldots, T)$ をホワイトノイズとする
- μ, α を定数とするとき,

$$X_t = \mu + \alpha t + \epsilon_t$$

  で与えられる時系列データ X_t, $t = 1, \ldots, T$ を**トレンドのあるホワイトノイズ**と呼ぶ
- 項 "$\mu + \alpha t$" は**トレンド**と呼ばれる
- トレンドのあるホワイトノイズは, 時間とともに平均が変動する時系列モデルの一つである
- トレンド項としてここでは t の 1 次式を考えているが, 高次の多項式や非線形関数 (指数関数, 三角関数など) を考えることもある
- 実行例 `trend.r`

ランダムウォーク

- X_1 を定数もしくは確率変数として, 式

$$X_t = X_{t-1} + \epsilon_t, \qquad (t = 2, \ldots, T)$$

  で帰納的に定義される確率過程 $X_t,\ t = 1, \ldots, T$ を**ランダムウォーク**と呼ぶ
- ここで $\epsilon_t,\ t = 2, ..., T$ は同一の分布に従う独立な確率変数列 (i.i.d.)
- ランダムウォークは分散が時間とともに増加する時系列モデルの一例である
- 実行例 `rw2.r`

自己回帰モデル (AR モデル)

- $\epsilon_t,\ t = p+1, ..., T$ を $\mathrm{WN}(0, \sigma^2)$ とする
- $a_1, \ldots, a_p$ を定数とする．$X_1, \ldots, X_p$ が初期値として与えられたとき，

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \epsilon_t, \qquad t = p+1, \ldots, T$$

  で帰納的に定まる確率過程を**次数 p の自己回帰過程**もしくは **AR(p) モデル**と呼ぶ (AR は autoregressive の略)
- $p = 1,\ a_1 = 1,\ \epsilon_t$ が独立で同一の分布に従う場合，ランダムウォークとなるため，AR モデルはランダムウォークの一般化と考えられる
- 実行例 `ar2.r`

移動平均モデル (MA モデル)

- ϵ_t, $t = q+1, ..., T$ を WN$(0, \sigma^2)$ とする
- $b_1, \ldots, b_q$ を定数とする．$X_1, \ldots, X_q$ が初期値として与えられたとき，

$$X_t = b_1\epsilon_{t-1} + \cdots + b_q\epsilon_{t-q} + \epsilon_t, \qquad t = q+1, \ldots, T$$

  で定まる確率過程を**次数 q の移動平均過程**もしくは **MA(q) モデル**と呼ぶ (MA は moving average の略)
- 実行例 `ma2.r`

自己回帰平均移動モデル (ARMA モデル)

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

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + b_1 \epsilon_{t-1} + \cdots + b_q \epsilon_{t-q} + \epsilon_t,$$
$$t = \max\{p, q\} + 1, \ldots, T$$

  で帰納的に定まる確率過程のモデルを**次数** (p, q) **の自己回帰平均移動モデル**もしくは **ARMA**(p, q) **モデル**と呼ぶ
- ARMA$(p, 0)$ モデルは AR(p) モデルであり，ARMA$(0, q)$ モデルは MA(q) モデルであるから，ARMA モデルは AR モデルと MA モデルの一般化である
- ARMA モデルは単純な形ながら異なる時点間の観測データの従属構造を柔軟に記述できるため，基本的な時系列モデルとして広く利用されている
- 実行例 `arma2.r`

(弱) 定常性

- 確率過程 $X_t,\ t = 1, \ldots, T$ が次の 2 つの性質をもつとき, **(弱) 定常**であるという:
  (i) X_t の平均は時点 t によらない.
  (ii) 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 ($h = 0$ の場合を考えればよい).
- 定常でない確率過程は**非定常**であるという. 前節で説明した確率過程の定常性は以下のようにまとめられる.
  - **定常過程** ホワイトノイズ, MA モデル
  - **非定常過程** 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確率過程** AR モデル, ARMA モデル

(弱) 定常性

- 非定常過程は平均や分散といった基本的な統計量が時間によって変動してしまうため扱いが難しい
- そのため, 非定常過程の分析の際には対数変換や階差をとる変換等によって定常過程とみなせるように変換したあと分析を実行することが多い (例えばランダムウォークは階差をとればホワイトノイズとなって定常過程となる)

自己共分散・自己相関

- X_t, $t = 1, \dots, T$ が定常過程の場合, その定義から X_t と X_{t+h} の共分散は時点 t によらずラグ $h \geq 0$ のみで定まる
  - この共分散をラグ h での**自己共分散**と呼ぶ
- また, 分散も時点によらないので, X_t と X_{t+h} の相関も時点 t によらずラグ $h \geq 0$ のみで定まる
  - この相関をラグ h での**自己相関**と呼ぶ
- 通常, ラグ h での自己共分散を観測データ $X_1, \dots, X_T$ から推定するには, 標本自己共分散

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

を用いる. ただし, $\bar{X} = \frac{1}{T} \sum_{t=1}^{T} X_t$ は標本平均である

自己共分散・自己相関

- 同様に，ラグ h での自己相関を観測データ $X_1, \ldots, X_T$ から推定するには，標本自己相関

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

  を用いる
- 自己共分散および自己相関は，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である
- R では関数 `acf()` によって計算できる
- 実行例 `acf2.r`, `eustock.r`

AR モデルのあてはめ

- R には, 与えられた時系列データに適切な定常 AR モデルをあてはめるための関数 `ar()` が用意されている
- 実行例 `ar-estimation.r`

ARMA モデルのあてはめ

- 指定された次数の定常 ARMA モデルを与えられた時系列データにあてはめるための関数として `arima()` が用意されている
- ただし, 関数 `ar()` と異なり関数 `arima()` には適切な次数を決定する機能は備わっていない
- そのため, 次数の決定は試行錯誤で行うか, パッケージ `forecast` の関数 `auto.arima()` を利用するとよい
- 実行例 `arma-estimation.r`

予測

- 推定されたモデルを使って数期先の時系列データの値を予測するには関数 `predict()` を使う
- n 期先までのデータを予測する場合, オプション `n.ahead` に n を指定すればよい
- 実行例 `ts-predict.r`