時系列の AR,MA,ARMA モデル

樋口さぶろお https://hig3.net

龍谷大学大学院 理工学研究科 数理情報学専攻

理論物理学特論　L09 (2022-11-16 Wed)

最終更新: Time-stamp: "2022-11-16 Wed 08:16 JST hig"

今日の目標

- 確率過程の自己回帰モデル AR(m) を説明できる.
- 白色雑音, 自己相関係数, 定常性の意味を説明できる.
- 確率過程の移動平均モデル MA(k) を説明できる.

L07-Q2

マルコフ連鎖の定常分布

次の転置推移確率行列を持つ 状態空間 $\{0,1,2\}$ 上のマルコフ連鎖を考えよう.

$$M = \begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{2}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

なお, M の固有値固有ベクトルは $\lambda = 1, \frac{3}{4}, \frac{1}{4}, \quad \vec{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$ であることを使ってよい.

1. 定常分布をすべて求めよう.
2. 初期分布 $\vec{p}(0) = \frac{1}{3}\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ のとき $\vec{p}(t)$ を求めよう.
3. 上のとき, $\log|\vec{p}(t) - \vec{p}(\infty)|$ の $t \to +\infty$ の振る舞いを求めよう.

1. 固有値 $\lambda = 1$ の固有ベクトルである確率ベクトルは $\frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}$ のみであり，これが唯一の定常分布
2. $$\vec{p}(t) = \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{1}{6}\begin{pmatrix}-1\\0\\1\end{pmatrix}\left(\frac{3}{4}\right)^t + \frac{1}{6}\begin{pmatrix}1\\-2\\1\end{pmatrix}\left(\frac{1}{4}\right)^t$$
3. 式でまともに計算すると，固有ベクトルの内積 $\vec{u}_2 \cdot \vec{u}_3$ とか出てきてたいへん，だけど，$t \to +\infty$ では，$\vec{p}(t) - \vec{p}(\infty)$ の第 3 項が (第 2 項と比べて) 無視できて，

$$\begin{aligned}\log|\vec{p}(t) - \vec{p}(\infty)| &= \log\left|-\frac{1}{6}\begin{pmatrix}-1\\0\\1\end{pmatrix}\left(\frac{3}{4}\right)^t\right| + \text{もっと小さい項}\\ &= t\log\frac{3}{4} + \log\frac{\sqrt{2}}{6} + \text{もっと小さい項}\end{aligned}$$

となる．

ギブスサンプリングの一般論

2 つの確率変数 X, Θ

- x, θ に初期値を代入
- while(1)
  - $X \sim X|\theta, \{y_i\}$ からサンプルして x を更新
  - $\Theta \sim \Theta|x, \{y_i\}$ からサンプルして θ を更新

L08-Q1

Quiz(ギブスサンプリング)

乱数やサイコロは使えず, コイン (ただし, 表のでる確率 p は自由に設定できる), すなわち $\mathrm{B}(1,p)$ の乱数だけが使える状況を考えよう.

離散型確率変数の同時分布 $f(\theta, y) =$

$y \backslash \theta$	1	2
10	1/10	2/10
20	3/10	4/10

を考える.

1. 初期値が $\theta = 1$ だったとき, ギブスサンプリングのループの 1 回目が終わった時点で得られる標本の分布を求めよう.
2. 上の同時分布が, ギブスサンプリングの定めるマルコフ連鎖の定常分布になっていることを確かめよう.

ここまで来たよ

8 メトロポリス・ヘイスティングス法とギブスサンプリング

9 時系列の AR,MA,ARMA モデル
- 自己回帰モデル AR(m)
- 自己回帰モデル AR(m)
- k 次の移動平均モデル MA(k) モデル

確率過程

定義 (確率過程)

時刻 t に対応した連続型確率変数の列 $\{X_t\}_{t=0,1,\dots}$ を確率過程という.

X_t は独立でも同分布でもない
サイズ 1 の標本抽出 (1 試行) で $X_0, X_1, \dots$ のぜんぶ (=サンプルパス sample path) が得られる.
t, X_t の位置づけは, 単回帰の (x, Y) と異なる.

単回帰, $Y = \beta_0 + \beta_1 x + \epsilon_x, \epsilon_x \sim \mathrm{N}(0, \sigma^2)$ は漸化式ではない. 右辺は x の 1 次関数

t, x は確率変数ではないという点は同じ

1 次の自己回帰モデル AR(1)

確率過程の単純な例

定義 (1 次の自己回帰モデル AR(1))

$t = 1, 2, \ldots$: 時刻
$\{\epsilon_t\}_{t=1,\ldots}$,$\{X_t\}_{t=0,1,\ldots}$: 連続型確率変数の列
$\phi_1 \in \mathbb{R}$: パラメタ
$X_t (t \geq 1)$ は次のように定まる.

$$X_t = \phi_1 \cdot X_{t-1} + \epsilon_t. \quad (t \geq 1)$$

ただし ϵ_t は次を満たす.

$$\mathrm{E}[\epsilon_t] = 0 \tag{WN1}$$
$$\mathrm{E}[\epsilon_t \epsilon_s] = \sigma^2 \delta_{ts} \tag{WN2}$$
$$\mathrm{E}[X_t \epsilon_s] = 0 \quad (t < s) \tag{PI}$$

定義 (白色雑音 (white noise))

WN1, WN2 を満たす ϵ_t を**白色雑音** (white noise) という.

形容詞 white は特徴がないことを表現.

AR(1) の定義方式への感想

たとえば, $\epsilon_t \sim \mathrm{N}(0, \sigma^2)$ (独立同分布) なら WN1, WN2 を満たすが, 分布を具体的に指定せず, 母期待値のみを制限する弱い条件になっている. 例えば, $\mathrm{E}[\epsilon_t^4]$ は様々な値である可能性を残している.

PI はジャンプ幅が場所によらない (よって壁はない) という条件.

例

- $t =$ シーズン, $X_t =$ タイガースの勝率の 5 割との差, $-1 < \phi < 0$.
- ランダムウォーク
- $t =$ 日, $X_t =$ 気温の, 平年平均気温との差, $0 < \phi < 1$.

AR(1) の漸化式の計算

$$
\begin{aligned}
X_t =& \phi_1 X_{t-1} + \epsilon_t \\
=& \phi_1(\phi_1 X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
=& \phi_1(\phi_1(\phi_1 X_{t-3} + \epsilon_{t-2}) + \epsilon_{t-1}) + \epsilon_t \\
=& \cdots \\
=& \phi_1^t X_0 + \phi_1^{t-1}\epsilon_1 + \phi_1^{t-2}\epsilon_2 + \cdots + \phi_1^0 \epsilon_t \\
=& \phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}
\end{aligned}
$$

時系列の共分散

定義 (母自己共分散, 母自己相関係数)

時系列 $\{X_t\}$ に対して, s 次の

- 母自己共分散 (auto covariance)$\mathrm{Cov}[X_t, X_{t-s}]$.
- 母自己相関係数 (autocorrelation)$r_t(s) = \frac{\mathrm{Cov}[X_t,X_{t-s}]}{\sqrt{\mathrm{V}[X_t]\mathrm{V}[X_{t-s}]}} \overset{\text{定常なとき}}{=} \frac{\mathrm{Cov}[X_t,X_{t-s}]}{\mathrm{Cov}[X_t,X_t]}$.

$s=0$ 次の母自己共分散はふつうの母分散, 母自己相関係数は 1.
s をラグ (lag) という.

時系列の定常性

定義 (時系列の (弱い意味での) 定常性)

時系列 $\{X_t\}$ が (弱い意味で) 定常 (stationary) であるとは, 次の両方を満たすこと.

- 母平均値 $\mathrm{E}[X_t]$ が t によらないこと.
- 各ラグ s に対して, s 次の母自己共分散 (auto covariance)$\mathrm{Cov}[X_t, X_{t-s}]$ が t によらないこと.

- 分布や高次のモーメントは t に依存してもよい (依存しない $\Leftrightarrow$ 強い意味での定常性)

AR(1) の母平均値, 母分散

$$\mathrm{E}[X_t] = \mathrm{E}[\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}] = \phi_1^t \mathrm{E}[X_0]$$

$$\begin{aligned}
\mathrm{V}[X_t] =& \mathrm{V}[\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}] \\
=& \mathrm{V}[\phi_1^t X_0] + \sum_{k=0}^{t-1} \phi_1^k \mathrm{V}[\epsilon_{t-k}] \\
=& \phi_1^{2t} \mathrm{V}[X_0] + \sum_{k=0}^{t-1} \phi_1^{2k} \sigma^2 \quad \text{WN2,PI} \\
=& \phi_1^{2t} \mathrm{V}[X_0] + \frac{1-\phi_1^{2t}}{1-\phi_1^2} \sigma^2
\end{aligned}$$

$|\phi_1| < 1, t \to +0$ で,
定常 ($\mathrm{E}[X_t] \to \mathrm{E}[X_0]$) の可能性,
$\mathrm{V}[X_t] \to \frac{1}{1-\phi_1^2}\sigma^2$.
以下, 簡単のため $\mathrm{E}[X_0] = \mathrm{E}[X_t] = 0$ と仮定.

独立と共分散

$\mathrm{E}[AB] = 0$ の意味
'偏りがない' とき, つまり $\mathrm{E}[A] = 0$ or $\mathrm{E}[B] = 0$ のとき, 共分散 $\mathrm{Cov}[A, B] = \mathrm{E}[AB]$.

独立ならば共分散ゼロ

$\mathrm{Cov}[A, B] = \mathrm{E}[(A - \mu_A)(B - \mu_B)] = \mathrm{E}[AB] - \mathrm{E}[A]\mathrm{E}[B] \overset{\text{独立}}{=} \mathrm{E}[A]\mathrm{E}[B] - \mathrm{E}[A]\mathrm{E}[B] = 0$

よって, $\mathrm{E}[AB] = 0$ は, 独立の必要条件 (独立よりちょっとだけ弱い条件, '2 次までは独立')

AR(1) の母自己共分散

$$\begin{aligned}\mathrm{Cov}[X_t, X_{t-s}] =& \mathrm{E}[X_t X_{t-s}] \\ =& \mathrm{E}[(\phi_1^t X_0 + \sum_{k=0}^{t-1} \phi_1^k \epsilon_{t-k}) \times (\phi_1^{t-s} X_0 + \sum_{\ell=0}^{t-s-1} \phi_1^\ell \epsilon_{t-s-\ell})]\end{aligned}$$

展開すると 0 にならないのは, $\mathrm{E}[X_0 X_0]$, $\mathrm{E}[\epsilon_k \epsilon_k] = \sigma^2$ のみ.

$$\begin{aligned}\mathrm{Cov}[X_t, X_{t-s}] =& \phi_1^{2t-s} \mathrm{E}[X_0 X_0] + (\phi_1^{2t-2s-2} + \phi_1^{2t-2s-4} + \cdots + \phi_1^0)\phi_1^s \sigma^2 \\ =& \phi_1^{2t-s} \mathrm{E}[X_0 X_0] + \frac{1-(\phi_1^2)^{t-s}}{1-\phi_1^2} \phi_1^s \sigma^2 \\ & \to \begin{cases} \frac{\phi_1^s \sigma^2}{1-\phi_1^2} & (|\phi_1| < 1) \\ \approx t & (|\phi_1| = 1) \\ \approx t\phi_1^{2t} & (|\phi_1| > 1) \end{cases}\end{aligned}$$

1 行前を修正

$|\phi_1| < 1$ のとき, $t \to +\infty$ で定常.
このとき sj 次の母自己相関係数は $r(s) = \frac{\phi_1^s \sigma^2}{1-\phi_1^2} \quad (|\phi_1| < 1)$ つまり等比数列.

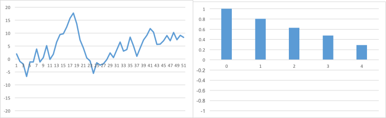
AR(1) モデルの母自己相関係数

k 次の母自己相関係数

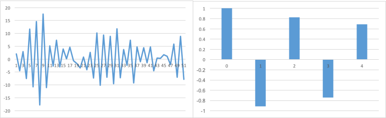
$$r(s) = \frac{\mathrm{Cov}[X_t, X_{t-s}]}{\sqrt{\mathrm{V}[X_t]}\sqrt{\mathrm{V}[X_{t-s}]}} \overset{\text{定常}}{=} \frac{\mathrm{E}[X_t X_{t-s}]}{\mathrm{E}[X_t X_t]} = \phi^k.$$

サンプルパスの t-X_t グラフと, 次回に説明する方法で推定した s-$r(s)$ グラフ (コレログラム)

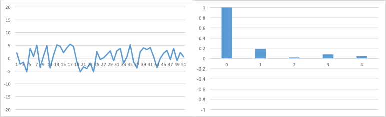
$\phi = 0.9,\ \sigma = 1$



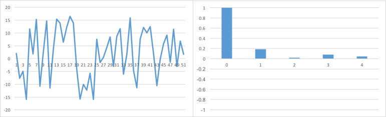
$\phi = -0.9,\ \sigma = 1$



$\phi = 0.2,\ \sigma = 1$



$\phi = 0.2,\ \sigma = 3$



AR(1) モデルとランダムウォーク

Listing 1: AR(1)

```
for(t){/*AR(1)*/
  x=phi*x+getrandom(getuniform());
}
```

AR(1) で $\phi = 1$ で, $\epsilon_t t$ が独立とすると, ランダムウォークで $\mathrm{E}[\epsilon_t] = 0$ で, 壁がないものを実現できる.

```
for(t){/*ランダムウォーク*/
  x=x+getrandom(getuniform());
}
```

モデルの包含関係

$\mathrm{E}[\epsilon_t] = 0$ で壁のないランダムウォーク $\subset$ AR(1) $\subset$ 自己回帰モデル $\subset$ 時系列

L09-Q1

Quiz(自己回帰モデル AR(2) の例)

2 次の自己回帰モデル AR(2) モデルを考える (授業中に使った記号で). 初期条件として, X_0, X_1 は定数 0 に等しいとしてよい $(P(X_0 = 0) = P(X_1 = 0) = 1)$.

1. 確率変数 X_2 を, ϵ_t と ϕ_k で表そう.
2. 母期待値 $\mathrm{E}[X_3^2]$ を σ と ϕ_k で表そう.
3. 母自己共分散 $\mathrm{Cov}[X_3, X_4]$ を σ と ϕ_k で表そう.

ここまで来たよ

8 メトロポリス・ヘイスティングス法とギブスサンプリング

9 時系列の AR,MA,ARMA モデル
- 自己回帰モデル AR(m)
- 自己回帰モデル AR(m)
- k 次の移動平均モデル MA(k) モデル

m 次の自己回帰モデル AR(m))

m 次の自己回帰モデル AR(m)

確率過程 $\{X_t\}_{t=0,1,2,3,\ldots}$. 各 X_t は (連続型) 確率変数.

$$X_t = \sum_{k=1}^{m} \phi_k X_{t-k} + \epsilon_t$$

ただし, ϵ_t は次満たす.

$$\mathrm{E}[\epsilon_t] = 0 \tag{WN1}$$

$$\mathrm{E}[\epsilon_t \epsilon_s] = \sigma^2 \delta_{ts} \tag{WN2}$$

$$\mathrm{E}[X_t \epsilon_s] = 0 \quad (t < s) \tag{PI}$$

Code example

Listing 2: arma

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1,0.5,0.2]          # [phi_0,phi_1,phi_2] 1は必ず含める
ma=[1]                  # 1は必ず含める
AR=ArmaProcess(ar,ma) # 自己回帰移動平均モデルを計算するオブジ
x=AR.generate_sample(nsample=50) # 時間の長さ=50
```

ここまで来たよ

8. メトロポリス・ヘイスティングス法とギブスサンプリング

9. 時系列の AR,MA,ARMA モデル
   - 自己回帰モデル AR(m)
   - 自己回帰モデル AR(m)
   - k 次の移動平均モデル MA(k) モデル

移動平均の定義

定義 (移動平均)

時系列 $\{X_t\}$.
$k = 2\ell + 1$ 次の移動平均 $m_{k,t} = \frac{1}{k}(X_{t-\ell} + \cdots + X_t + \cdots + X_{t+\ell})$.

- X_t にランダム項 ϵ_t が含まれる場合, その効果が小さくなり, 他の部分がよく見える.
- X_t に周期成分がある場合, 次数をその周期に取ると, 周期成分が平均され, 他の部分がよく見える.
- このため, 標本から移動平均を計算することがよく行われる.

移動平均モデルの定義のアイデア

$$X_t = \phi_1 \frac{1}{1} X_{t-1} + \phi_2 \frac{1}{2}(X_{t-1} + X_{t-2}) + \phi_3 \frac{1}{2}(X_{t-1} + \cdots + X_{t-3}) + \cdots + + \phi_k \frac{1}{2}(X_{t-}$$

としては?

k 次の移動平均モデル MA(k)

定義 (k 次の移動平均モデル MA(k))

$t = 1, 2, \ldots$: 時刻
$\{\epsilon_t\}_{t=1,\ldots}$,$\{X_t\}_{t=0,1,\ldots}$: 連続型確率変数の列
$\theta_k \in \mathbb{R}$: パラメタ
$X_t (t \geq 1)$ は次のように定まる.

$$X_t = \beta_0 + \theta_0 \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_k \epsilon_k. \quad (t \geq k)$$

ただし $\theta_0 = 1$. ϵ_t は次を満たす.

$$\mathrm{E}[\epsilon_t] = 0 \quad \text{(WN1)}$$
$$\mathrm{E}[\epsilon_t \epsilon_s] = \sigma^2 \delta_{ts} \quad \text{(WN2)}$$
$$\mathrm{E}[X_t \epsilon_s] = 0 \quad (t < s) \quad \text{(PI)}$$

移動平均モデル MA(k) の母平均値・母自己共分散

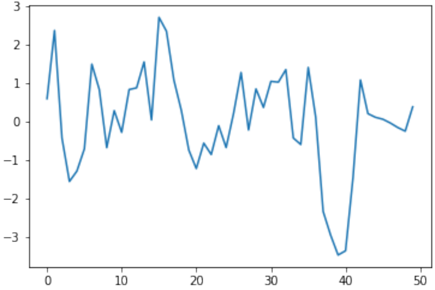
$$\mathrm{E}[X_t] = \beta_0.$$

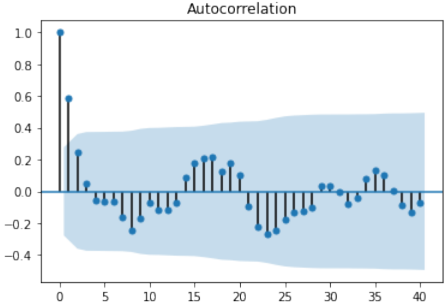
$$\begin{aligned}
\mathrm{V}[X_t] =& \mathrm{V}[\theta_0\epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_k\epsilon_k] \\
=& \theta_0^2\mathrm{V}[\epsilon_t] + \theta_1^2\mathrm{V}[\epsilon_{t-1}] + \cdots + \theta_k^2\mathrm{V}[\epsilon_k] \\
=& (\theta_0^2 + \theta_1^2 + \cdots + \theta_k^2)\sigma^2.
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}[X_t, X_{t-s}] =& \mathrm{E}[(\theta_0\epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_k\epsilon_{t-k}) \\
& \times (\theta_0\epsilon_{t-s} + \theta_1\epsilon_{t-s-1} + \cdots + \theta_k\epsilon_{t-s-k})] \\
=& \cdots \\
=& \begin{cases} 0 & (k < s) \\ (\theta_0\theta_s + \theta_1\theta_{s+1} + \cdots + \theta_{k-s}\theta_k)\sigma^2 & (k \geq s) \end{cases}.
\end{aligned}$$

移動平均モデル MA(k) のサンプルパスとコレログラム

サンプルパスから, 次回に説明する方法から推定した結果.





L09-Q2

Quiz(移動平均モデル MA(1) の例)

1 次の移動平均モデル MA(1) モデルを考える (授業中に使った記号で).

1. 確率変数 X_2 を, ϵ_t, θ_1 で表そう.
2. 母期待値 $\mathrm{E}[X_2^2]$ を σ と θ_1 で表そう.
3. 母自己共分散 $\mathrm{Cov}[X_2, X_3]$ を σ と θ_1 で表そう.
4. 母自己共分散 $\mathrm{Cov}[X_2, X_4]$ を求めよう.

Code example

Listing 3: arma

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1]                  # 1は必ず含める
ma=[1,0.5,0.2]          # [theta_0,theta_1,theta_2] 1は必ず含める
MA=ArmaProcess(ar,ma) # 自己回帰移動平均モデルを計算するオブジ
x=MA.generate_sample(nsample=50) # 時間の長さ=50
```

m,k 次の自己回帰移動平均モデル ARMA(m,k)

漸化式の右辺が, AR(m) と MA(k) の和になっている時系列

m,k 次の自己回帰和分移動平均モデル ARIMA(m,k)

階差が ARMA(m,k) にしたがう時系列