時系列データ, 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　L10 (2022-11-30 Wed)

最終更新: Time-stamp: "2022-12-01 Thu 18:22 JST hig"

今日の目標

- 時系列データの統計量を求められる
- 確率過程の自己回帰モデル AR(m) 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.

L09-Q1
Quiz 解答: 自己回帰モデル AR(1) の例

1. モデルの定義より $X_2 = \phi_1 X_1 + \epsilon_1 = \phi_1(\phi_1 \cdot 0 + \epsilon_1) + \epsilon_2 = \phi_1\epsilon_1 + \epsilon_2$.
2. $\mathrm{E}[(\phi_1\epsilon_1 + \epsilon_2)^2] = \phi_1^2\sigma^2 + 2\cdot\phi_1\cdot 0 + \sigma^2$.
3. $\mathrm{E}[X_1] = \mathrm{E}[X_2] = 0$ より, $\mathrm{Cov}[X_1, X_2] = \mathrm{E}[X_1 X_2] = \phi_1\sigma^2$.

ここまで来たよ

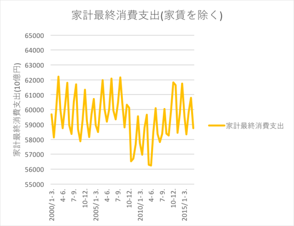
9 時系列の AR,MA,ARMA モデル

10 時系列データ, MA,ARMA モデル
- 時系列解析
- 標本移動平均
- 標本自己共分散・標本自己相関係数
- m 次の自己回帰モデル AR(m)
- k 次の移動平均モデル MA(k)

時系列解析 Time Series Analysis

時系列　時間 t に依存する量の列 $x(0), x(1), x(2), \ldots, x(t), \ldots$.
以前の値が, 今の値に影響. $x(t)$ $t = 0, 1, 2, 3, \ldots$ は独立でない.
例

- 特定の銘柄の毎日の株価のデータ
- 週ごとの売上のデータ
- 1 分おきの気温のデータ
- 1 年ごとの太陽黒点の個数のデータ
- 時刻 t のランダムウォーカーの座標 $X(t)$



時系列解析　時系列を解析する手法群 経済統計学でさかん.
目的 時系列を再現する. $t \leq T$ のデータから $t > T$ を予測する.
標本 (データ) を解析 $\rightarrow$ **時系列モデルを作成** $\rightarrow$ 再現・予測

ランダムウォーク　確率モデル (3Q4)

時系列の 3 要素

現実の時系列は次の 3 つの重ね合わせになっていることが多い

$x(t) = \boxed{\text{トレンド}} + \boxed{\text{周期的変動}} + \boxed{\text{ランダム成分}}$

- トレンド (長期的傾向) 期間を通して時間に比例して増減する傾向. 一過的な増減
- 短い周期の周期的変動 季節, 週, 月, 年

  フーリエ級数解析, フィルタ (パターン情報処理)
- ランダム成分 (ノイズ) 時刻ごとに独立な確率変数

ここまで来たよ

標本移動平均

標本平均 moving average とは, 時系列 $x(t)$ から平滑化 (smoothing) した別の時系列 $y(t)$ を作る手法

定義 ($(2\ell+1)$ 次の移動平均)

$$y_{2\ell+1}(t) = \frac{1}{2\ell+1}\sum_{t'=t-\ell}^{t+\ell} x(t')$$

.

$x(0)$　$x(1)$　$x(2)$　$x(3)$　$x(4)$　$x(5)$　$x(6)$　$x(7)$　$x(8)$　$x(9)$　$x(10)$　x

定義 (2ℓ 次の移動平均)

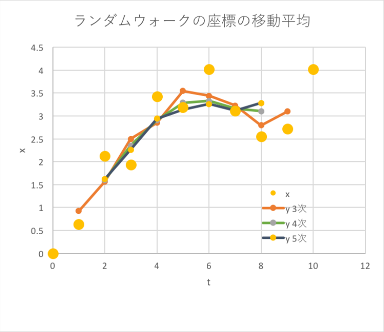
$$y_{2\ell}(t) = \frac{1}{2\ell}\left(\frac{1}{2}\cdot x(t-\ell+1)+x(t-\ell+2)+\cdots+x(t)+\cdots+x(t+\ell-2)+\frac{1}{2}x(t+\ell-1)\right)$$

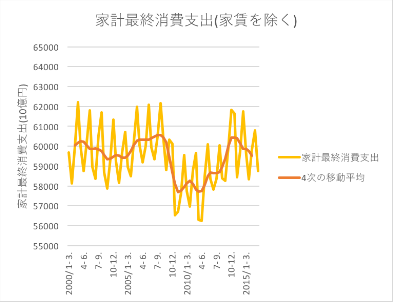
$x(0)$　$x(1)$　$x(2)$　$x(3)$　$x(4)$　$x(5)$　$x(6)$　$x(7)$　$x(8)$　$x(9)$　$x(10)$　x

離散フーリエ級数の第 0 項

移動平均の性質

- 次数が高くなるほど滑らかになる
- 元のデータの「真ん中へん」を通る.
- 時間帯の端までは描けない





移動平均の性質 2

- トレンド (長期的傾向) 期間を通して時間に比例して増減する傾向. 一過的な増減 $\rightarrow$ 移動平均ではっきり見えるようになる
- 短い周期の周期的変動 季節, 週, 月, 年 $\rightarrow$ (周期くらいの) 移動平均で消える. (もとのデータ)−(移動平均) ではっきり見える フーリエ級数解析, フィルタ (パターン情報処理)
- ランダム成分 (ノイズ) 時刻ごとに独立な乱数 $\rightarrow$ 移動平均で小さくなる

ラグ k の後方移動平均 $y(t) = \frac{1}{k}\sum_{t'=t-k+1}^{t} x(t')$
ラグ k の前方移動平均 $y(t) = \frac{1}{k}\sum_{t'=t}^{t+k-1} x(t')$

L10-Q1

Quiz(移動平均)

次の時系列データから, 3,4 次の移動平均を求めよう.

t	0	1	2	3	4	5	6	7	8	9
x	1.8	-1.6	2.6	-1.2	3.0	-1.2	3.4	-0.8	3.8	0.0
y_3										
y_4										

ここまで来たよ

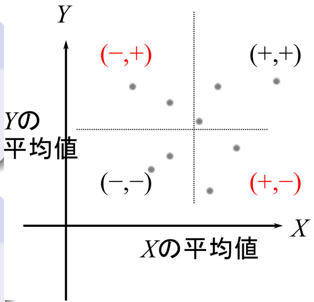
復習: 標本共分散

定義 (母共分散 (covariance))

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])]$$

定義 (標本共分散 (covariance))

$$C_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})\times(y_i-\overline{y})$$



$(+,-) = (x_i - \overline{x}$の符号$, y_i - \overline{y}$の符号$)$.

標本自己共分散, 標本自己相関係数

定義 (k 次の標本自己共分散, 標本自己相関係数)

時間 t を, ラグ k だけずらした $x(t), x(t-k)$ を 2 変量データだと思って, 標本共分散, 標本相関係数を考えたもの

$$C(k) = \frac{1}{T-k} \sum_{t=k+1}^{T} (x(t) - \overline{x})(x(t-k) - \overline{x})$$

$$\text{ただし標本平均値}\overline{x} = \frac{1}{T-k} \sum_{t=k+1}^{T} x(t)$$

定常な時系列データに対する標本ナントカ

‘時系列分析’ では, 定常過程について, 1 個の時系列データを, 一定の長さに分割して複数個のデータからなる標本のように扱うことが行われる.

例: 標本平均値 $\overline{X}_t$, 不変標本分散 $S^2_{X_t}$.
例: 横:t 縦: 標本内のデータ番号. 標本自己相関係数を求めるとき

	A	B	C	D	E	F	G	H	I	J
1	n ＼ t	0	1	2	3	4	5	6	7	8
2	1	2	-0.2522	-1.4293	0.41339	[illegible]	[illegible]	0.05638	1.16461	0.77765
3	2	2	0.40156	-0.1803	1.02945	[illegible]	[illegible]	1.1405	0.1365	0.27561
4	3	2	1.74901	0.20238	-0.7695	[illegible]	[illegible]	0.66943	0.10917	1.30642
5	4	2	0.45349	-0.3762	-0.6185	[illegible]	[illegible]	-1.4453	0.09464	1.55283
6	5	2	1.01492	0.94089	-0.5959	[illegible]	[illegible]	0.5173	0.11447	1.38334
7	6	2	1.78265	1.70528	2.07552	[illegible]	[illegible]	0.10902	-0.7164	-1.1645
8	7	2	0.69062	1.27761	1.15377	[illegible]	[illegible]	0.24885	-1.323	-1.1454
9	8	2	0.74554	0.27013	1.05844	[illegible]	[illegible]	-0.0232	1.09015	1.91198
10	9	2	1.21618	-0.2882	0.81639	[illegible]	[illegible]	-1.1622	-1.5981	-2.3211
11	10	2	1.20421	-0.487	-0.2274	[illegible]	[illegible]	1.46621	0.13736	0.13426

本当はこういう標本が欲しい

	A	B	C	D	E	F	G	H	I	J
1	n ＼ t	0	1	2	3	4	5	6	7	8
2	1	2	-0.2522	-1.4293	0.41339	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
3	2	2	0.40156	-0.1803	1.02945	1.31571	-0.3059	1.1405	0.1365	0.27561
4	3	2	1.74901	0.20238	-0.7695	-1.3943	-0.0067	0.66943	0.10917	1.30642
5	4	2	0.45349	-0.3762	-0.6185	-1.7157	-0.461	-1.4453	0.09464	1.55283
6	5	2	1.01492	0.94089	-0.5959	-0.8685	0.30052	0.5173	0.11447	1.38334
7	6	2	1.78265	1.70528	2.07552	0.06083	0.429	0.10902	-0.7164	-1.1645
8	7	2	0.69062	1.27761	1.15377	1.44955	-0.4663	0.24885	-1.323	-1.1454
9	8	2	0.74554	0.27013	1.05844	-0.1876	0.06592	-0.0232	1.09015	1.91198
10	9	2	1.21618	-0.2882	0.81639	1.84743	0.18484	-1.1622	-1.5981	-2.3211
11	10	2	1.20421	-0.487	-0.2274	0.34252	0.62665	1.46621	0.13736	0.13426

定常ならこれでもいいじゃん

	A	B	C	D	E	F	G	H	I	J
1	n ＼ t	0	1	2	3	4	5	6	7	8
2			2	-0.2522	-1.4293	0.41339	[illegible]	[illegible]	[illegible]	[illegible]
3	1	2	-0.2522	-1.4293	0.41339	0.93494	[illegible]	[illegible]	[illegible]	[illegible]
4	2	2	0.40156	-0.1803	1.02945	1.31571	-0.3059	1.1405	0.1365	0.27561
5	3	2	1.74901	0.20238	-0.7695	-1.3943	-0.0067	0.66943	0.10917	1.30642
6	4	2	0.45349	-0.3762	-0.6185	-1.7157	-0.461	-1.4453	0.09464	1.55283
7	5	2	1.01492	0.94089	-0.5959	-0.8685	0.30052	0.5173	0.11447	1.38334
8	6	2	1.78265	1.70528	2.07552	0.06083	0.429	0.10902	-0.7164	-1.1645
9	7	2	0.69062	1.27761	1.15377	1.44955	-0.4663	0.24885	-1.323	-1.1454
10	8	2	0.74554	0.27013	1.05844	-0.1876	0.06592	-0.0232	1.09015	1.91198
11	9	2	1.21618	-0.2882	0.81639	1.84743	0.18484	-1.1622	-1.5981	-2.3211

Excel 的にはこうやると楽

$k = 1$ の例

x	y
$x(1)$	—
$x(2)$	$x(2-1)$
$x(3)$	$x(3-1)$
$\vdots$	$\vdots$
$x(T-1)$	$x(T-1-1)$
$x(T)$	$x(T-1)$
—	$x(T)$

k の例

x	y
$x(1)$	—
$\vdots$	
$x(k+1)$	$x(1)$
$\vdots$	$\vdots$
$x(T)$	$x(T-k)$
$\vdots$	
—	$x(T)$

定義 (k 次の標本自己相関係数 autocorrelation)

$$r(k) = \frac{\text{自己共分散}}{\sqrt{\text{分散}}\sqrt{\text{分散}}} = \frac{C(k)}{C(0)}$$

$C(0)$ は $x(t)$ をサイズ T の標本と思ったときの分散.

定義 (コレログラム correlogram)

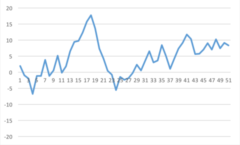
横軸 ラグ k, 縦軸 k 次の自己相関係数 $r(k)$ のグラフ.

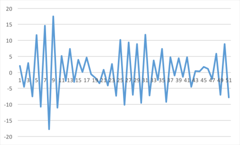
- 多くの定常モデル (自己回帰モデルなど) では, k が大きいほど (遠い時刻ほど), 標本自己相関係数 $r(k)$ の絶対値は小さくなる.

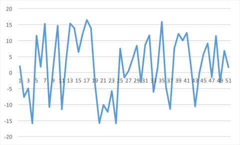
L10-Q2

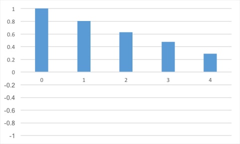
Quiz(コレログラム)

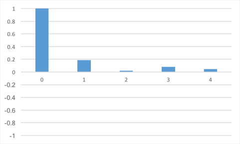
次の時系列データと，コレログラムの間の対応をつけよう．

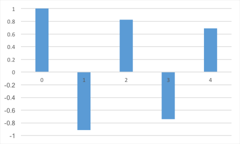
1. 

2. 

3. 

a. 

b. 

c. 

離散フーリエ変換

データ $\{X_t\}_{t=0,1,\ldots,L-1}$ の周期性を知るには, 離散フーリエ変換も使える.
離散フーリエ変換

$$\tilde{X}_s = \sum_{t=0}^{L-1} e^{-\frac{2\pi st}{L}} X_t$$

で, $\frac{1}{L} \cdot |\tilde{X}_s|$ は, 周期 L/s の成分の寄与の大きさ.

フーリエ解析

ここまで来たよ

9 時系列の AR,MA,ARMA モデル

10 時系列データ, MA,ARMA モデル
- 時系列解析
- 標本移動平均
- 標本自己共分散・標本自己相関係数
- m 次の自己回帰モデル AR(m)
- k 次の移動平均モデル MA(k)

m 次の自己回帰モデル AR(m)

定義 (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.

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

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

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

L10-Q3

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.

Code example

Listing 1: arma

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1,-0.5,-0.2]        # [phi_0=1,-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

移動平均の定義

定義 (移動平均)

時系列 $\{X_t\}$.
k 次の後方移動平均 $Y_{t,k} = \frac{1}{k}(X_{t-k+1} + \cdots + X_{t-1} + X_t)$.

移動平均モデルの定義のアイデア

$$X_t = \phi_1 Y_{t-1,1} + \phi_2 Y_{t-1,2} + \cdots + \phi_k Y_{t-1,k} + \epsilon_t$$

としては?
漸化式を解いていくと, けっきょくすべて $\epsilon_{t'}$ $t-k \leq t' \leq t$ に帰着する.

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))

$t = 0, 1, 2, \ldots$: 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.

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

ただし $\theta_0 = 1$. ϵ_t は次を満たす.

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

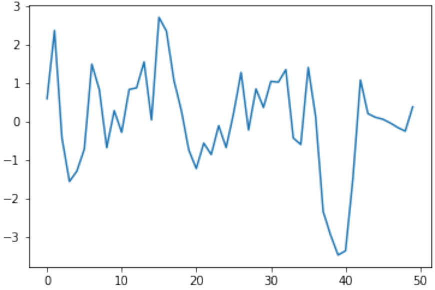
移動平均モデル MA(k) の母平均値・母自己共分散

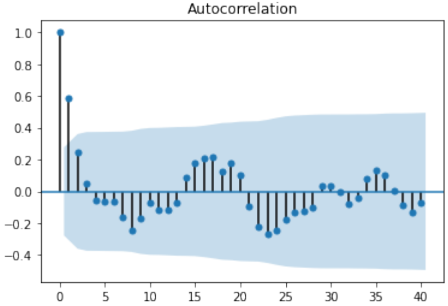
$$\mathrm{E}[X_t] = \beta_0.$$

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

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

移動平均モデル MA(k) のサンプルパスとコレログラム





L10-Q4

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).
$\beta_0 = 0$ とする.

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 2: arma

```
from statsmodels.tsa.arima_process import ArmaProcess
ar=[1,-phi1,-phi2]             # 1は必ず含める マイナスをつける
ma=[1,theta1,theta2]           # [theta_0,theta_1,theta_2] 1は必ず
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
```

ar, ma は, 漸化式の L のラグ多項式 の係数のリスト.

定義 (ラグ演算子)

ラグ演算子 L を $LX_t = X_{t-1}$ で定める.

$$
\begin{aligned}
X_t =& \phi_1 X_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \\
(1 - \phi_1 L^1 - \phi_2 L^2) X_t =& (1 + \theta_1 L^1 + \theta_2 L^2) \epsilon_t
\end{aligned}
$$

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