時系列解析と自己回帰モデル

樋口さぶろお

龍谷大学理工学部数理情報学科

計算科学☆実習 B L11(2017-07-03 Mon)

最終更新: Time-stamp: "2017-07-03 Mon 11:04 JST hig"

今日の目標

- 移動平均, 標本自己共分散, 標本自己相関係数の意味が説明でき, 計算できる.
- 自己回帰モデルの定義が説明でき, 確率シミュレーションに使える.

http://hig3.net

L10-Q1

Quiz 解答:ラグランジュ表現とオイラー表現

1. 6 羽なのでサイズは 6.
   各要素は, `x[]={1,1,3,3,3,8};` (順序はこうである必要はない. 自由にペンギン番号をつけてよい)
2. 座標が $x = 0, 1, 2, \ldots, 9$ の計 10 か所なので, サイズは 10.
   各要素は `u[]={0,2,0,3,0,0,0,0,1,0};` (順序はこうである必要がある)

L10-Q3

Quiz 解答:ラグランジュ表現

```
int sum=0,count=0;
for(k=0;k<SAMPLESIZE;k++){
  sum+=x[k];
  if( x[k]<=5) count++;
}
ex=(double)sum/SAMPLESIZE;
px=(double)count/SAMPLESIZE;
```

L10-Q4

Quiz 解答:オイラー表現

```
double ex=0.0, px=0.0;
for(x=0;x<XMAX;x++){
  ex+=p[x]*x;
  if(x<=5) px+=p[x];
}
```

L10-Q5

Quiz(確率シミュレーションと中心極限定理)

B 湖の毎日の水位の変化 R は, 毎日独立に, -1 cm 以上 2cm の範囲でランダムに定まり, どの値も同様に確からしい. 0 日に水位は 100cm だった.

1. 30 日の水位はどんな分布?
2. 30 日の水位が 120cm 以上 125cm 未満である確率を求めよう.

ただし, 標準正規分布の累積分布関数 $\Phi(z) = \frac{1}{2\pi} \int_{-\infty}^{z} \mathrm{e}^{-u^2/2} \mathrm{d}u$ を使って答えてよい. 計算機でシミュレーションして答えてもよい.

t 日の水位を $X(t)$ とすると,

$$X(t+1) = X(t) + R(t+1), \quad X(0) = 100.$$

ただし, $R(t)$ は確率変数で, 確率密度関数

$$f(r) = \begin{cases} 1/3 & (-1 \le r < 2) \\ 0 & (\text{他}) \end{cases}$$

にしたがう.

$\vdots$

ここまで来たよ

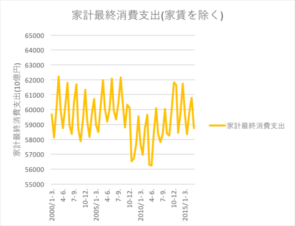
10 オイラー表現とラグランジュ表現

11 時系列解析と自己回帰モデル
- 標本の時系列解析:移動平均と自己相関係数
- 確率過程の時系列モデル

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 $\rightarrow$ $\boxed{}$ $\rightarrow$ 再現・予測
ランダムウォークで言ったらデータから R の分布 $f_R(t)$ を知って, 確率シミュレーションやマルコフ連鎖…で推定すること.

移動平均 Moving Average

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

$(2\ell+1)$ 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

2ℓ 次の移動平均

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

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

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

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

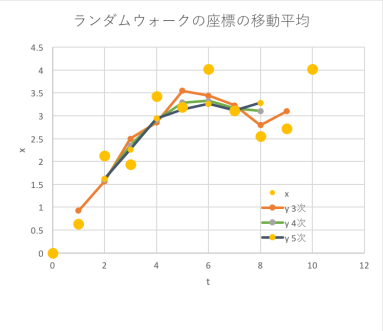
- トレンド (長期的傾向) 期間を通して時間に比例して増減する傾向. 一過的な増減 → 移動平均ではっきり見えるようになる
- 短い周期の周期的変動 季節, 週, 月, 年 → (周期くらいの) 移動平均で消える. (もとのデータ)−(移動平均) ではっきり見える

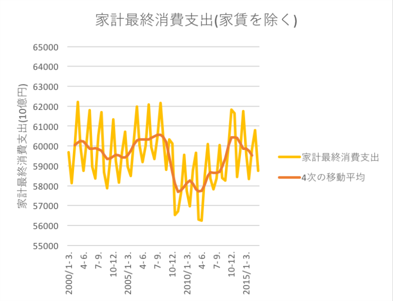
フーリエ級数解析, フィルタ (パターン情報処理)

- ランダム成分 (ノイズ) 時刻ごとに独立な乱数 → 移動平均で消える. → 次の自己相関係数で気にする

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





L11-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.

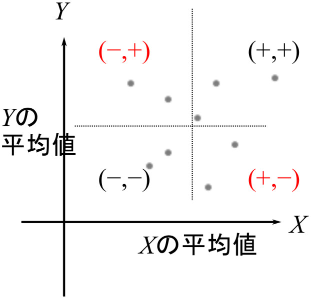
t	0	1	2	3	4	5	6	7	8	9
x	1.8	-1.6	2.6	-1.2	3.0	-1.2	3.4	-0.8	3.8	0.0

y_3										
y_4										

復習:標本共分散と標本相関係数

標本共分散 (covariance)

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



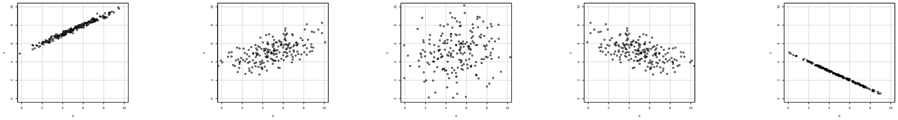
$$(+, -) = (x_i - \overline{x}\text{の符号}, y_i - \overline{y}\text{の符号}).$$

塚田確率統計 §1.8, §3.6

相関係数

$$\text{標本相関係数}\ r = \frac{C_{xy}}{s_x s_y}$$

s_x, s_y :標本標準偏差



強い正の相関 $r = 0.99$ | 弱い正の相関 $r = 0.55$ | 無相関 $r = 0$ | 弱い負の相関 $r = -0.55$ | 強い負の相関 $r = -0.99$

相関

'正の相関': x が大きい $\Leftrightarrow$ y が大きい

'負の相関': x が大きい $\Leftrightarrow$ y が小さい

強い/弱い: 傾向がはっきりしている/していない

2 変量データ　塚田確率統計 §1.8, §3.6

標本自己共分散, 標本自己相関係数

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

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

kj 次の標本自己共分散 autocovariance

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

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

kj 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.

コレログラム correlogram

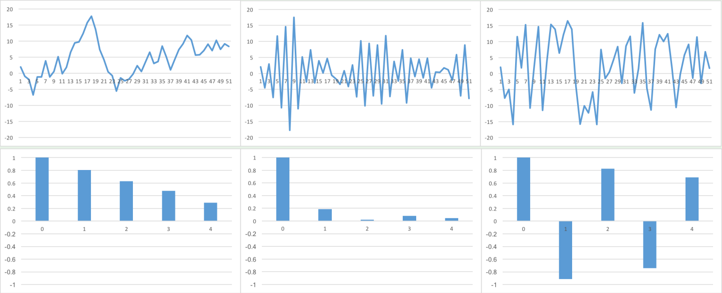
横軸 ラグ k, 縦軸 k 次の自己相関係数の棒グラフのこと.

- 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.
- ランダムウォークのとき, $r(k) =$ 一定.

L11-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.



ここまで来たよ

m 次の自己回帰モデル=AR モデル Autoregression

m 次の自己回帰モデル AR(m)

$X(t)$: 連続型確率変数, $t = 0, 1, 2, 3, \ldots$

$$X(t) = \sum_{k=1}^{m} a_k X(t-k) + R(t)$$

ただし, $R(t)$ はすべて同じ分布で次を満たす.

$$\begin{aligned}
\mathrm{E}[R(t)] &= 0, \\
\mathrm{E}[R(t)X(s)] &= 0 \quad (t > s), \\
\mathrm{E}[R(t)R(s)] &= \sigma^2 \times \delta_{t,s} = \sigma^2 \times \begin{cases} 1 & (t = s) \\ 0 & (\text{他}) \end{cases}
\end{aligned}$$

これを満たす $R(t)$ をホワイトノイズ, 白色雑音 という.

AR(1) モデルとランダムウォーク

$\mathrm{E}[R] = 0$ なランダムウォークは, AR(1) モデルとみなせる.
$a_1 = 1$. $\mathrm{E}[R(t)] = 0$, $\mathrm{V}[R(t)] = \sigma^2$.

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

AR(1) $a_1 = \phi$ (一般の値). $\mathrm{E}[R(t)] = 0$, $\mathrm{V}[R(t)] = \sigma^2$.

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

L11-Q3

Quiz(AR(1) モデルの分布)

AR(1) モデル

$$X(t+1) = \phi \times X(t) + R(t+1), \quad R(t)$$

を考える. $\phi = 1$ ととったときはランダムウォークになる.

1. $X(2)$ を $X(0), R(1), R(2)$ で書き表そう.
2. $X(0) = a$, すなわち, $P(X(0) = a) = 1$ とする. また, $R(t) \sim \mathrm{N}(0, \sigma^2)$ とする (AR(1) モデルの条件を満たすことをチェック). このとき, $X(2)$ のしたがう分布を求めよう.
3. (2) と同じ条件下で, $X(t)$ $(t \geq 1)$ のしたがう分布を求めよう.

定常過程

定常過程

確率過程 $X(t)$ で

- $\mathrm{E}[X(t)]$ が t によらない
- $\mathrm{E}[X(t)X(s)]$ が差 $t-s$ だけにより, t によらない

とき, 定常過程という.

要するに

ある時間範囲, たとえば $t \to \infty$ のみで定常過程であることはある.

- がありと定常過程ではない
- があると定常過程ではない
- ランダムウォークは定常過程

母自己共分散, 母自己相関係数

母自己共分散, 母自己相関係数

$\mu = \mathrm{E}[X(t)]$.
k 次の母自己共分散 $C(k) = \mathrm{E}[(X(t) - \mu)(X(t+k) - \mu)]$
$k = 0$ 次の母自己共分散 $C(0) = \mathrm{E}[(X(t) - \mu)^2]$=ふつうの分散.
k 次の母自己相関係数 $r(k) = \frac{C(k)}{C(0)}$.

定常な確率過程に対する母ナントカと標本ナントカ

定常過程については, 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ってよい.

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

AR(1) モデルの自己共分散, 自己相関係数

$$
\begin{aligned}
X(t) =& \phi X(t-1) + R(t) \\
=& \phi(\phi X(t-2) + R(t-1)) + R(t) \\
=& \cdots = \phi^t X(0) + \phi^{t-1} R(1) + \cdots \phi R(t-1) + R(t). \\
\mathrm{E}[X(t)] =& \phi^t \mathrm{E}[X(0)].
\end{aligned}
$$

以下 $\mathrm{E}[X(t)] = 0$ と仮定.

$$
\begin{aligned}
\mathrm{E}[X(t)X(t+k)] =& \mathrm{E}[(\phi^t X(0) + \phi^{t-1} R(1) + \cdots \phi R(t-1) + R(t)) \\
& \times(\phi^{t+k} X(0) + \phi^{t+k-1} R(1) + \cdots + \phi^{k+1} R(t-1) + \phi^k R(t) \\
& \quad + \cdots + R(t+k))] \\
=& \phi^{2t+k} \mathrm{E}[X(0)X(0)] + (\phi^{2t+k-2} + \phi^{2t+k-4} + \cdots \phi^k)\sigma^2 \\
\mathrm{E}[X(t)X(t)] =& \phi^{2t} \mathrm{E}[X(0)X(0)] + (\phi^{2t-2} + \phi^{2t-4} + \cdots + \phi^0)\sigma^2.
\end{aligned}
$$

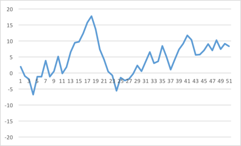
$$
\frac{r(k) = \mathrm{E}[X(t)X(t+k)]}{\mathrm{E}[X(t)X(t)]} = \phi^k.
$$

AR(1) モデルの母自己相関係数

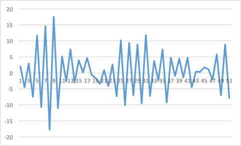
実は, 定常な AR(1) モデルでは $r(k) = a_1^{|k|} = \phi^{|k|}$.
$X(t)$ のサンプルパスや コレログラムから, $a_1 = \phi$, σ^2 は予想できる.

これらが見た目で区別できるようになりたい.

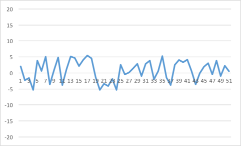
$\phi = 0.9$, $\sigma = 1$



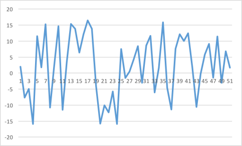
$\phi = -0.9$, $\sigma = 1$



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



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



お知らせ

- 月昼 樋口オフィスアワー (1-502)
- チューター/Math ラウンジ 月火水木昼 1-614

→ Learn Math Moodle → 計算科学☆演習 B
→ 2017-07-12 までの参加予定

http://hig3.net

計算科学の今後の予定

- 2017-07-05 水 3 説明+プレゼンテーション準備
- 2017-07-10 月 4 説明+プレゼンテーション準備 (3-B105)
- 2017-07-12 水 3 プレゼンテーション 1 15 ピーナッツ (小教室)
- 2017-07-17 月 4 説明+プレゼンテーション準備 (3-B105)
- 2017-07-19 水 3 振り返り+プレゼンテーション準備 (3-B105)
- 2017-07-24 月 集中補講日 計算科学なし
- 2017-07-26 水 3 プレゼンテーション 2 15 ピーナッツ (小教室)
- 2017-07-31 月 4 ファイナルトライアル (筆記) 25 ピーナッツ
- 2017-08-02 水 3 演習の時間はファイナルトライアルなし