時系列解析

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 L12(2016-07-11 Mon)

最終更新: Time-stamp: "2016-07-11 Mon 17:25 JST hig"

今日の目標

- 移動平均, 標本自己共分散, 標本自己相関係数が Excel で計算できる
- 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

L11-Q1
Quiz 解答:確率シミュレーションと中心極限定理

1. $\mathrm{E}[R] = \frac{1}{2}, \mathrm{V}[R] = \frac{9}{12} = \frac{3}{4}$. $R(t)$ は独立なので,
$\mathrm{E}[X(30)] = \mathrm{E}[X(0)] + 30 \cdot \frac{1}{2} = 115, \mathrm{V}[X(30)] = \frac{3}{4} \cdot 30 = \frac{90}{4}$.
2. $x = X(30)$ のしたがう確率分布の確率密度関数は,$T = 30$ が十分に大きいと考えると, 中心極限定理より

$$f(x; 115, \tfrac{90}{4}) = \frac{1}{\sqrt{2\pi \cdot \frac{90}{4}}} \mathrm{e}^{-\frac{(x-115)^2}{2\cdot(90/4)}}.$$

よって, 求める確率は, 変数変換 $z = \frac{x-115}{\sqrt{\frac{90}{4}}}$ より,

$$P = \int_{120}^{125} f(x; 115, \tfrac{90}{4})\, \mathrm{d}x = \int_{5/\sqrt{90/4}}^{10/\sqrt{90/4}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}\, \mathrm{d}z.$$

正規分布表より,

$$P = Q(\frac{10}{sqrt90/4}) - Q(\frac{5}{\sqrt{90/4}}) = Q(1.05) - Q(2.11) = 0.1469 - 0.0174$$

L11-Q2

Quiz 解答:確率シミュレーションと中心極限定理

1. $s = r - 4$ と変数変換すると, 上限が $s = 0$ となってだいぶ楽.
$\mathrm{E}[R] = \frac{2}{3}$, $\mathrm{E}[R^2] = \frac{11}{6}$, $\mathrm{V}[R] = \frac{25}{18}$.
よって, $\mathrm{E}[X(30)] = 100 + \frac{2}{3} \times 30 = 120$,
$\mathrm{V}[X(30)] = \frac{25}{18} \times 30 = (\frac{5}{3}\sqrt{15})^2$.

2

$$P=\int_{120}^{130}\frac{1}{\sqrt{2\pi(\frac{5}{3}\sqrt{15})^2}}\mathrm{e}^{-\frac{(x-120)^2}{2(\frac{5}{3}\sqrt{15})^2}}\,\mathrm{d}x$$

$$=\int_{0}^{10/(\frac{5}{3}\sqrt{15})}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{z^2}{2}}\,\mathrm{d}z=Q(0)-Q(10/(\tfrac{5}{3}\sqrt{15})).$$

ここまで来たよ

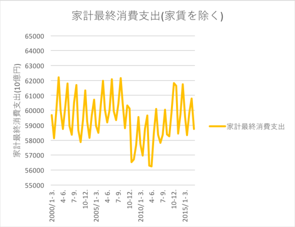
3 連続座標ランダムウォークと中心極限定理

4 時系列解析
- 標本の時系列解析
- 標本の時系列解析:移動平均
- 標本の時系列解析: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 $\rightarrow$ $\rightarrow$ 再現・予測

ここまで来たよ

3 連続座標ランダムウォークと中心極限定理

4 時系列解析
- 標本の時系列解析
- 標本の時系列解析:移動平均
- 標本の時系列解析:自己相関係数
- 確率過程の時系列モデル

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

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

.

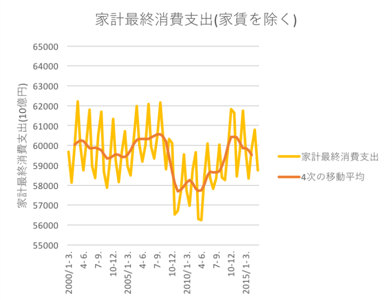
2ℓ 次の移動平均

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

移動平均の利用目的

- ノイズを相殺して真の傾向を見やすく
- 周期的な変動 (例. 季節変動) を消去して真の傾向を見やすく

フーリエ級数解析，フィルタ (パターン情報処理)



- 次数が高くなるほど滑らかになる
- 移動平均は元のデータの真ん中へんを通る.

実習 (移動平均)

与えられたデータの移動平均を求めてグラフを描いて提出.

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

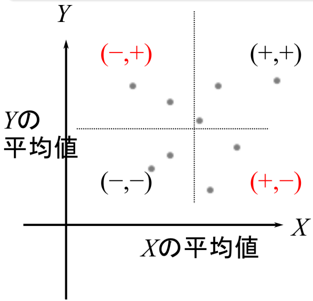
- トレンド 期間を通して時間に比例して増減する傾向. 一過的な増減
  $\rightarrow$ 移動平均ではっきり見えるようになる
- 周期的な変動 季節, 週, 月, 年
  $\rightarrow$ 移動平均で消す. (もとのデータ)−(移動平均) ではっきり見える
- ランダム成分
  $\rightarrow$ 自己相関係数で気にする

ここまで来たよ

復習:標本共分散と標本相関係数

標本共分散 (covariance)

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



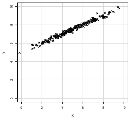
$$(+, -) = (x_i - \overline{x}\text{の符号}, y_i - \overline{y}\text{の符号}).$$

確率統計☆演習 I(2015)L04

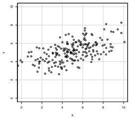
相関係数

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

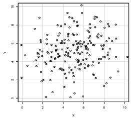
s_x, s_y :標本標準偏差



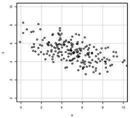
強い正の相関
$r = 0.99$



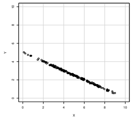
弱い正の相関
$r = 0.55$



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 確率統計☆演習 I(2015)L04

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 $y(t), y(t-k)$ 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
$y(1)$	—
$y(2)$	$y(2-k)$
$y(3)$	$y(3-k)$
$\vdots$	$\vdots$
$y(T-1)$	$y(T-2)$
$y(T)$	$y(T-1)$
—	$y(T)$

標本自己共分散

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

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

標本自己相関係数

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

$C(0)$ は $y(t)$ をサイズ T の標本と思ったときの分散.

コレログラム

横軸 ラグ k, 縦軸 k 次の自己相関係数の棒グラフのこと.

- 多くのモデル (自己回帰モデルなど) では, k が大きいほど $r(k)$ の絶対値は小さくなる.

実習 (標本自己相関係数とコレログラム)

標本自己相関係数を求めてコレログラムを描いて提出. ずらして横の列にコピー. 分析ツール $\to$ 相関係数, 共分散行列

こういう分析に意味があるのは, y が定常のときに限る. $y(t)$ $(t = 1, 2, \ldots, T)$ をサイズ T の標本と考えている (あとで).

Excel での行と列の入れ替え

ペーストするときに「行と列を入れ替える」をチェック

分析ツールでは, 行・列どちらに標本のデータが繰りかえされているかを選べる

ここまで来たよ

3 連続座標ランダムウォークと中心極限定理

4 時系列解析
- 標本の時系列解析
- 標本の時系列解析:移動平均
- 標本の時系列解析:自己相関係数
- 確率過程の時系列モデル

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)

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

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

ただし, R は独立同分布で次を満たす.

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

これを満たす $R(t)$ をホワイトノイズ, 白色雑音 という.

定常過程

定常過程

$\mathrm{E}[Y(t)]$, $\mathrm{E}[Y(t)Y(s)]$ が差 $t-s$ だけにより, t によらないとき,
定常過程という

- $\boxed{}$, があると定常過程ではない
- (ランダム性に起因しない)$\boxed{}$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$\boxed{}$

AR(m) モデルが定常かどうかは a_k による.

AR(1) モデルとランダムウォーク

ランダムウォーク $a_1 = 1$. $\mathrm{E}[R(t)] = 0$, $\mathrm{V}[R(t)] = \sigma^2$.
$\mathrm{E}[R(t)] = \mu \neq 0$ なら, ホワイトノイズ $(R(t) - \mu)$ と トレンド μ に分解.

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

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

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

定常過程に対する標本の考え方

連続した 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．

母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

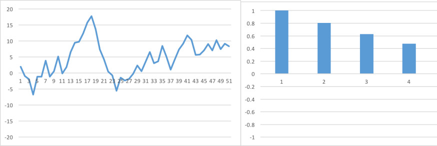
Excel 的にはこうやると楽

AR(1) モデルの母自己相関係数

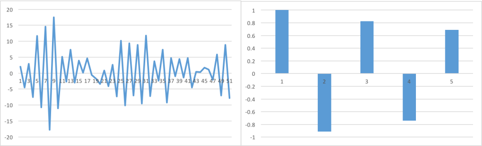
実は, 定常な AR(1) モデルでは $r(k) = \phi^{|k|}$.
$Y(t)$ のサンプルパスと コレログラムから, a_1, σ^2 は予想できる.

これらが見た目で区別できるようになりたい.

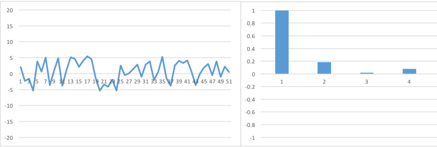
$\phi = 0.9,\ \sigma = 1$



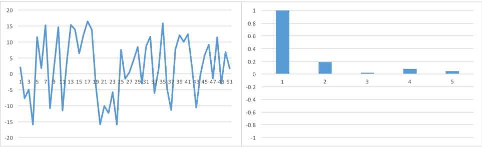
$\phi = -0.9,\ \sigma = 1$



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



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



夏のプチテスト (プログラミング)

- 2016-07-27 水 3 夏のプチテスト (プログラミング)
- 14 ピーナッツ. (旧カリキュラムの人は演習の 28 ピーナッツ/100)
- 春, 初夏のプチテストと同様の非参照プログラミングのテスト. チームでなく個人別.
- 出題計画 (2016-07-20 水に確定します). デバッガーはプログラムの完成に役立ちますが, debugger1, 操作方法など, デバッガーの使用が必須な問題は出題しません.
  - 連続型確率変数の乱数生成, 連続型確率変数の標本のヒストグラム作成 (cont15,inverse01)
  - 連続座標のランダムウォークの確率シミュレーション (contrwsim01 の一部分)
  - 未定

お知らせ

- 全学授業アンケート manaba から.
- もっと大きい点数の, だけどもっと大変な任意プレゼンテーション準備中
- 2016-07-29 金に実習の補講が通知されますが, これは, 夏のプチテスト (プログラミング) の日が台風などで全学休講になった場合の予備日で, 台風が来ないかぎりは実施しません.
- 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

`https://manaba.ryukoku.ac.jp`
マイページの下の方に manaba 出席カード提出