ランダムウォーク, 自己回帰モデル, 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　L08(2018-06-12 Tue)

最終更新: Time-stamp: "2018-06-12 Tue 19:22 JST hig"

今日の目標

- ランダムウォークの座標の母平均値, 母分散が求められる
- 自己回帰モデルの座標の母平均値, 母分散が厳密に求められる
- 自己回帰モデルの自己相関係数が求められる

http://hig3.net

L07-Q1

Quiz 解答:大きな転置推移確率行列をかける関数

ソースコード 1: 大きな転置推移確率行列をかける関数

```
int multiply_trans(double q[], double p[],int m){
  int x;
  q[0]=          7.0/10*p[0]+2.0/10*p[0+1];
  for(x=1;x<m-1;x++){
    q[x]=3.0/10*p[x-1]+5.0/10*p[x]+2.0/10*p[x];
  }
  q[m-1]=3.0/10*p[m-2]+8.0/10*p[m-1];
  return 0;
}
```

L07-Q2

Quiz 解答:一様分布

1. $\mathrm{E}[1]=1$ より, $C=\frac{1}{b-a}$.
2. $\mathrm{E}[X]=\frac{a+b}{2}$.
3. $\sqrt{\mathrm{V}[X]}=\frac{b-a}{\sqrt{12}}\simeq\frac{b-a}{3.5}$.

ここまで来たよ

7 連続型確率変数の擬似乱数

8 ランダムウォーク, 自己回帰モデル, 時系列解析
- (連続型) ランダムウォークの座標の母平均値と母分散
- 自己回帰モデル

(連続型) ランダムウォーク I

(連続型) ランダムウォークの定義

ランダムウォークの座標 $X(t)$: 次で決まる確率変数. $t = a, a+1, \dots$
$R(t)$: 独立同分布に従う (連続型) 確率変数. $t = a+1, a+2, \dots$

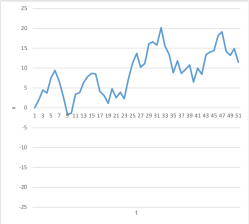
$$X(t) = X(t-1) + R(t), \qquad \text{初期条件 } P(X(a) = b) = 1$$

日本語で言うと,

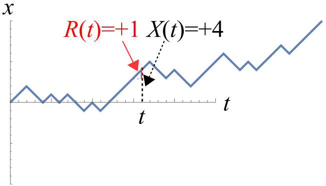
- x 軸上を移動するランダムウォーカーを考える
- ウォーカーは, 時刻 $t = a$ に, $x = b$ から出発する (確率が 1 である)
- ウォーカーは各時刻に, 確率変数 $R(t)$ だけ移動する

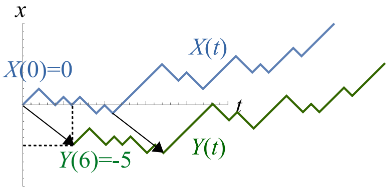
連続型 $R(t)$ の例 $R(t) \sim \mathrm{U}(c, d)$,
$\mathrm{E}[R(t)] = \mu = \frac{c+d}{2}$, $\mathrm{V}[R(t)] = \sigma^2 = \frac{(d-c)^2}{12}$

$X(T)$：時刻 T のランダムウォーカーの座標 (を確率変数とみたもの)
$(X(0), X(1), X(2), \ldots, X(T))$：パス (path) (を確率変数とみたもの)
連続座標ランダムウォーク



離散座標ランダムウォーク, 初期条件 $X(0) = 0$　初期条件 $X(6) = -5$





ランダムウォークの座標の母平均値と母分散

確率統計☆演習 I(2017)L06

$$
\begin{aligned}
X(t) =& X(t-1) + R(t) \\
=& (X(t-2) + R(t-1)) + R(t) \\
=& \cdots = X(a) + R(a+1) + \cdots + R(t)
\end{aligned}
$$

連続/離散型ランダムウォークの母平均値/母分散

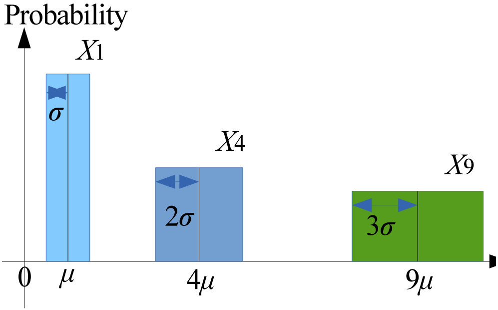
$\mathrm{E}[R(t)] = \mu, \mathrm{V}[R(t)] = \sigma^2$ のとき,

$$
\begin{aligned}
\mathrm{E}[X(T)] =& \mathrm{E}[X(a) + R(a+1) + \cdots + R(T)] = b + (T-a)\mu \quad (\mathrm{E}[]\text{ の性質}), \\
\mathrm{V}[X(T)] =& \mathrm{V}[X(a) + R(a+1) + \cdots + R(T)] \\
=& \mathrm{V}[R(a+1) + \cdots + R(T)] \quad (X(a) = b \text{ は定数}) \\
=& (T-a)\sigma^2 \quad (R(t) \text{ は互いに独立})
\end{aligned}
$$

$a = b = 0$ という簡単な場合, $X(t)$ の母平均値, 母分散は t に正比例する.

$X(t)$ の母標準偏差 $= \sqrt{\mathrm{V}[X(T)]} = \boxed{}$

ってことは, ランダムウォークの座標の確率分布の時間変化はこんな感じ?



L08-Q1

Quiz(ランダムウォーカーの到達点の座標の母平均・母分散)

時刻 $t = 3$ に, $x = 5$ から出発するランダムウォークの座標 $X(t)$ を考える.
時刻ごとに, 時刻 1 あたりの座標の増分 $R(t)$ は

$$\mathrm{E}[R(t)] = -3, \mathrm{V}[R(t)] = 5$$

を満たす確率変数である. 時刻ごとの増分は独立である.

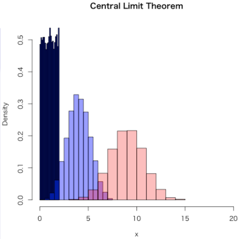
1. $X(20)$ の母平均値を求めよう.
2. $X(20)$ の母分散を求めよう.
3. この母平均値と母分散を持つ正規分布の確率密度関数 $f(x)$ を書こう.

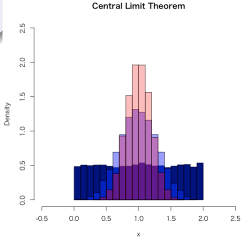
中心極限定理 西川確率統計 §4.2 確率統計☆演習 I(2017)L10

中心極限定理 (いいかげんバージョン)

$X_1, \dots, X_n$ が母平均値 μ, 母分散 σ^2 の独立同分布に従うとき, $n \to +\infty$ で

- $U_n = X_1 + \cdots + X_n$, の確率分布は, の $\boxed{\qquad\qquad\qquad\qquad}$ に似る





ランダムウォークの場合.
$n \to t$, $X_i \to R(i)$. $U_n \to X(n)$
と思って適用できる.
問: さっきのランダムウォークの $X(20)$ の確率分布を描いて.

ここまで来たよ

7 連続型確率変数の擬似乱数

8 ランダムウォーク, 自己回帰モデル, 時系列解析
- (連続型) ランダムウォークの座標の母平均値と母分散
- 自己回帰モデル

m 次の自己回帰モデル=AR モデル Auto Regression

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, $R(t)$ は同分布にしたがい,

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

なお, 最初の 2 つの式を満たす $R(t)$ をホワイトノイズ, 白色雑音 という.

自己回帰モデルで記述できそうな現象の例

- $t =$ 日, $X(t) =$ 気温の, 平年平均気温との差, $0 < \phi < 1$.
- $t =$ シーズン, $X(t) =$ タイガースの勝率の 5 割との差, $-1 < \phi < 0$.

$\mathrm{E}[R] = 0$ なランダムウォーク $\subset$ AR(1) $\subset$ 自己回帰モデル $\subset$ 時系列

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) モデルとランダムウォーク

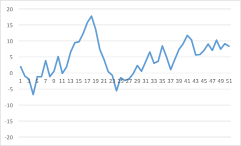
AR(1) $a_1 = \phi$ (一般の値).

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

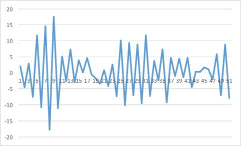
AR(1) で $\phi = 1$ で, $R(t)$ が独立とすると, ランダムウォークで $\mathrm{E}[R] = 0$ であるもの実現できる.

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

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



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



L08-Q2

Quiz(AR(1) モデルの例)

AR(1) モデルで, 出発点を $X(0) = 100.0$ とする. 得られた乱数の値を, $R(1) = 15.0, R(2) = -8.0$ とする. $X(1), X(2)$ を, 3 つの場合 $\phi = 1.0, 0.9, -0.9$ について求めよう.

定常過程

定常過程

確率過程 $X(t)$ が

- $\mathrm{E}[X(t)]$ 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

であるとき, 定常過程という.

要するに $\boxed{}$
ある極限, たとえば $t \to \infty$ のみで定常過程であることはある.

- ランダムウォークは定常過程 $\boxed{}$
- 一般に, AR(m) モデルが定常かどうかは a_k による.

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 $r(k) = \frac{C(k)}{C(0)}$.

k をラグ lag という
横軸ラグ k, 縦軸自己相関係数 $r(k)$ の棒グラフを コレログラムという.

AR(1) モデルの自己共分散

$$
\begin{aligned}
X(t) =& \phi X(t-1) + R(t)\\
=& \phi(\phi X(t-2) + R(t-1)) + R(t)\\
=& \cdots = \phi^t X(0) + \phi^{t-1} R(1) + \cdots \phi R(t-1) + R(t).\\
\mathrm{E}[X(t)] =& \phi^t \mathrm{E}[X(0)]. \quad \text{(初期条件の影響は}\cdots\text{)}
\end{aligned}
$$

以下 $\mathrm{E}[X(t)] = 0$ と仮定.

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

展開すると対角項 $\mathrm{E}[R(t')R(t')]$ しか残らない

$$
\begin{aligned}
&= \phi^{2t+k} \mathrm{E}[X(0)X(0)] + (\phi^{2t+k-2} + \phi^{2t+k-4} + \cdots \phi^k)\sigma^2\\
&\overset{t\to\infty}{\to} \begin{cases} \frac{\phi^k \sigma^2}{1-\phi^2} & (|\phi| < 1)\\ (C + t\sigma^2 \text{ のように発散}) & (|\phi| = 1)\\ (t \text{ より速く発散}) & (|\phi| > 1) \end{cases}
\end{aligned}
$$

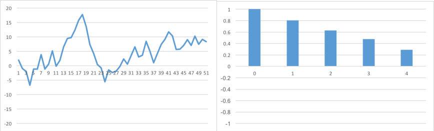
AR(1) モデルの母自己相関係数

$$\mathrm{E}[X(t)X(t)] = \phi^{2t}\mathrm{E}[X(0)X(0)] + (\phi^{2t-2} + \phi^{2t-4} + \cdots + \phi^{0})\sigma^2.$$

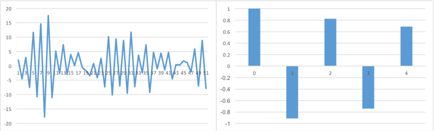
$$k \text{ 次の母自己相関係数 } r(k) = \frac{\mathrm{E}[X(t)X(t+k)]}{\mathrm{E}[X(t)X(t)]} = \phi^k.$$

t-$X(t)$ グラフの例と k-$r(k)$ グラフ (コレログラム)

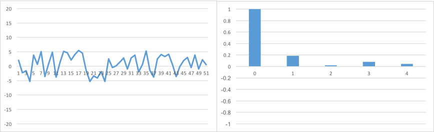
$\phi = 0.9,\ \sigma = 1$



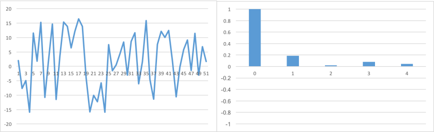
$\phi = -0.9,\ \sigma = 1$



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



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



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	0.93494	1.39985	0.05638	1.16461	0.77765
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

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	0.93494	1.39985	0.05638	1.16461	0.77765
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	0.93494	1.39985	0.05638	1.16461
3	1	2	-0.2522	-1.4293	0.41339	0.93494	1.39985	0.05638	1.16461	0.77765
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

お知らせ

提出場所

https://learn.math.ryukoku.ac.jp/moodle

モバイルアプリ

https://download.moodle.org/mobile

通信量を抑えられるスキャナアプリ．　おすすめ: CamScanner on iOS/Android

https://www.camscanner.com/

- チューター/Math ラウンジ 月火水木昼 1-614
- 2018-06-20 水 数理情報セミナー履修説明会
- 2018-06-22 金 初夏のプチテスト (プログラミング実技)
- Visual Studio 自宅インストールおすすめ中．計算科学☆実習のページから．