ランダムウォークの座標の標本抽出と推定

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 L03(2017-04-24 Mon)

最終更新: Time-stamp: "2017-04-24 Mon 18:24 JST hig"

今日の目標

- 標本から, ランダムウォークの座標の母平均値, 母分散, 母期待値, 母比率が点/区間推定できる.
- ランダムウォークの座標の標本抽出するプログラムが書ける.

http://hig3.net

L02-Q1
Quiz 解答:ランダムウォークの確率と座標の期待値

1. $$P(X(3)=x)=\begin{cases}\binom{3}{0}p^0(1-p)^3 & (x=0)\\ \binom{3}{1}p^1(1-p)^2 & (x=1)\\ \binom{3}{2}p^2(1-p)^1 & (x=2)\\ \binom{3}{3}p^3(1-p)^0 & (x=3)\end{cases}$$
2. $\mathrm{E}[X(3)]=(1-p)^3 0+3p(1-p)^2 1+3p^2(1-p)^1 2+p^3 3=3p.$
3. $\mathrm{E}[X(3)^2]=(1-p)^3 0^2+3p(1-p)^2 1^2+3p^2(1-p)^1 2^2+p^3 3^2=6p^2+3p.$
   $\mathrm{V}[X(3)]=\mathrm{E}[X(3)^2]-\mathrm{E}[X(3)]^2=3p(1-p).$
4. $\mathrm{E}[\mathbf{1}_{[X>1]}(X(3))]=(1-p)^3 0+3p(1-p)^2 0+3p^2(1-p)^1 1+p^3 1.$

L02-Q2
Quiz 解答:ランダムウォーカーの到達点の座標の母平均・母分散

1. $\mu = \mathrm{E}[R(t)] = (-1)\cdot\frac{5}{9} + 0\cdot\frac{1}{9} + (+1)\cdot\frac{3}{9} = -\frac{2}{9}$.
2. $\mathrm{E}[(R(t))^2] = (-1)^2\cdot\frac{5}{9} + 0^2\cdot\frac{1}{9} + (+1)^2\cdot\frac{3}{9} = \frac{8}{9}$. $\mathrm{E}[R(t))^2] = (-\frac{2}{9})^2$.
   $\sigma^2 = \mathrm{V}[R(t)] = \mathrm{E}[(R(t))^2] - \mathrm{E}[R(t)]^2 = \frac{68}{81}$.
   次のように直接に計算しても同じ結果になる.
   $\mathrm{V}[R(t)] = \mathrm{E}[(R(t)-\mu)^2] =$
   $((-1)-(-\frac{2}{9}))^2\cdot\frac{5}{9} + (0-(-\frac{2}{9}))^2\cdot\frac{1}{9} + ((+1)-(-\frac{2}{9}))^2\cdot\frac{3}{9}$.
3. $\sigma_{R(t)} = \sqrt{\frac{68}{81}} = \frac{2\sqrt{17}}{9}$.
4. 一般に $\mathrm{E}[X(T)] = T\cdot\mathrm{E}[R(t))$. $T = 20$ とすると,
   $\mathrm{E}[X(20)] = 20\mathrm{E}[R(t)] = -\frac{40}{9}$.
5. 一般に $\mathrm{V}[X(T)] = T\cdot\mathrm{V}[R(t))$. $T = 20$ とすると,
   $\mathrm{V}[X(20)] = 20\mathrm{V}[R(t)] = \frac{1360}{81}$.
6. $\sigma_{X(20)} = (\mathrm{V}[X(20)])^{1/2} = (\frac{1360}{81})^{1/2} = \frac{4}{9}\sqrt{85}$

L02-Q3

Quiz 解答:ランダムウォーカーの到達点の座標の母平均・母分散

1. $R(-4) + \cdot + R(3) \sim \mathrm{B}(8, \frac{2}{3})$ が, 値 $5 - 2 = 3$ をとる確率を求める.
   $\binom{8}{3} (\frac{2}{3})^3 (\frac{1}{3})^5$.
2. $\mathrm{E}[X(3)] = \mathrm{E}[X(-5) + R(-4) + \cdots + R(3)] = 2 + 8\frac{2}{3}$.
3. $\mathrm{V}[X(3)] = \mathrm{V}[X(-5) + R(-4) + \cdots + R(3)] = 8\frac{2}{3}\frac{1}{3}$.
4. $\sqrt{8\frac{2}{3}\frac{1}{3}}$.

L02-Q4

Quiz 解答:ランダムウォーカーの到達点の座標の母平均・母分散

$\mathrm{E}[R(t)] = -\frac{2}{9}$, $\mathrm{V}[R(t)] = \frac{68}{81}$ に注意する.

1. $\mathrm{E}[X(7)] = \mathrm{E}[5 + R(4) + R(5) + R(6) + R(7)] = 5 + 4 \times (-\frac{2}{9}) = \frac{37}{9}$.
2. $\mathrm{V}[X(7)] = \mathrm{V}[5 + R(4) + R(5) + R(6) + R(7)] = 4 \times \frac{68}{81} = \frac{272}{81}$.
3. $\left(4 \times \frac{68}{81}\right)^{1/2} = \frac{4}{9}\sqrt{17}$.

$\mathrm{V}[X(7)] = \mathrm{E}[X(7)^2] - \mathrm{E}[X(7)]^2$ に行った人は, そこまでは正しいけど, $\mathrm{E}[X(7)^2]$ を計算しようとしたら容易なことじゃないよ. $X(7)$って 9 通りの値をとるんだからそれ全部加えないと…
それなのに, $\mathrm{E}[X(7)^2] \neq \mathrm{E}[R(t)^2]$ や,
$\mathrm{E}[X(7)^2] = \mathrm{E}[(5 + R(4) + R(5) + R(6)^2 + R(7)^2] \neq$
$\mathrm{E}[R(4)^2] + \mathrm{E}[R(5)^2] + \mathrm{E}[R(6)^2] + \mathrm{E}[R(7))^2]$ の $\neq$ を $=$ であることにして計算しちゃった人が一定数いました. そんなの成立する理由ないでしょ. NG ワードで減点したいくらい. 確率論としても, ウォーカーの意味をつけて物理的に考えてもへんです.

ここまで来たよ

2. 離散座標ランダムウォークの座標の確率・母平均値・母分散
   - ランダムウォークの座標分布の正規近似

3. ランダムウォークの座標の標本抽出と推定
   - 標本からの推定
   - 標本抽出するプログラム
   - 母比率の推定の例
   - 擬似乱数生成の仕組みとシード

独立同分布を利用して中心極限定理で近似 (T が大きいとき)

ここでは, $X(0) = 0$ に限って考える. 塚田確率統計 §5.3 塚田確率統計 §5.3.1

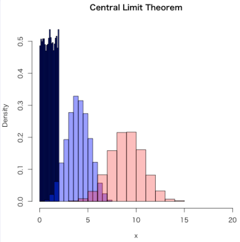
確率統計☆演習 I(2016)L9

中心極限定理 (いいかげんバージョン)

$R(1), \ldots, R(T)$ が母平均値 μ_{R}, 母分散 σ_{R}^2 の独立同分布に従うとき,

- $X(T) = R(1) + \cdots + R(T)$, の確率分布は, $T \to +\infty$ で, 正規分布 $\mathrm{N}(T \cdot \mu_{\mathrm{R}}, T \cdot \sigma_{\mathrm{R}}^2)$ に似る

確率統計☆演習 I(2015)L09



$\rightsquigarrow$

ランダムウォークの座標 $X(T)$ の確率分布は, T が大きいとき, 母平均値 $T \cdot \mu_{\mathrm{R}}$, 母分散 $T \cdot \sigma_{\mathrm{R}}^2$ の正規分布にほぼ従う.

ここまで来たよ

2 離散座標ランダムウォークの座標の確率・母平均値・母分散
- ランダムウォークの座標分布の正規近似

3 ランダムウォークの座標の標本抽出と推定
- 標本からの推定
- 標本抽出するプログラム
- 母比率の推定の例
- 擬似乱数生成の仕組みとシード

こんなこと考えたいんだった

ランダムウォークの座標 $X(t)$ について,
(前回) 手計算で以下の母ナントカを厳密に求めよう.
(今回) 標本から以下の母ナントカを点推定/区間推定しよう.

- $\mathrm{E}[X(2)]$, $\mathrm{E}[\mathrm{e}^{X(2)}]$, $X(2) > 1$ となる確率 (比率)
- $\mathrm{E}[X(1002)]$, $\mathrm{E}[\mathrm{e}^{X(1002)}]$, $X(1002) > 51$ となる確率 (比率)
- $X(50) = 12$ かつ $X(100) = 25$ となる確率 (比率)

確率シミュレーションによる母ナントカの推定

X: 一般の確率変数.
母ナントカの公式は忘れたけどコンピュータはあるとする.
擬似乱数を使ってサイズ N の標本 $X^{(1)}, X^{(2)}, \dots, X^{(N)}$ を作って, 母平均値 $\mathrm{E}[X]$ を標本平均値

$$\overline{X} = \frac{1}{N} \sum_{n=1}^{N} X^{(n)}$$

で推定すれば?

点推定

母平均値の推定

$$\text{標本平均値 } \overline{X} = \frac{1}{N}(X^{(1)} + \cdots + X^{(N)}) = \frac{1}{N}\sum_{n=1}^{N} X^{(n)}$$

が, 母平均値 $\mathrm{E}[X]$ の ‘よい’ 推定値になっている. 塚田確率統計 p.110

母分散の推定

$$\text{不偏標本分散 } S^2 = \frac{1}{N-1}[(X^{(1)} - \overline{X})^2 + \cdots + (X^{(N)} - \overline{X})^2]$$

$$= \frac{N}{N-1}\left[\frac{1}{N}\sum_{n}(X^{(n)})^2 - \left(\overline{X}\right)^2\right]$$

が, 母分散 $\mathrm{V}[X]$ の ‘よい’ 推定値になっている. 塚田確率統計 p.120

母期待値の推定

標本期待値

$$\overline{\phi(X)} = \frac{1}{N}\sum_{n=1}^{N}\phi(X^{(n)})$$

が 母期待値 $\mathrm{E}[\phi(X)]$ の ‘よい’ 推定値になっている.

理由: 確率変数 $Y = \phi(X)$ の母平均値の推定と同じこと.

標比率の推定

ベルヌーイ分布 $\mathrm{B}(1,p)$ に従う Y について，サンプルのデータ N 個中 k 個が 1 であるとき，

$$\text{標本比率 } \hat{p} = \frac{k}{N}$$

が Y の母比率 p のよい推定値になっている. 塚田確率統計 p.130

母比率の推定

サンプルのデータ N 個中 k 個が「条件…を満たす」とき，

$$\text{標本比率 } \hat{p} = \frac{k}{N}$$

が「…」の母比率 $\mathrm{E}[\mathbf{1}_{[\cdots]}(X)]$ のよい推定値になっている. 塚田確率統計 p.130

理由 $\phi(X) = \mathbf{1}_{[\cdots]}(X)$ と思えば，$Y = \phi(X)$ はベルヌーイ分布にしたがう.

L03-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)

仕組みのよくわからないランダムウォークで標本抽出したところ, $X(3)^{(n)}$ が

$$3, 3, 3, 1, 1, 1, 1, 1, -1, -3$$

だった ($N = 10$).

1. $\mathrm{E}[X(3)]$ を推定しよう.
2. $\mathrm{V}[X(3)]$ を推定しよう.
3. $\mathrm{E}[X(3)^3]$ を推定しよう.
4. $X(3) > 1$ となる確率 (母比率) を推定しよう.

区間推定

ランダムウォークで $X(T)$ は近似的に正規分布に従う.
そこで, 正規分布を仮定した区間推定 塚田確率統計 §7.3 も使える.
L03-Q2

Quiz(母平均値の区間推定 (母分散未知) 母分散の区間推定)

あるエスプレッソメーカーの作る 1 杯分のエスプレッソの体積 X は, 未知の母平均値 $\mu\mathrm{cm}^3$ と母分散 $\sigma^2(\mathrm{cm}^3)^2$ の正規分布にしたがう確率変数である. $n = 3$ 杯いれてみたところ, 体積は,

$$28\mathrm{cm}^3, \quad 30\mathrm{cm}^3, \quad 32\mathrm{cm}^3$$

だった.

1. 母平均値 μ を, t 分布を用いて, 信頼係数 $1 - \alpha = 0.99$ で区間推定しよう.
2. 母分散 σ^2 を, カイ二乗分布を用いて, 信頼係数 $1 - \alpha = 0.95$ で区間推定しよう.

確率シミュレーション

確率シミュレーション

確率的現象を, 擬似乱数を使ってそのままコンピュータ上で再現し (simulate), くり返し実行して標本抽出し, 母ナントカを推定すること.

- とりあえずなんでも計算 (ていうか $\boxed{}$) できちゃう
- 要 $\boxed{}$

最終的な誤差= $\boxed{\boxed{\text{統計誤差}}}$ + $\boxed{\text{数値計算誤差}}$

対比される計算方法

先週のように, $\mathrm{E}[X(t)]$ のような母ナントカを公式で書き, それをプログラムで計算する作戦.

最終的な誤差= $\boxed{\text{数値計算誤差}}$

ここまで来たよ

2. 離散座標ランダムウォークの座標の確率・母平均値・母分散
   - ランダムウォークの座標分布の正規近似

3. ランダムウォークの座標の標本抽出と推定
   - 標本からの推定
   - 標本抽出するプログラム
   - 母比率の推定の例
   - 擬似乱数生成の仕組みとシード

欲しい出力

$X(t)^{(n)}$

t: $\boxed{}$

(n): $\boxed{}$

	$t=0$	$t=1$	$\cdots$	$t=T$
$n=1$	$X(0)^{(1)}$,	$X(1)^{(1)}$,	$\cdots$	$X(T)^{(1)}$, 改行
$n=2$	$X(0)^{(2)}$,	$X(1)^{(2)}$,	$\cdots$	$X(T)^{(2)}$, 改行
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n=N$	$X(0)^{(N)}$,	$X(1)^{(N)}$,	$\cdots$	$X(T)^{(N)}$, 改行

$X(0), X(1), X(2), \ldots, X(T)$ の標本を抽出するプログラム

```
                                        /*1*/
for(n=0;n<N;n++){
                                        /*2*/

   for(t=0;t<T;t++){
                                        /*3*/
      x=x+getrandom(getuniform());
                                        /*4*/
   }
                                        /*5*/
}
                                        /*6*/
```

問: `srand(seed)`, `x=0`, `printf("%d,",x)` はどこ?
問: 他に何がいる?

標本から推定

	$t=0$	$t=1$	$\cdots$	$t=T$
$n=1$	$X(0)^{(1)}$,	$X(1)^{(1)}$,	$\cdots$	$X(T)^{(1)}$, 改行
$n=2$	$X(0)^{(2)}$,	$X(1)^{(2)}$,	$\cdots$	$X(T)^{(2)}$, 改行
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n=N$	$X(0)^{(N)}$,	$X(1)^{(N)}$,	$\cdots$	$X(T)^{(N)}$, 改行

Excel の関数: average, var, if(条件, 真のときの式, 偽の時の式), sum

Excel を使わないで標本期待値の計算

$$\overline{\phi(X(T))} = \frac{1}{N}\sum_{n=1}^{N}\phi(X(T)^{(n)})$$

```
                                    /*1*/
for(n){
                                    /*2*/
  for(t){
                                    /*3*/
    x=x+getrandom(getuniform());
                                    /*4*/
  }
                                    /*5*/
}
                                    /*6*/
```

sum1=0, sum1+=x*x*x ($\phi(x) = x^3$ のとき),
printf("%f",(double)sum1/N)?

ここまで来たよ

2. 離散座標ランダムウォークの座標の確率・母平均値・母分散
   - ランダムウォークの座標分布の正規近似

3. ランダムウォークの座標の標本抽出と推定
   - 標本からの推定
   - 標本抽出するプログラム
   - 母比率の推定の例
   - 擬似乱数生成の仕組みとシード

$\mathbf{1}_{[\text{条件}]}(X(T))$ のこと考えると…

sum1 は $\boxed{}$ と名付けたほうがいいかも.

母比率の推定の例

L03-Q3

例題

$t = 2$ に $x = 10$ から出発したランダムウォーカーが, $t = 20$ で, 領域 $x < 0$ にいる確率を推定しよう.
`double getuniform()` と (未知の)`int getrandom(double y)` は与えられているとして, `main()` だけをかけばよい.

ランダムウォークの言葉づかいの習慣

$X(2)$：初期条件, ランダムウォーカーの出発点 (を確率変数とみたもの)

- 「ランダムウォーカーが時刻 $t = 2$ に $x = 3$ から出発した」⇔
- 「$t = 20$ で $x < 0$ にいる」⇔

L03-Q4

Example

$t = 0$ に $x = 3$ から出発したランダムウォークが, $T = 10$ で $|X(T)| < 3$ である確率を推定して出力するプログラムを書こう.

ここまで来たよ

2. 離散座標ランダムウォークの座標の確率・母平均値・母分散
   - ランダムウォークの座標分布の正規近似

3. ランダムウォークの座標の標本抽出と推定
   - 標本からの推定
   - 標本抽出するプログラム
   - 母比率の推定の例
   - 擬似乱数生成の仕組みとシード

離散型擬似乱数列の生成

```
#include <stdlib.h>

/* 0以上 RAND_MAX 以下の正の整数をランダムに選んで返す関数 */
int rand();
/* その初期化 */
void srand(unsigned seed);
```

```
/* [0,1) 一様乱数 */
double getuniform(){
  return rand()/(1.0+RAND_MAX);
}
```

getuniform() の性質

- '値域' は $[0,1)$. $0 \leq \texttt{getuniform()} < 1$.
- ($\texttt{getuniform()} < r$ となる確率)$= r$. $\quad (0 \leq r \leq 1)$

計算機の頭の中どうなってんの?

擬似乱数列＝'ほぼ' ランダムな数列

(毎回別々の)seed を指定することは必要

- seed に応じて, 毎回異なる乱数列を得られる.
- 特定の乱数列に対する動作を再現できる. デバッグでは必須.
- seed を適切に設定すると, 複数回の実行で, 別々の (独立な) 標本抽出が行える.

L03-Q5

Quiz(rand() の振る舞い)

次のプログラムで, seed を無作為に入力するとき, A が出力される確率は?

```
int getrandom(double y){
  if( y<1/3.0 ){
    return 0;
  }else{
    return 1;
  }
}

int main(){
  int seed;
  scanf("%d",&seed);
  srand(seed);
  if( getrandom()==getrandom() ){
    printf("A\n");
  }
  return 0;
}
```

1. 0
2. $1/2$
3. $5/9$ に近い
4. $6/9$ くらい
5. 1

L03-Q6

Quiz(rand() の振る舞い)

次のプログラムで, seed を無作為に入力するとき, A が出力される確率は?

```
int getrandom(double y){
  if( y<2/13.0 ){
    return 0;
  }else if ( y<6/13.0){
    return 3;
  } else {
    return 4;
  }
}

int main(){
  int seed;
  scanf("%d",&seed);
  srand(seed);
  if( getrandom()>1 ){
    if( getrandom()>3.5 ){
      printf("A\n");
    }
  }
  return 0;
}
```

1. 0
2. $\frac{7}{13}$
3. $\frac{77}{13^2}$ に近い
4. $\frac{121}{13^2}$ くらい
5. 1

お知らせ

- 2017-04-26 水 3 実習の春のプチテスト
- 2017-05-01 月 4 臨時教室変更 3-B105 実習室
- チューター/Math ラウンジ 月火水木昼 1-614
- 2017-05-08 月 統計検定申込締切
- 2017-06-18 統計検定