ランダムウォークの座標の標本抽出

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　L02(2019-04-18 Thu)

最終更新: Time-stamp: ”2019-04-29 Mon 22:11 JST hig”

今日の目標

- 擬似乱数列でランダムウォークの座標を標本抽出するプログラムを書ける
- 標本ナントカから母ナントカを点推定, 区間推定できる

L01-Q1

Quiz 解答:連続的な確率変数の母平均値・母分散・母標準偏差・確率 (一様分布)

1. $\mathrm{E}[\cos(\pi X)] = \int_{5/2}^{3} 2\cos(\pi x)\,\mathrm{d}x = [\frac{2}{\pi}\sin(\pi x)]_{5/2}^{3} = -\frac{2}{\pi}$
2. $P(\frac{22}{8} < X < \frac{23}{8}) = \mathrm{E}[\mathrm{I}_{[\frac{22}{8}<X<\frac{23}{8}]}(X)] = \int_{22/8}^{23/8} 2\,\mathrm{d}x = \frac{1}{4}.$

L01-Q2

L01-Q3

L01-Q4

Quiz 解答:期待値

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

値 $R = 3$ が返される確率の検算.
$P(R = 3) = P(Y < 2/13$ でない かつ $Y < 6/13) = P(2/13 \le Y < 6/13) = \int_{2/13}^{6/13} 1 \, \mathrm{d}y = 4/13$.

ここまで来たよ

- 略解:ランダムウォークと離散型確率分布にしたがう擬似乱数

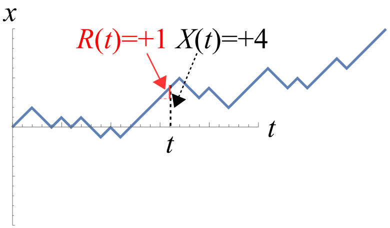
1. ランダムウォークの座標の標本抽出
   - 標本からの推定
   - 確率シミュレーション
   - ランダムウォークの確率シミュレーション

ランダムウォーク (確率過程の例)

ランダムウォーク

$$X(t) = X(t-1) + R(t), \quad X(0) = x_0$$

階差数列 $R(t)$: 独立同分布にしたがう確率変数.
ランダムウォーカー座標 $X(t)$: 確率変数.



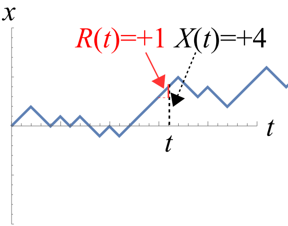
ランダムウォークの座標の母期待値，母比率は?

$R(t)$ やランダムウォークの座標 $X(1000)$ は確率変数.

ありうる問 (確率)

(手計算で) 母ナントカを厳密に求めよう.

- 母平均値 $\mathrm{E}[R(t)]$, 母期待値 $\mathrm{E}[\mathrm{e}^{R(t)}]$, 母比率 $P(R(t) > 1)$
- 母平均値 $\mathrm{E}[X(1000)]$, 母期待値 $\mathrm{E}[\mathrm{e}^{X(1000)}]$, 母比率 $P(X(1000) > 1)$
- 母比率 $P(X(50) = 12$ かつ $X(100) = 25)$



別のタイプの問 (推定)

確率分布 $f(x)$ の式を知らない (例:誰かが作った getrandom の中身不明) けど, データ (標本) だけはあるとする. 母ナントカを推定したい.

点推定

母平均値の推定 前園確率統計 5.1 確率統計☆演習 I(2018)L10

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

が, 母平均値 $\mathrm{E}[X]$ の ‘よい’ 推定値になっている.

母分散の推定 前園確率統計 5.1 確率統計☆演習 I(2018)L10

$$\begin{aligned}\text{不偏標本分散 } S^2 &= \frac{1}{N-1}[(X^{(1)} - \overline{X})^2 + \cdots + (X^{(N)} - \overline{X})^2] \\ &= \frac{N}{N-1}\left[\frac{1}{N}\sum_{n}(X^{(n)})^2 - (\overline{X})^2\right]\end{aligned}$$

が, 母分散 $\mathrm{V}[X]$ の ‘よい’ 推定値になっている.

母期待値の推定 前園確率統計 5.1 確率統計☆演習 I(2018)L10

$$\text{標本期待値}\ \overline{\phi(X)} = \frac{1}{N}\sum_{n=1}^{N}\phi(X^{(n)})$$

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

確率変数 $Y = \phi(X)$ の母平均値の推定と思おう.

母比率の推定 前園確率統計 5.1 確率統計☆演習 I(2018)L10

X のサンプルのデータ N 個中 k 個が「条件 $\cdots$ を満たす」とき,

$$\text{標本比率}\hat{p} = \frac{k}{N} = \frac{1}{N}\sum_{n=1}^{N}\mathrm{I}_{[\cdots]}(X^{(n)})$$

が母比率 $P(\text{「条件}\cdots\text{」}) = \mathrm{E}[\mathrm{I}_{[\cdots]}(X)]$ の 'よい' 推定値になっている.

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

母平均値の信頼区間 (母分散未知)

前園確率統計 p.82　確率統計☆演習 I(2018)L11

(母分散未知の) 正規分布 $\mathrm{N}(\mu, ?^2)$ にしたがう母集団から, サイズ N の標本を得たとき, 母平均値 μ の 信頼係数 $1-\alpha$ の信頼区間は

$$\overline{X} - t(N-1;\alpha/2) \times \sqrt{s^2/N} < \mu < \overline{X} + t(N-1;\alpha/2 \times \sqrt{s^2/N}.$$

ただし, s^2:不偏標本分散, $t(N-1;\alpha/2)$: 自由度 $N-1$ の t 分布の上側確率が $\alpha/2$ となる点 (表から求める).

母比率の信頼区間 (母分散未知)

前園確率統計 §5.4　確率統計☆演習 I(2018)L11

X のサイズ N の標本で, 標本比率 $\hat{p} = k/N$ のとき, 母比率の信頼係数 $1-\alpha = 0.95$, 信頼係数 $1-\alpha = 0.99$ の信頼区間は

$$\hat{p} - 1.96 \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})} < p < \hat{p} + 1.96 \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})},$$

$$\hat{p} - 2.58 \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})} < p < \hat{p} + 2.58 \times \sqrt{\frac{1}{N}\hat{p}(1-\hat{p})}.$$

L02-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,
$X(3)^{(n)} \quad n = 1, 2, \cdots, 10$ が

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

だった.

1. 母平均値 $\mathrm{E}[X(3)]$ を点推定しよう.
2. 母分散 $\mathrm{V}[X(3)]$ を点推定しよう.
3. 母期待値 $\mathrm{E}[X(3)^3]$ を点推定しよう.
4. 母比率 $P(X(3) > 1)$ を点推定しよう.

ここまで来たよ

略解:ランダムウォークと離散型確率分布にしたがう擬似乱数

1. ランダムウォークの座標の標本抽出
   - 標本からの推定
   - 確率シミュレーション
   - ランダムウォークの確率シミュレーション

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), 繰り返して実行して標本抽出して, 母ナントカを推定すること.

- いつでも推定はできちゃう
- 要 $\boxed{}$

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

対比される計算方法 (いま使わない)

$\mathrm{E}[X] = \sum_{x=1}^{3} x f(x)$ のような母ナントカの公式をプログラムで計算する.

$f(x)$ が求まったら苦労しないよ

最終的な誤差 $=$ $\boxed{\text{数値計算誤差}}$

確率変数 $R(1)$ の標本

$R(1)^{(n)}$

$t = 1$: $\boxed{}$

(n): $\boxed{}$

n	$t = 1$
1	$R(1)^{(1)}$, 改行
2	$R(1)^{(2)}$, 改行
$\vdots$	$\vdots$
N	$R(1)^{(N)}$, 改行

ソースコード 1: 擬似乱数

```
/*
rand1.c -- -1 or +1 を確率1/4, 3/4で選ぶ乱数
Time-stamp: "2019-04-16 Tue 18:04 JST hig"
*/
#define _CRT_SECURE_NO_WARNINGS // Visual C++用おまじない
#include <stdio.h>
#include <stdlib.h>   /* srand(), rand() を使うのに必要 */

/* 関数プロトタイプ宣言 */
int getuniform();
int getrandom(double y);

int main(){
  int seed;         /* 疑似乱数のシード */
  int n;            /* カウンタ 標本内通し番号*/
  int nmax=100;     /* 疑似乱数を得る回数=サンプルサイズN */

  scanf("%d",&seed);
  srand(seed);          /* シードの設定 */
  for(n=0;n<nmax;t++){ /* 数式とnは1ずれてる*/
    /* srand(seed);  */ /*ここに置くと? */
    printf("%d,%d\n",n,getrandom(getuniform());
  }
  return 0;
}
/** [0,1) 一様疑似乱数を返す */
double getuniform(){
  return rand()/(RAND_MAX+1.0);
}
/** -1 or +1 を確率1/4, 3/4 で返す乱数 */
int getrandom(double y){
  if( y < 0.25 ){
    return -1;
  } else {
    return +1;
  }
}
```

ここまで来たよ

- 略解:ランダムウォークと離散型確率分布にしたがう擬似乱数

1. ランダムウォークの座標の標本抽出
    - 標本からの推定
    - 確率シミュレーション
    - ランダムウォークの確率シミュレーション

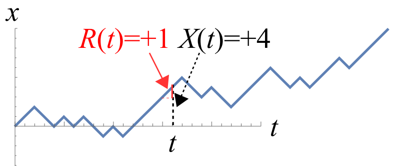
ランダムウォークの座標 $X(t)$ の標本

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

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=1;t<=T;t++){
                                        /*3*/
    x=x+getrandom(getuniform());
                                        /*4*/
  }
                                        /*5*/
}
                                        /*6*/
```

解説動画

https://www.youtube.com/watch?v=H-sIaVJmeiA

問: `srand(seed)`,x=0,t=0,`printf("%d,",x)` はどこ?

問: 他に何がいる?

標本期待値は Excel を使わなくても C の中でも計算できる!

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

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

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

標本比率の C による計算

$\phi(x) = \mathrm{I}_{[\text{条件}]}(x)$ に相当する `phi` は?

sum1 は $\boxed{}$ と名付けたほうがいいかも.

L02-Q2

Quiz(確率シミュレーション)

$t = 2$ 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て出力するプログラムを書こう. ただし,

$$X(t) = X(t-1) + R(t)$$

で, `int getrandom(getuniform())` が独立同分布にしたがう確率変数 $R(t)$ を返すものとする. main と phi の中だけ書こう.

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」$\Leftrightarrow$ $\boxed{}$
- 「$t = 20$ で $x < 0$ にいる」$\Leftrightarrow$ $\boxed{}$

L02-Q3

Quiz(rand() の振る舞い)

seed を与えられると, $t = 3$ に $x = 4$ から出発したランダムウォークが, $T = 10$ で $|X(T)| < 5$ である確率を推定して出力するプログラムを書こう. ただし,

$$X(t) = X(t-1) + R(t),$$

で, 独立同分布にしたがう確率変数 $R(t)$ を, `int getrandom(getuniform())` が返すものとして, main と phi の中だけ書こう.

予習復習問題のやり方+今後の予定

Learn Math Moodle
`https://learn.math.ryukoku.ac.jp/moodle`

Moodle App for iOS/Android

URL をきかれたら `https://learn.math.ryukoku.ac.jp/moodle` で登録.

お知らせ

実習室整備状況, 登録者数により未確定です. 当日に My 時間割を確認してください. いずれにせよ, 教科書と (PC につながる) イヤフォン用意してください.

- 2019-04-23 火 4 実習
- Math ラウンジ 月火水木昼 1-614
- 樋口オフィスアワー火 5(1-507/1-542)