母分布と標本分布, 期待値の計算

樋口さぶろお

龍谷大学理工学部数理情報学科

計算科学☆演習 II L02(2014-04-18 Fri)

今日の目標

1. 母分布 (確率分布) から手でグラフが描ける.
2. (演習) 標本から Excel でヒストグラムが描ける.
3. 母分布から手と数式で母期待値が計算できる.
4. 標本から電卓や Excel で母期待値, 母平均値が推定できる.

http://hig3.net

ここまで来たよ

L01-S3
Quiz 解答:擬似乱数の使いかた

ソースコード 1: 乱数

```
int getrandom(double y){
  if( y<1.0/3.0 ){
    return -1;
  } else if( y<1.0/3.0+1.0/2.0 ){
    return 0;
  } else {
    return +1;
  }
}
```

ここまで来たよ

母分布と標本分布

例 A

R	確率
1	$p(1) = \frac{1}{6}$
2	$p(2) = \frac{1}{3}$
3	$p(3) = \frac{1}{2}$

ここからグラフを描く方法 $\times 2$

母分布　標本分布

母分布と標本分布の比較

	母分布	標本分布
サイズ	1 つに定まっている．本物． → サイズ ∞	‘やる’ たびに違う．実験結果． → 標本サイズ を自由に決められる
例	表 関数 $p(r_j)$ 1 個の getrandom() 1 個のサイコロ 1 個の赤ひげ危機一髪 何かブラックボックス	 実行してみる ふってみる 刺してみる 使ってみる

この他にも, 母ナントカ, 標本ナントカ がいろいろある．

母=population, 標本=sample

母ナントカと標本ナントカの関係

標本抽出 母ナントカから標本ナントカを作る

- サンプルに 1 個ずつデータを追加 (試行という) していって作る
-
- 母分布がブラックボックスでも, とりあえず使えるデータが手に入る
- 推定に使うため, 好きなサイズのサンプルを作る

推定 標本ナントカから母ナントカをあてる

-
- サンプルが違えば推定結果も違う
- 母分布が (正確には) わからないとき, なるべく正確な推定をしたい.

L02-Q1

Quiz(母分布と標本分布)

以下のようにして得られるのは母分布?標本分布? 母分布のものを何個でも選ぼう.

1. getrandom() 内の定義を読んで，この値が返る確率はいくつ，と考える.
2. getrandom() を for で何度も実行して，どの値が何回返ったか数える.
3. 赤玉白玉のたくさんはいった箱から，中を見ないで取りだして，元に戻す．このとき赤玉である確率を求めるのに，箱の中の赤玉白玉の総数を数える.
4. 宝くじのあたる確率を，宝くじを買ってみることで求める
5. あるコインの表がでる確率がきっちり $1/2$ かどうか調べるために，コインを何回も投げてみる.

ここまで来たよ

1. ランダムウォークと擬似乱数生成
   - Quiz 解説
2. 母分布と標本分布, 期待値の計算
   - 母分布と標本分布
   - (相対) 度数分布表とヒストグラム
   - ヒストグラム
   - 期待値
   - 標本平均値

(相対) 度数分布表

サンプル $\to$ 度数分布表 $\to$ 相対度数表 $\to$ ヒストグラム

例 A

ここまで来たよ

1. ランダムウォークと擬似乱数生成
   - Quiz 解説
2. 母分布と標本分布, 期待値の計算
   - 母分布と標本分布
   - (相対) 度数分布表とヒストグラム
   - ヒストグラム
   - 期待値
   - 標本平均値

ヒストグラム

粗い (相対) 度数分布表を**粗い**グラフ=ヒストグラムにする方法
一定幅の階級, ビン $a < r \leq a + d$ に分けて, その中に入る個数を数える.
d: 階級, ビンの幅
データから Excel で描く方法 $\rightarrow$ 演習
例 AKB48 グループの身長を標本抽出して, ヒストグラムを描こう
$145, 152, 153, \ldots.$

階級	度数	相対度数
145 より大きく 150 以下	7	0.09
150 より大きく 155 以下	17	0.22
155 より大きく 160 以下	29	0.37
160 より大きく 165 以下	19	0.24
165 より大きく 170 以下	4	0.05
170 より大きく 175 以下	1	0.01
175 より大きく 180 以下	0	0.00
180 より大きく 185 以下	1	0.01
合計	78	1.00

ps3id_raicho_1182 さん（最終更新日時:2012/3/20）投稿日：
2012/2/15 http://note.chiebukuro.yahoo.co.jp/detail/n32745

L02-Q2

Quiz(ヒストグラム)

下の 1 変量データについて, 度数分布表とヒストグラムを作ろう. ただし, 階級は, $a = 100, 105, \ldots$, 幅が $d = 5$, 'a より大きく $a + d$ 以下' で定めよう.

X
104
108
108
108
108
112
114
115
115
118

ここまで来たよ

1. ランダムウォークと擬似乱数生成
   - Quiz 解説
2. 母分布と標本分布, 期待値の計算
   - 母分布と標本分布
   - (相対) 度数分布表とヒストグラム
   - ヒストグラム
   - 期待値
   - 標本平均値

母期待値

R: 確率変数

例

R	確率
r_1	$p(r_1)$
r_2	$p(r_2)$
$\cdots$	$\cdots$
r_m	$p(r_n)$

j: とりうる値につけた番号.

当然, 確率の合計 $\sum_{j=1}^{m} p(r_j) = 1.$ (全事象)

$f(R)$: R の関数
例 $f(R) = 2R, R^2, \mathrm{e}^R$, (場合分けで書かれた関数), $\cdots$.

$f(R)$ の母期待値

$$\mathrm{E}(f(R)) = \sum_{j=1}^{m} p(r_j) \times f(r_j)$$

.

よく使われる期待値
確率の合計 $= \mathrm{E}(1) = 1$. つまり $f(R) = 1$,
母平均値 $= \mu = \mathrm{E}(R)$. つまり $f(R) = R$,
母分散 $= \sigma^2 = \mathrm{E}((R-\mu)^2)$. つまり $f(R) = (R-\mu)^2$.

例 A

条件が成立する確率

特徴関数

$$\mathbf{1}_{[\text{条件}]}(r) = \begin{cases} 1 & (r \text{ について条件が成立}) \\ 0 & (r \text{ について条件が不成立}) \end{cases}$$

とすると，

$$\text{条件が成立する確率} = \mathrm{E}(\mathbf{1}_{[\text{条件}]}(r))$$

例 A で，条件 $R \geqq 2$ が成立する確率?

標本から母期待値を推定する

まず母分布を推定して, そこから母期待値を計算することもできるが, もっと楽な方法がある.

標本期待値

母期待値 $\mathrm{E}(f(R))$ は,

$$\text{標本期待値 } \overline{f(R)} = \frac{1}{N}\sum_{i=1}^{N} f(R^{(i)})$$

と推定できる. ‘母期待値を標本期待値で推定する’
(i): データ番号. 値に重複あり. N:標本サイズ.

Excel で計算する方法 $\rightarrow$ 演習

L02-Q3

Quiz(標本抽出と推定)

標本抽出と推定について, 正しい文の番号を答えよう.

1. *母*期待値は, 標本抽出のたびに変化する
2. 標本期待値は, *母*期待値の求め方の一つである.
3. *母*期待値は標本期待値に等しいから, *母*期待値を質問されたら標本期待値を答えればよい.
4. *母*期待値を推定するには, 標本期待値が使える.
5. 標本期待値は, 標本抽出のたびに変化する.

ここまで来たよ

標本平均値

平均値や分散は，期待値の特別な場合だから，母平均値や母分散も標本から推定できる．$\rightsquigarrow$ 標本平均値，標本分散
サイズ N のサンプル $R^{(1)}, R^{(2)}, \ldots, R^{(N)}$ が与えられたとき，

標本平均値

$$\text{標本平均値}\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R^{(i)} = \frac{1}{N}(R^{(1)} + R^{(2)} + \cdots + R^{(N)}).$$

Excel では標本平均値は `average`

例 A

L02-Q4

Quiz(期待値)

確率変数 R は, 値 $R=-2$ を確率 $3/11$ で, 値 $R=-1$ を確率 $1/11$ で, 値 $R=0$ を確率 $2/11$ で, 値 $R=+1$ を確率 $5/11$ でとる.

1. 母期待値 $\mathrm{E}(R^2)$ を求めよう.
2. 母期待値 $\mathrm{E}(R)$ を求めよう.
3. 条件 $\mathrm{e}^R > 0.9$ が成立する確率を求めよう.

L02-Q5

Quiz(標本期待値)

ゆがんだサイコロがある．これを振ってでる目の数を確率変数 R とする．振ることによって，次のようなサンプルを得た．
1,6,1,2,1,3,4,2,2,1

1. (相対) 度数分布表を作ろう
2. ヒストグラムを作ろう
3. 母期待値 $\mathrm{E}(\frac{1}{R})$ を推定しよう．
4. 母期待値 $\mathrm{E}(R)$ を推定しよう．