統計学 2

9. 中心極限定理

矢内　勇生

2019年5月16日

高知工科大学　経済・マネジメント学群

講義要項変更

- 講義要項（シラバス）の内容を変更したので、確認すること
  - 変更点：第14回の内容

今日の目標

- 乱数生成の方法を理解する！
- 中心極限定理を理解する！
  - なぜ正規分布（標準正規分布）ばかり使うのか？

乱数 (random numbers)

- 確率・統計を理解するには、乱数を使うのが一番
  - 実際に実験する
    - サイコロを振る、コインを投げる、etc.
  - 乱数表を使う
  - Rで乱数を生成する

Rで乱数を作る

- Rを乱数生成器 (random number generator) として使う
  - Rで作れるのは擬似乱数 (pseudo-random numbers)
  - メルセンヌ・ツイスタ (Mersenne Twister) が利用されている

Rで生成できる乱数の例 (1)

★ 基本形は r (random) + 分布名の最初の数文字

- 二項分布 (binomial distribution)：`rbinom()`
- 正規分布 (normal distribution)：`rnorm()`
- 一様分布 (uniform distribution)：`runif()`
- カイ二乗分布 (chi-squared distribution)：`rchisq()`
- t 分布 (Student's t distribution)：`rt()`

Rで生成できる乱数の例 (2)

★ 特定の対象の集合から無作為（ランダム）に引く関数

```
sample()
```

Ｒで実際にやってみよう！

- 授業のウェブページ
  - 乱数生成と中心極限定理

正規分布ばかり使うのはなぜか

- 確率分布は、正規分布だけではない
    - 例）一様分布、二項分布
- なぜ正規分布を使って統計的推定・検定を行うのか？

➡ 中心極限定理

中心極限定理 (Central Limit Theorem: CLT)

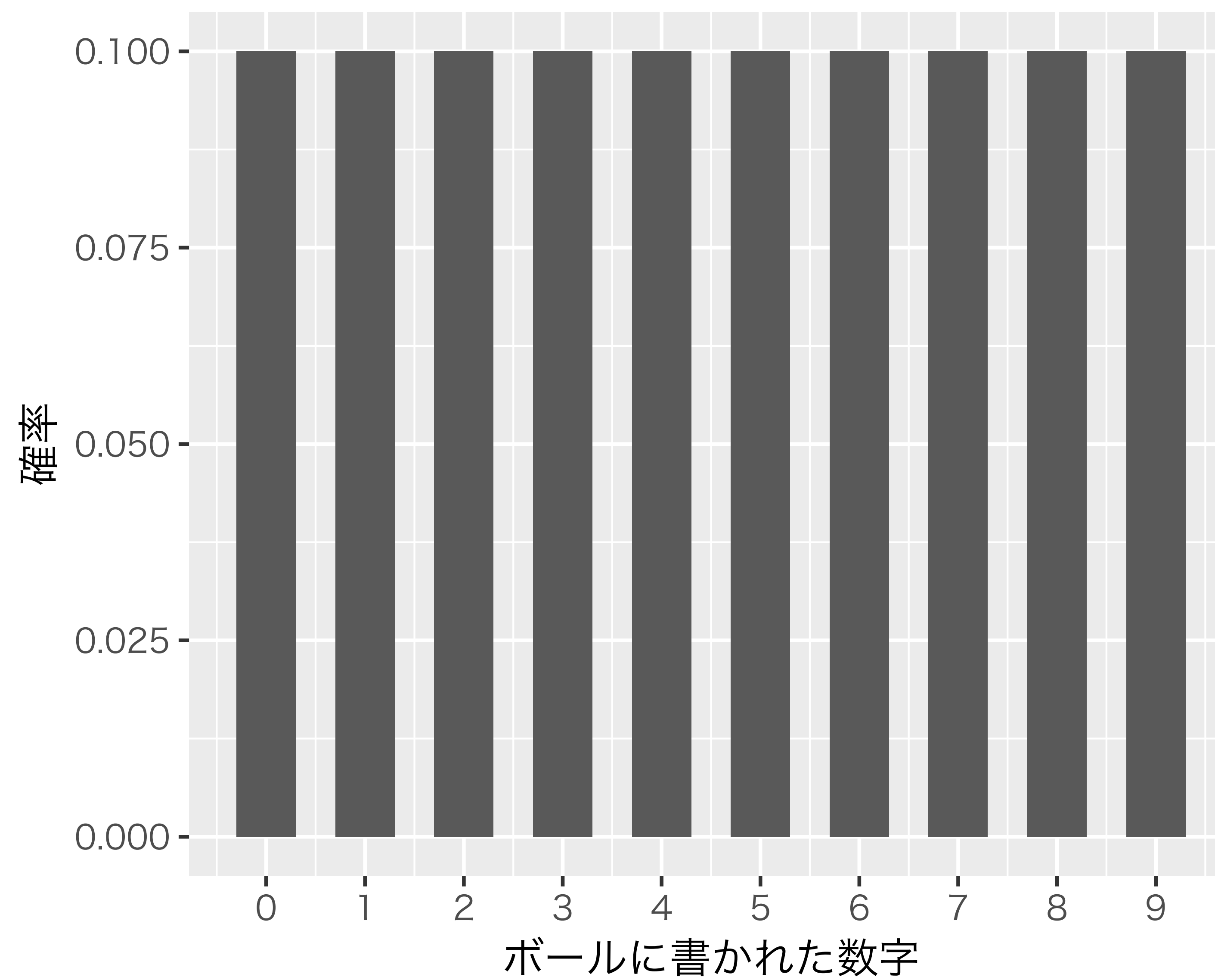
- 標本サイズ n が十分大きければ、元の確率分布がどんなものであっても、誤差は近似的に正規分布に従う

➡ 正規分布以外の確率分布に従う変数であっても、n が大きければ正規分布を使って推定が行える！

★ シミュレーションで示す

離散一様分布

- バッグの中に番号が書かれたボールが10個入っている
  - 番号：0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- この分布の平均 = (9-0)/2 = 4.5



平均値の推定

- バッグ内のボールに書かれた数を知らないとする
- バッグからボールを引いて、平均を当てたい（推定したい）
- バッグからボールを n 回引き、出た数の平均値を推定に使う
- ただし、1度引いたボールはすぐにバッグの中に戻す（復元抽出法）

例：ボールを2回選ぶ

- 1回目の選び方：10通り
- ２回目の選び方：10通り

➡ 選び方は全部で 10 x 10 = 100 通り

- 2個のボールの合計：0 から18までの19通り
- 平均 = 合計 / 2 ：{0, 0.5, 1, ..., 9} の19通り

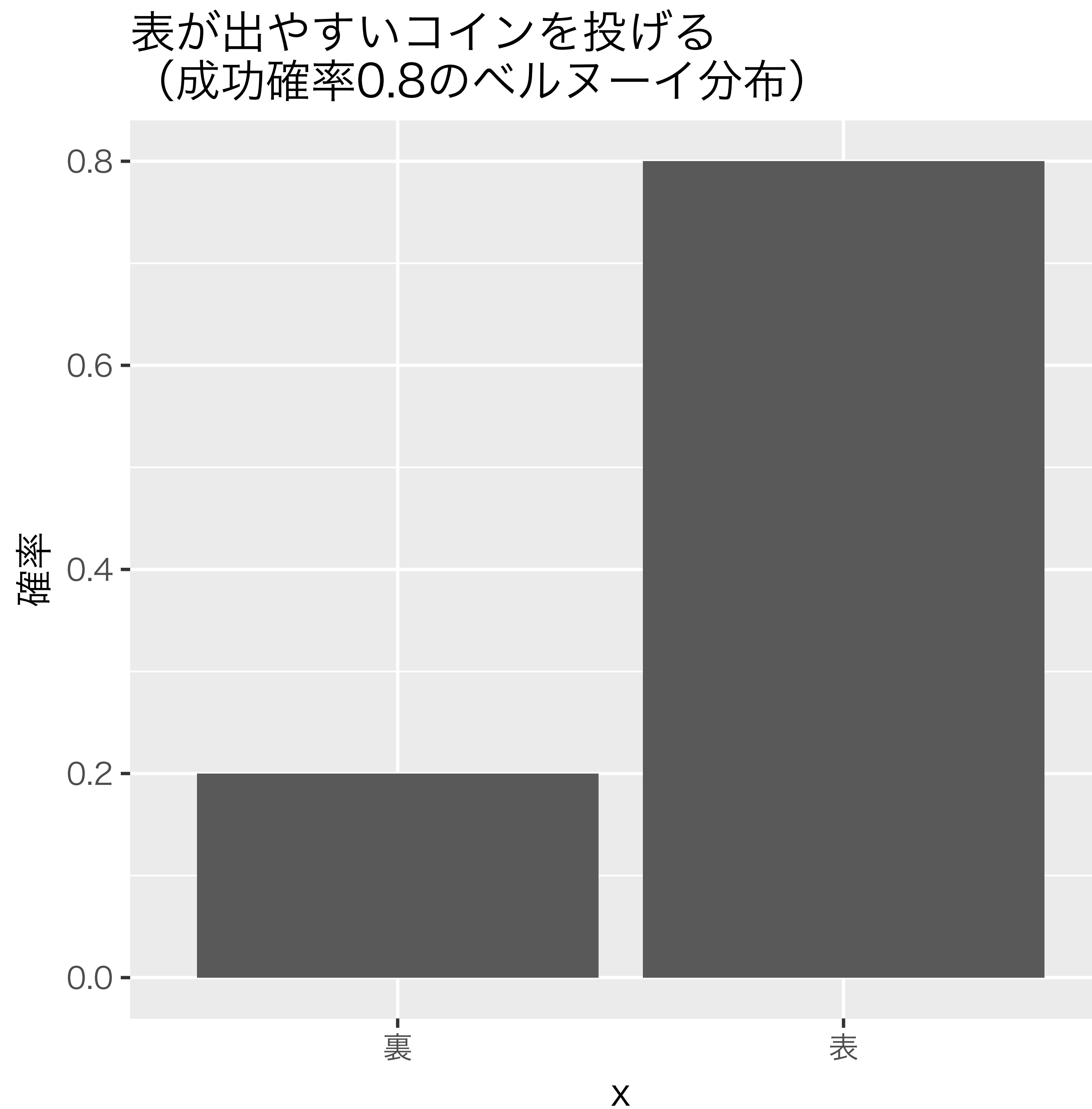
合計	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
平均	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9
確率	1/100	2/100	3/100	4/100	5/100	6/100	7/100	8/100	9/100	10/100	9/100	8/100	7/100	6/100	5/100	4/100	3/100	2/100	1/100

シミュレーション

- 「ボールを2個選んで平均値を求める」という作業を1000回繰り返してみる
- 平均値（推定値）の分布はどのような形になる？
- 1回ごとに選ぶ個数 (n) を増やすとどうなる？

ベルヌーイ分布

- コインを1回投げる
- 表が出る確率 p は0.8
- 裏が出る確率 は0.2



表が出る確率の推定

- 表が出る確率を知らないとする
- コインをn 回投げ、表が出た割合をpの推定値として使う

例：コインを2回投げる

- 1回目の結果：2通り
（表 or 裏）
- ２回目の結果：2通り
（表 or 裏）

➡ 選び方は全部で 2 x 2 = 4 通り

- 表が出る回数：{0, 1, 2} の3通り
- 割合 = 表の回数 / 2 ：{0, 0.5, 1} の3通り

<table>
<tr><th>1投目</th><th>裏</th><th>裏</th><th>表</th><th>表</th></tr>
<tr><th>2投目</th><th>裏</th><th>表</th><th>裏</th><th>表</th></tr>
<tr><td>表の回数</td><td>0</td><td colspan="2">1</td><td>2</td></tr>
<tr><td>平均</td><td>0</td><td colspan="2">0.5</td><td>1</td></tr>
<tr><td>確率</td><td>0.2x0.2
=0.04</td><td colspan="2">0.2x0.8+0.8x0.2
=0.32</td><td>0.8x0.8
=0.64</td></tr>
</table>

シミュレーション（追加課題）

- 「コインを2回投げて表の割合を求める」という作業を1000回繰り返してみる
- 平均値（推定値）の分布はどのような形になる？
- 1回ごとに投げる回数 (n) を増やすとどうなる？

今日のまとめ

- Rを使うと、様々な方法で乱数を生成することができる
  - 確率・統計分布の理解に役立つ
  - シミュレーションができる
- 中心極限定理のおかげで正規分布を使った推論ができる