高知工科大学 経済・マネジメント学群

統計学 2

6. シミュレーション

やない ゆうき
矢内 勇生

https://yukiyanai.github.io

yanai.yuki@kochi-tech.ac.jp

このトピックの目標

- 乱数生成の方法を理解する！
- for ループを理解する
- 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？

乱数を利用する

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()
```

for ループ

for ループとは？

- Rで特定の計算を繰り返し行う場合に用いる方法の1つ
- 長所
  - コードがわかりやすい
  - 入れ子にできる
- 短所
  - コードが長くなる
  - 実行速度が遅くなりがち

for ループの例

- 3行4列の行列Aの要素を順番に表示 (print) する
- for を使って、i行j列を順番に表示
  - まず、i を1に固定
    - j を 1, 2, 3, 4 と順番に動かす
  - 次に、i を2に固定
    - j を1, 2, 3, 4 と動かす
  - 最後に、i を 3 に固定
    - j を 1, 2, 3, 4 と動かす

```
A <- matrix(1:12, nrow = 3)


for (i in 1:3) {
  for (j in 1:4) {
    print(A[i, j])
  }
}
```

繰り返しの実行

- for ループ以外にも繰り返しを実現する方法はある
  - while ループ
  - apply, map などの関数（Rらしい関数）
    - 詳しくは、副読本 の「[Rプログラミングの基礎]」の章を参照

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

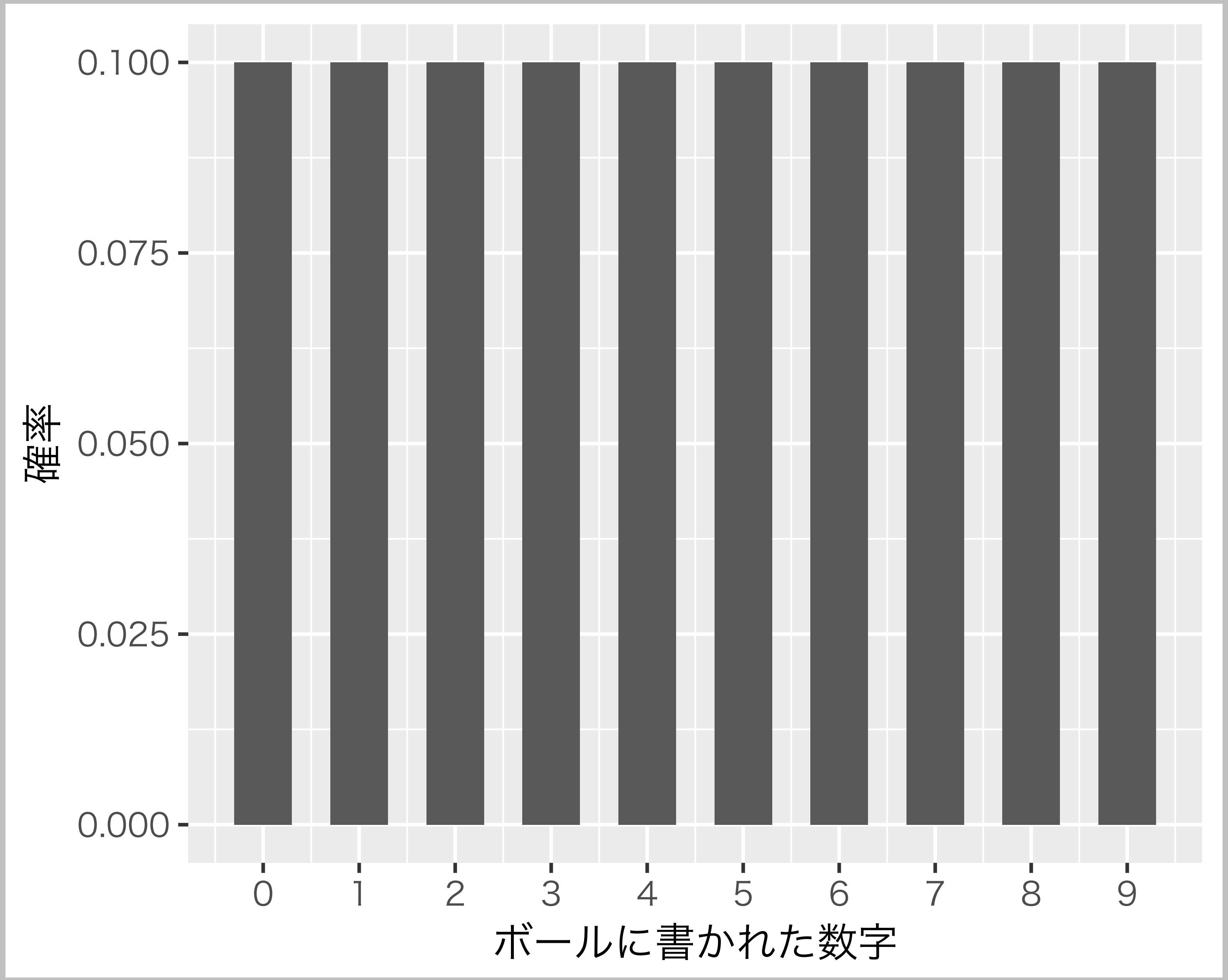
中心極限定理 (Central Limit Theorem; CLT)

- 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、正規分布を利用することができる

★ 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

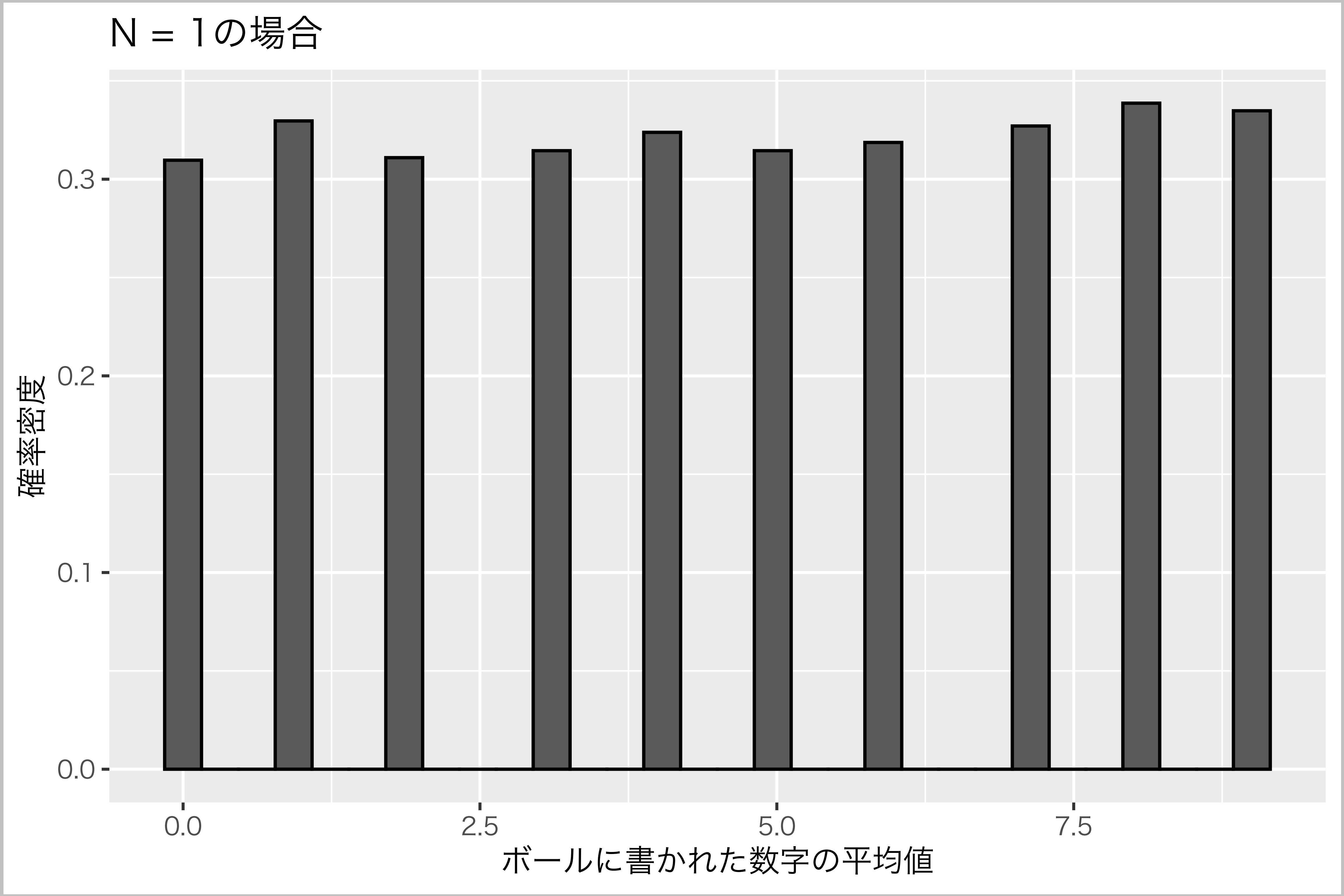
➡選び方は全部で $10 \times 10 = 100$ 通り

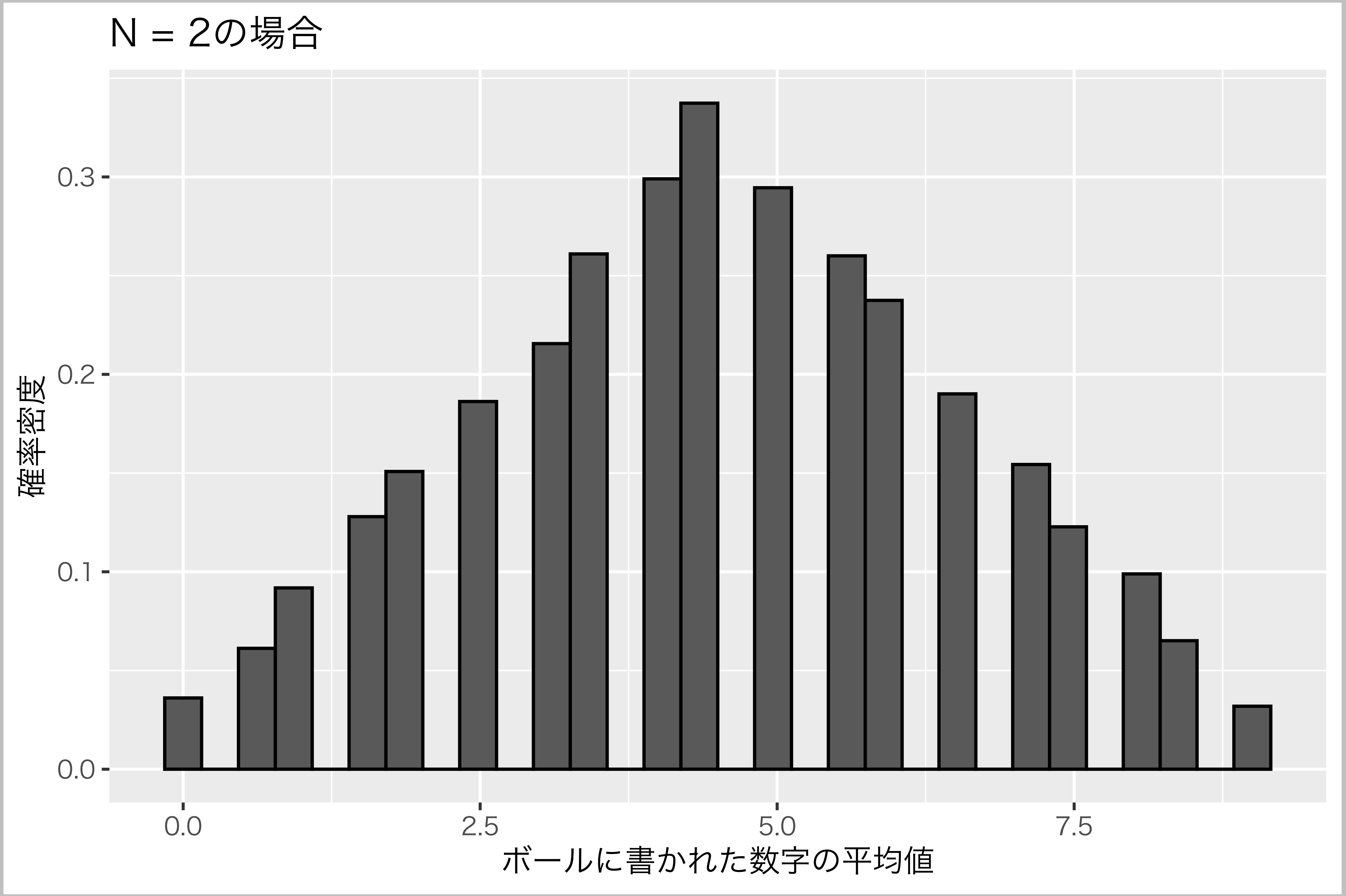
- 2個のボールの合計：0 から18までの19通り
- 平均 = 合計 / 2 ：{0, 0.5, 1, ..., 9} の19通り

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

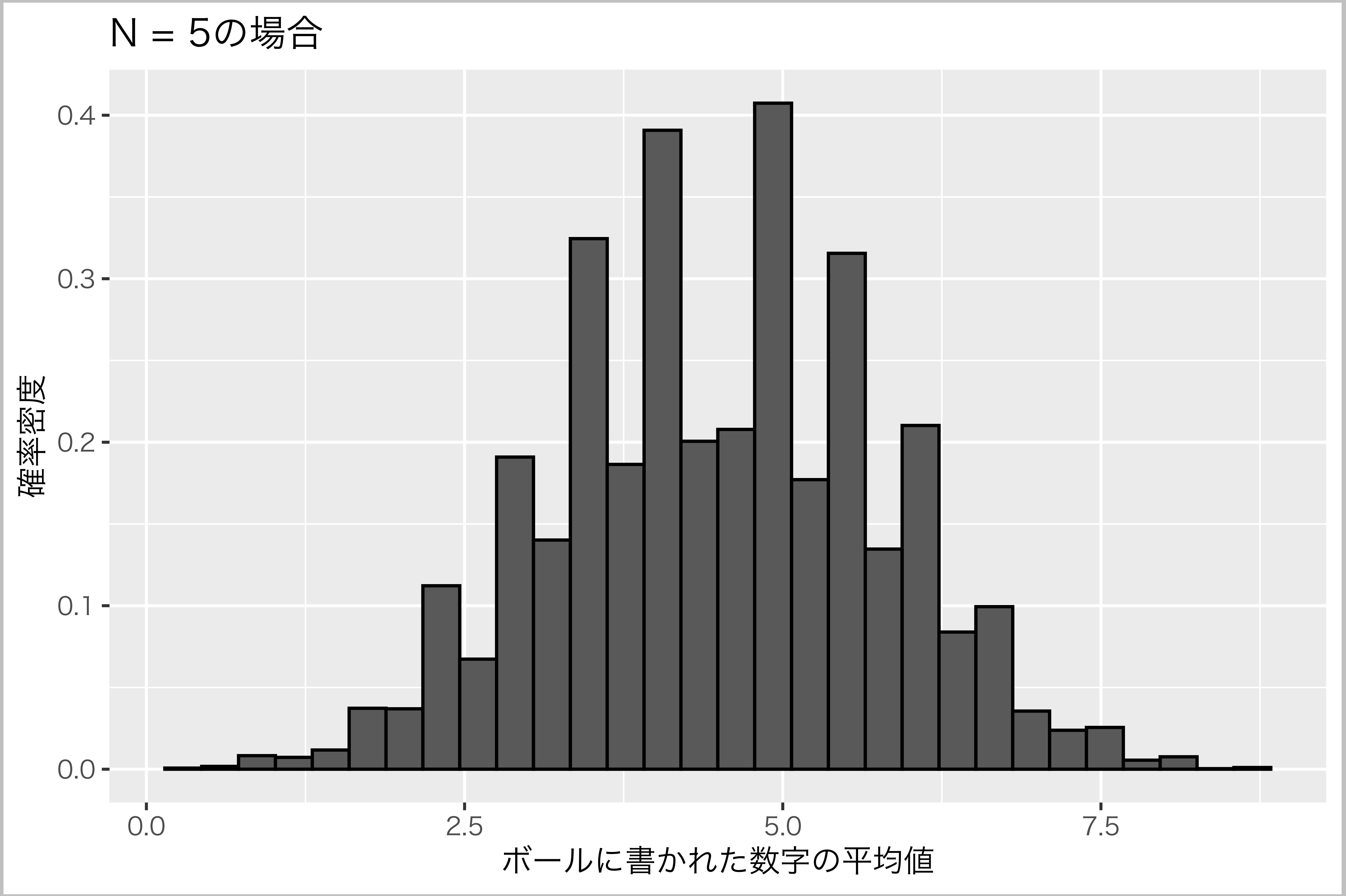
シミュレーション

- 「ボールを N 個選んで平均値を求める」という作業を 10,000 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？

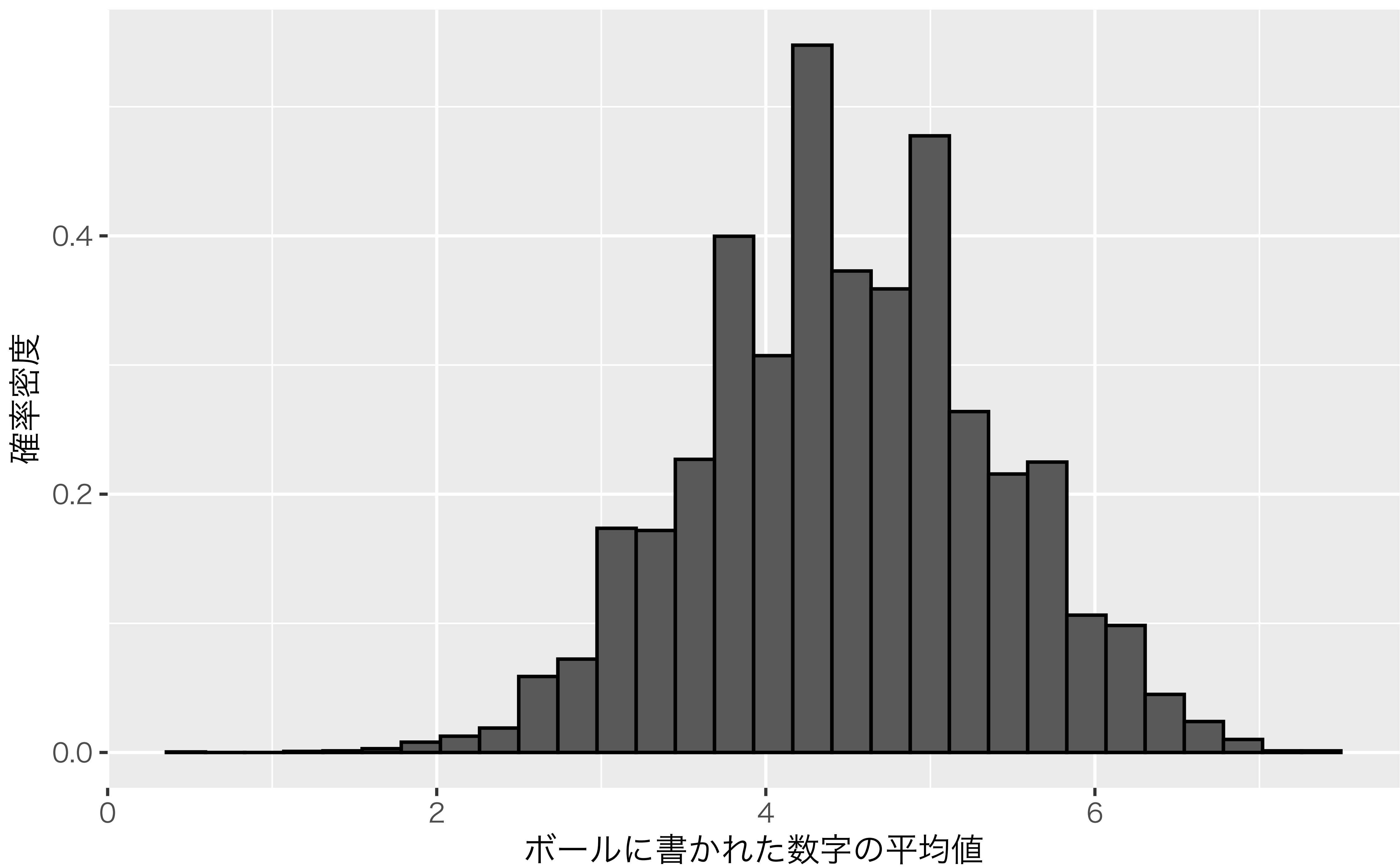

N = 1の場合
0.3
0.2
0.1
0.0
確率密度
0.0
2.5
5.0
7.5
ボールに書かれた数字の平均値

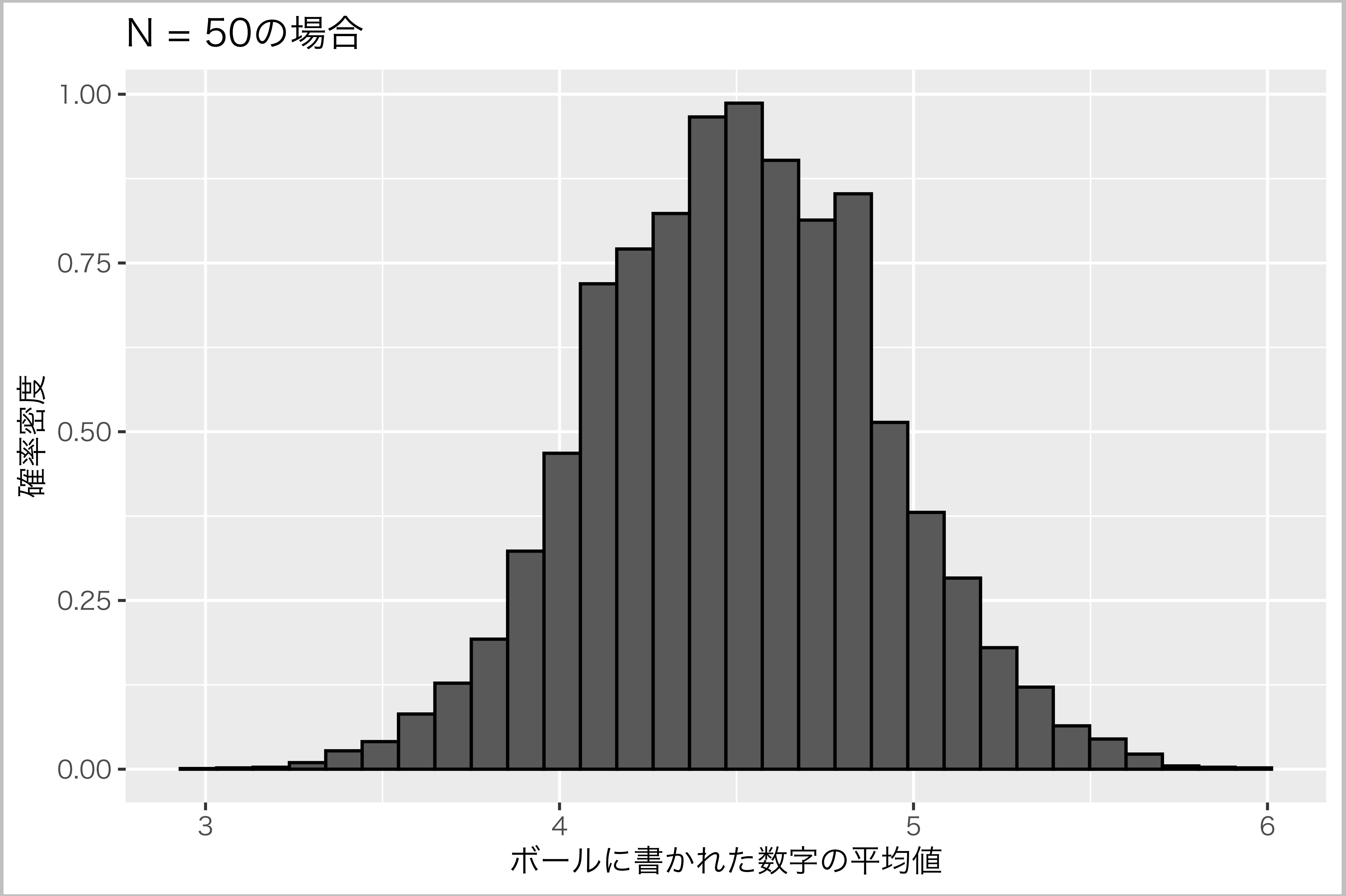


N = 2の場合
確率密度
0.3
0.2
0.1
0.0
0.0
2.5
5.0
7.5
ボールに書かれた数字の平均値

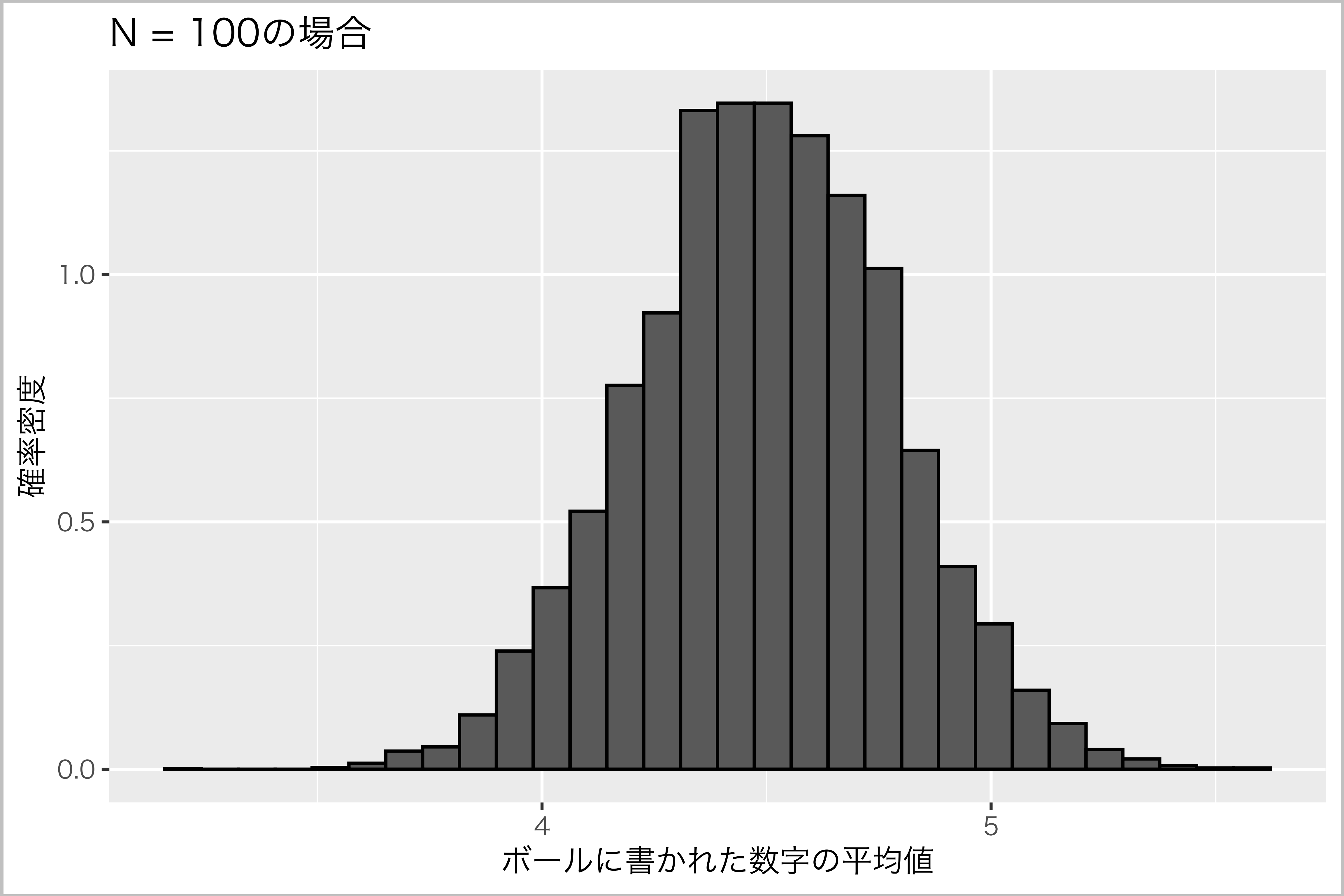

N = 5の場合
0.4
0.3
0.2
0.1
0.0
確率密度
0.0
2.5
5.0
7.5
ボールに書かれた数字の平均値

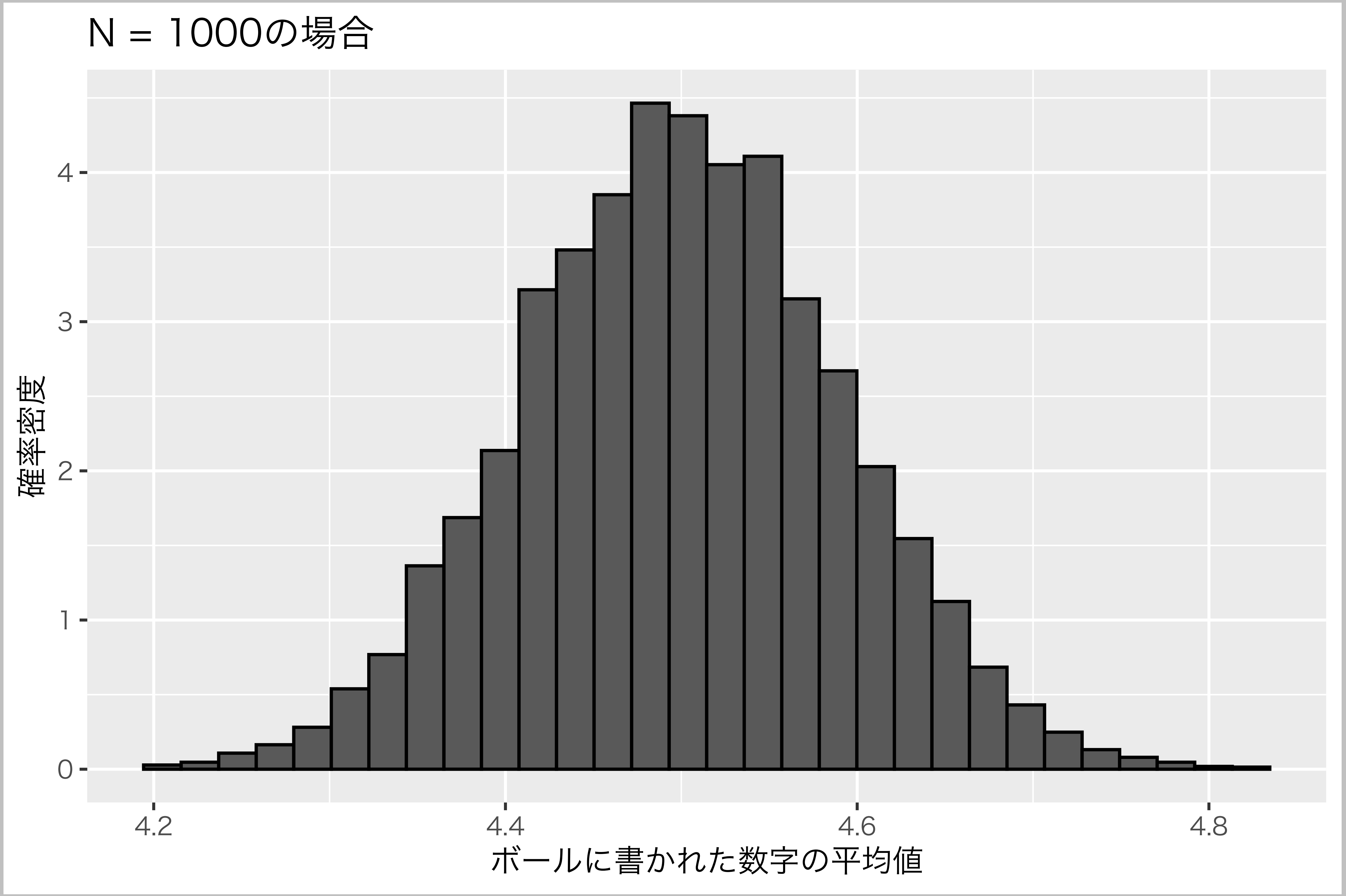
N = 10の場合





N = 50の場合
1.00
0.75
0.50
0.25
0.00
確率密度
3
4
5
6
ボールに書かれた数字の平均値


N = 100の場合
確率密度
1.0
0.5
0.0
4
5
ボールに書かれた数字の平均値


N = 1000の場合
確率密度
4
3
2
1
0
4.2
4.4
4.6
4.8
ボールに書かれた数字の平均値

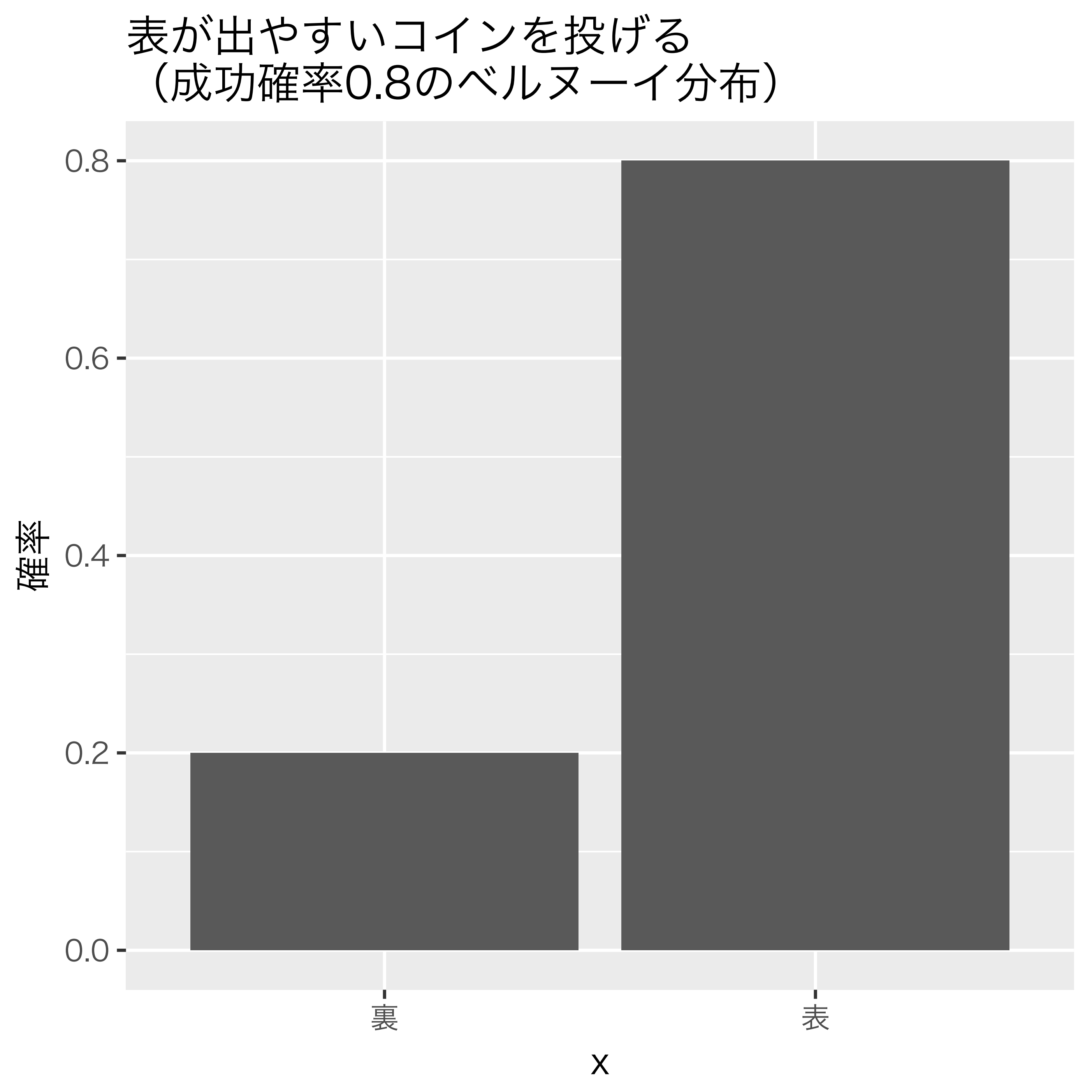
ベルヌーイ分布

- コインを1回投げる
- 表が出る確率 p は、

  $p = 0.8$

- 裏が出る確率 $1 - p$ は

  $1 - p = 0.2$



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 $2 \times 2 = 4$ 通り

- 表が出る回数：{0, 1, 2} の3通り

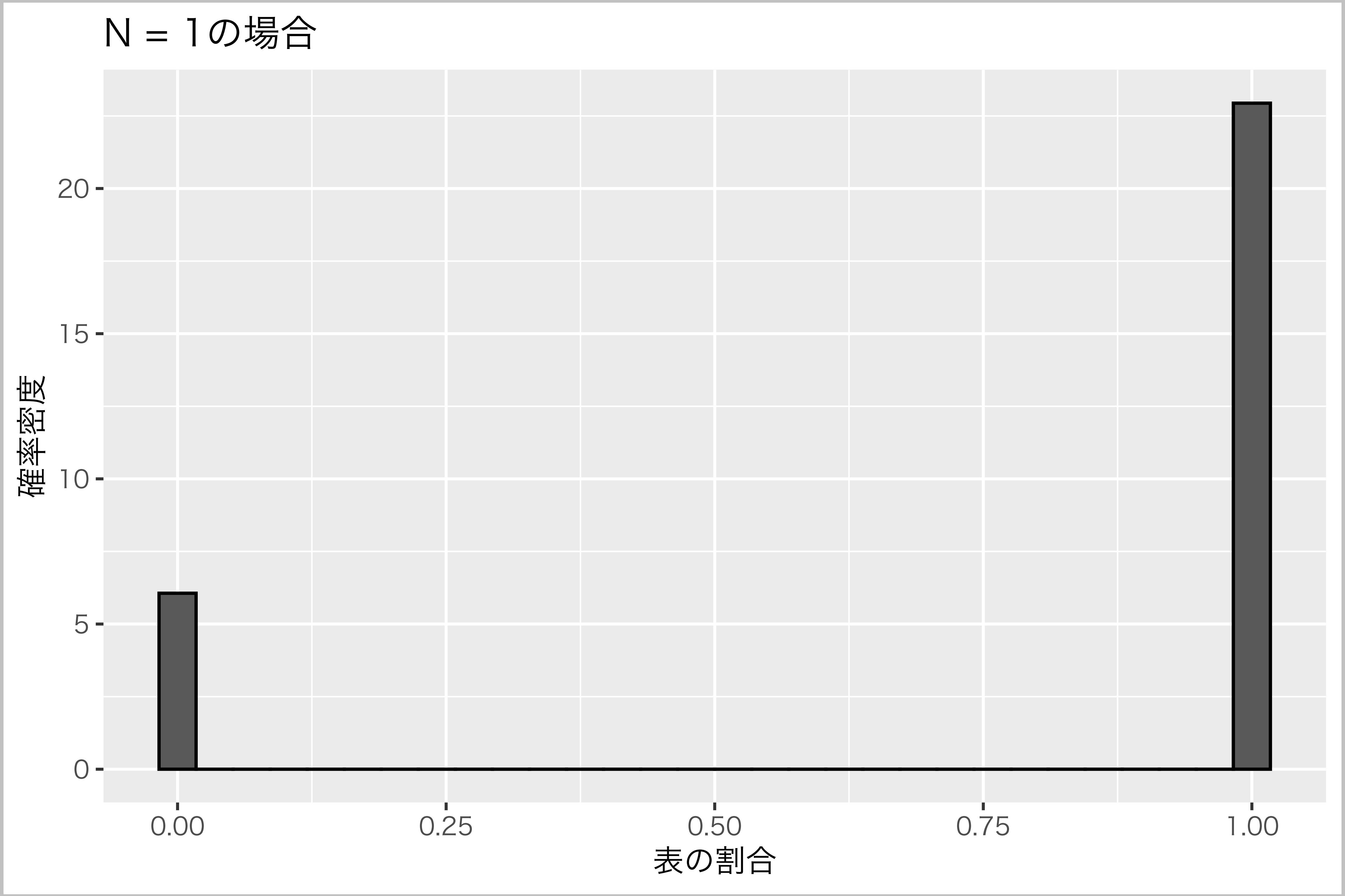
- 割合 = 「表の回数 / 2」 ： {0, 0.5, 1} の3通り

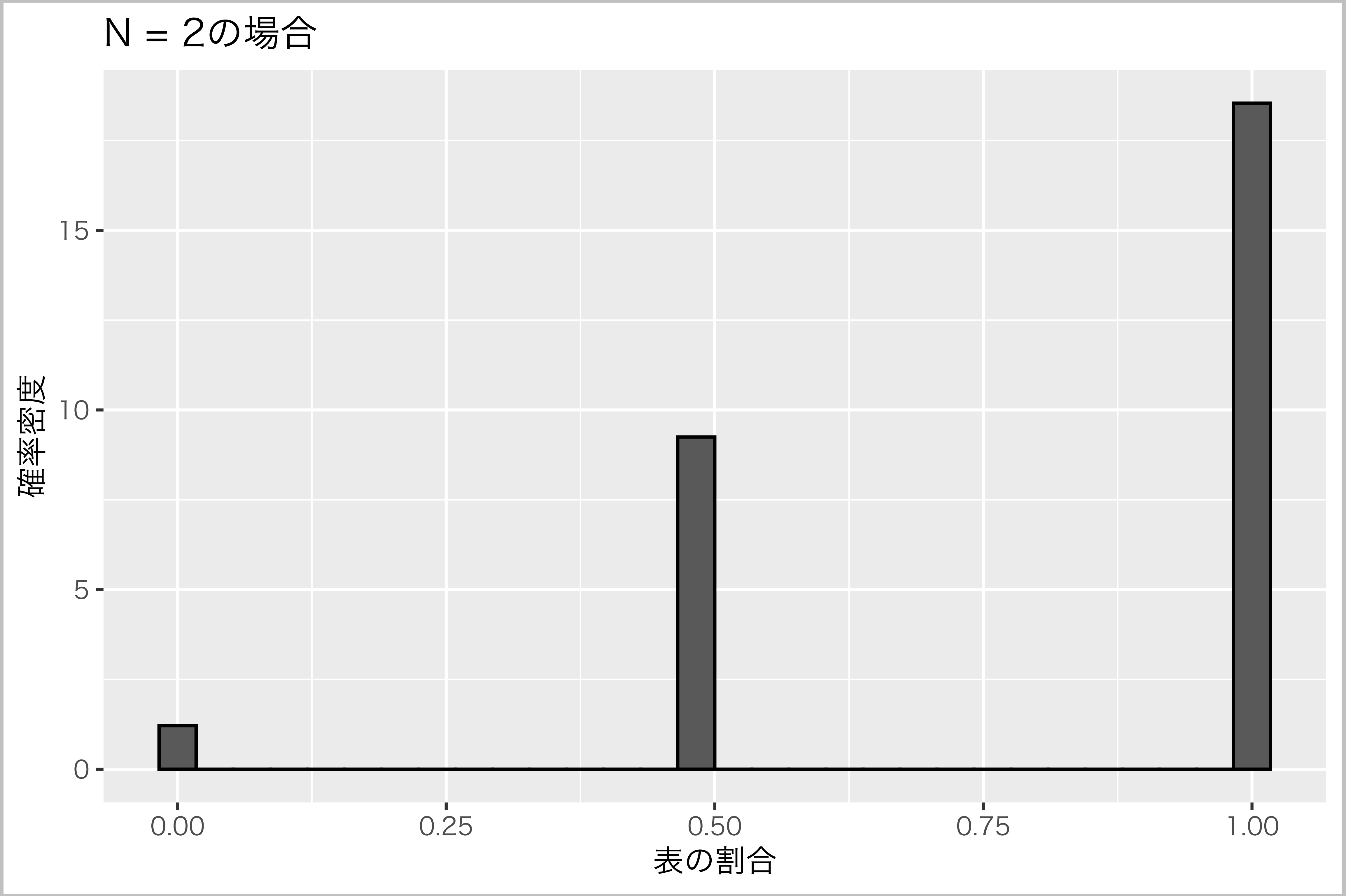
表が出る確率 $p = 0.8$

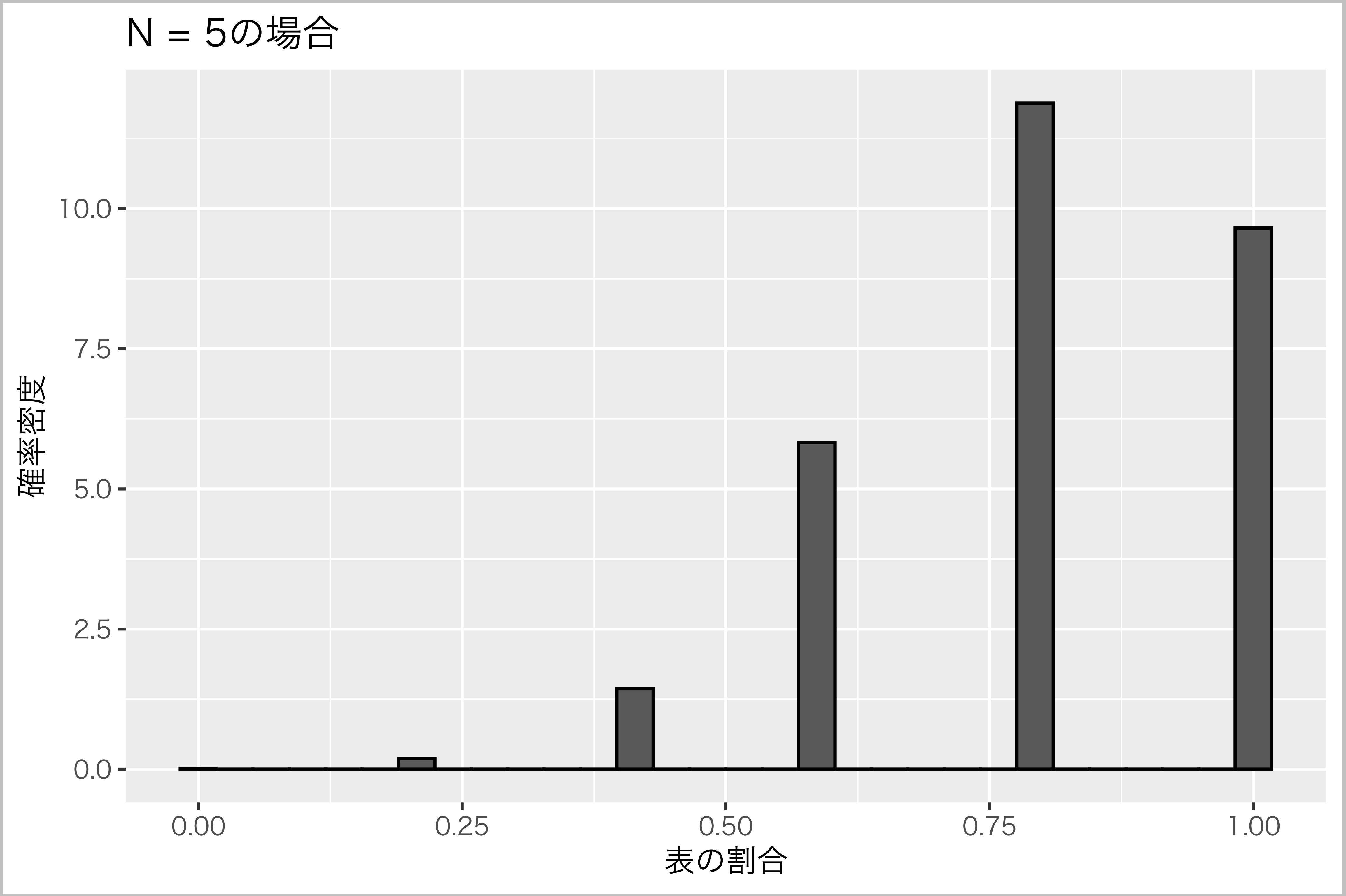
1投目	裏	裏	表	表
2投目	裏	表	裏	表
表の回数	0	1		2
平均	0	0.5		1
確率	0.2x0.2 =0.04	0.2x0.8+0.8x0.2 =0.32		0.8x0.8 =0.64

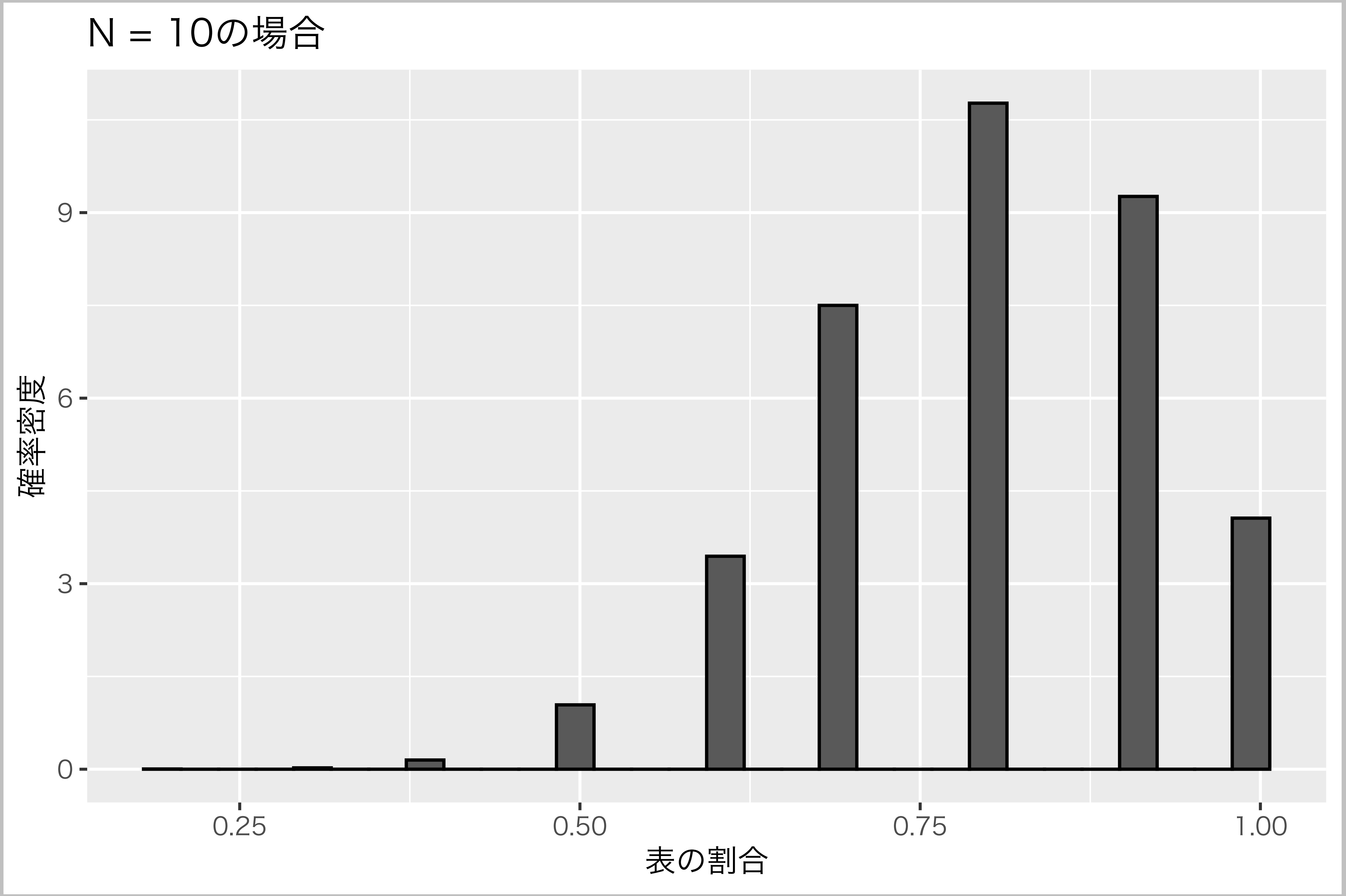
シミュレーション

- 「コインをN 回投げて表の割合を求める」という作業を10,000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？

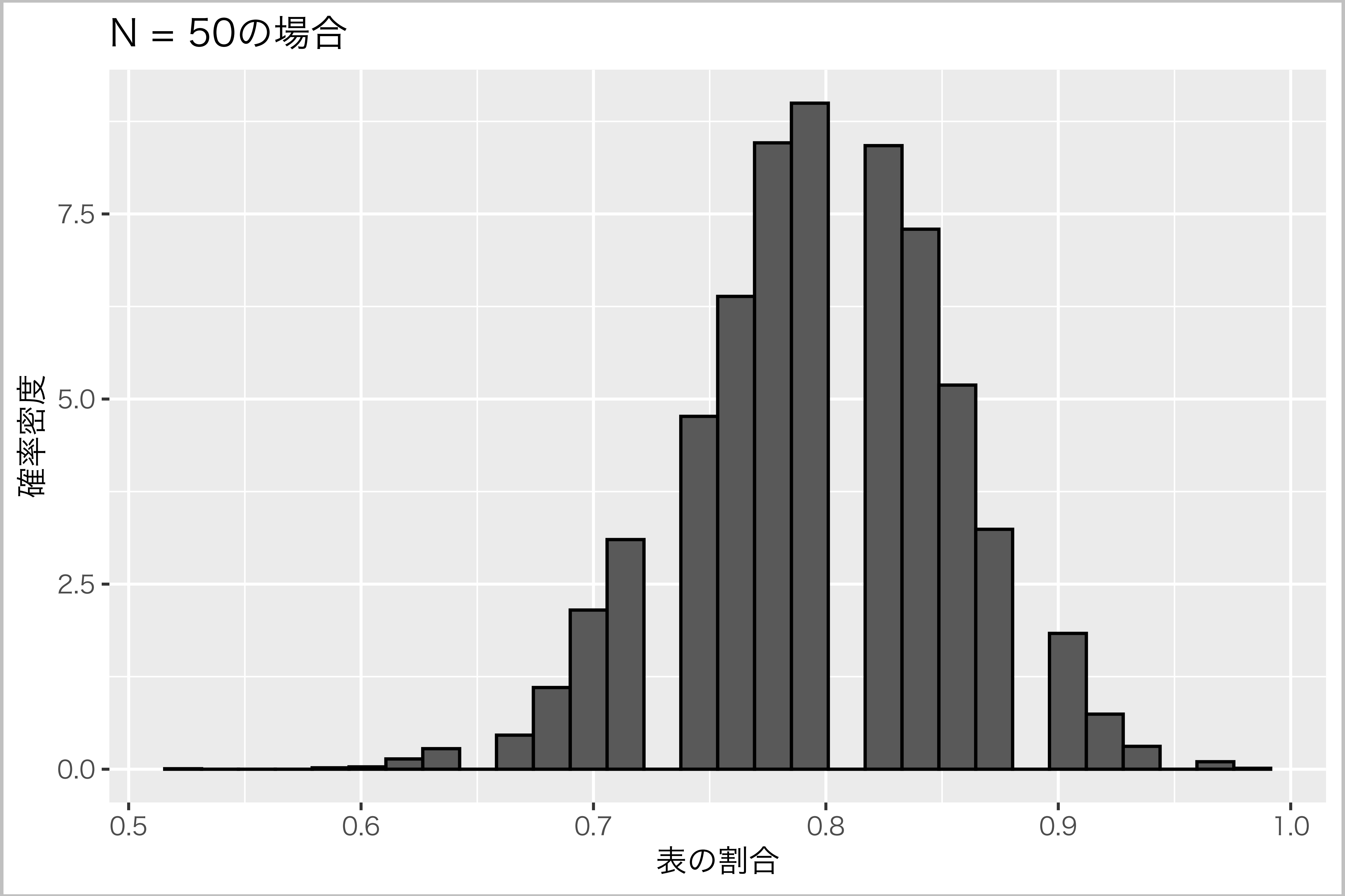

N = 1の場合
確率密度
20
15
10
5
0
0.00
0.25
0.50
0.75
1.00
表の割合

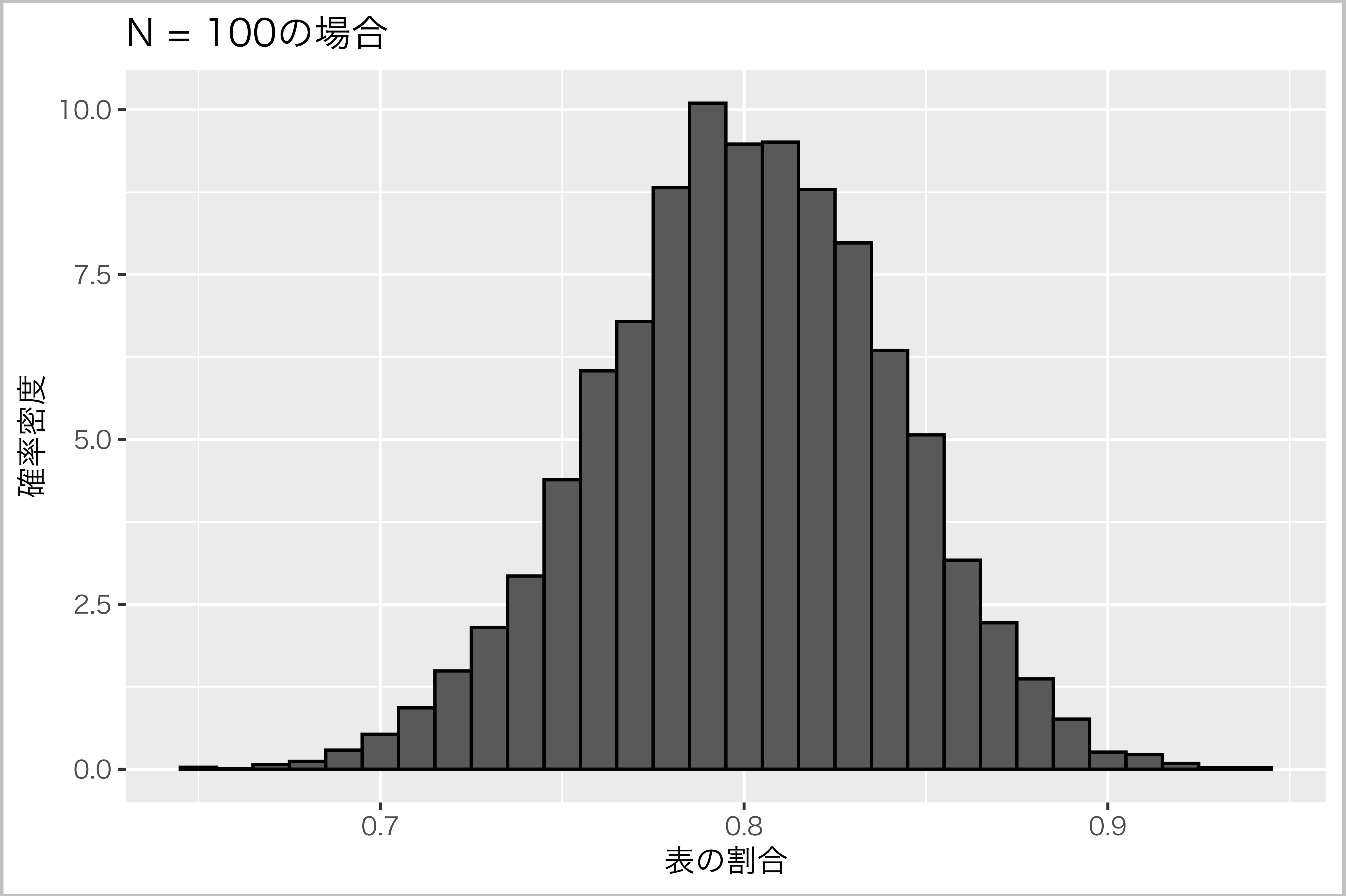

N = 2の場合
確率密度
15
10
5
0
0.00
0.25
0.50
0.75
1.00
表の割合

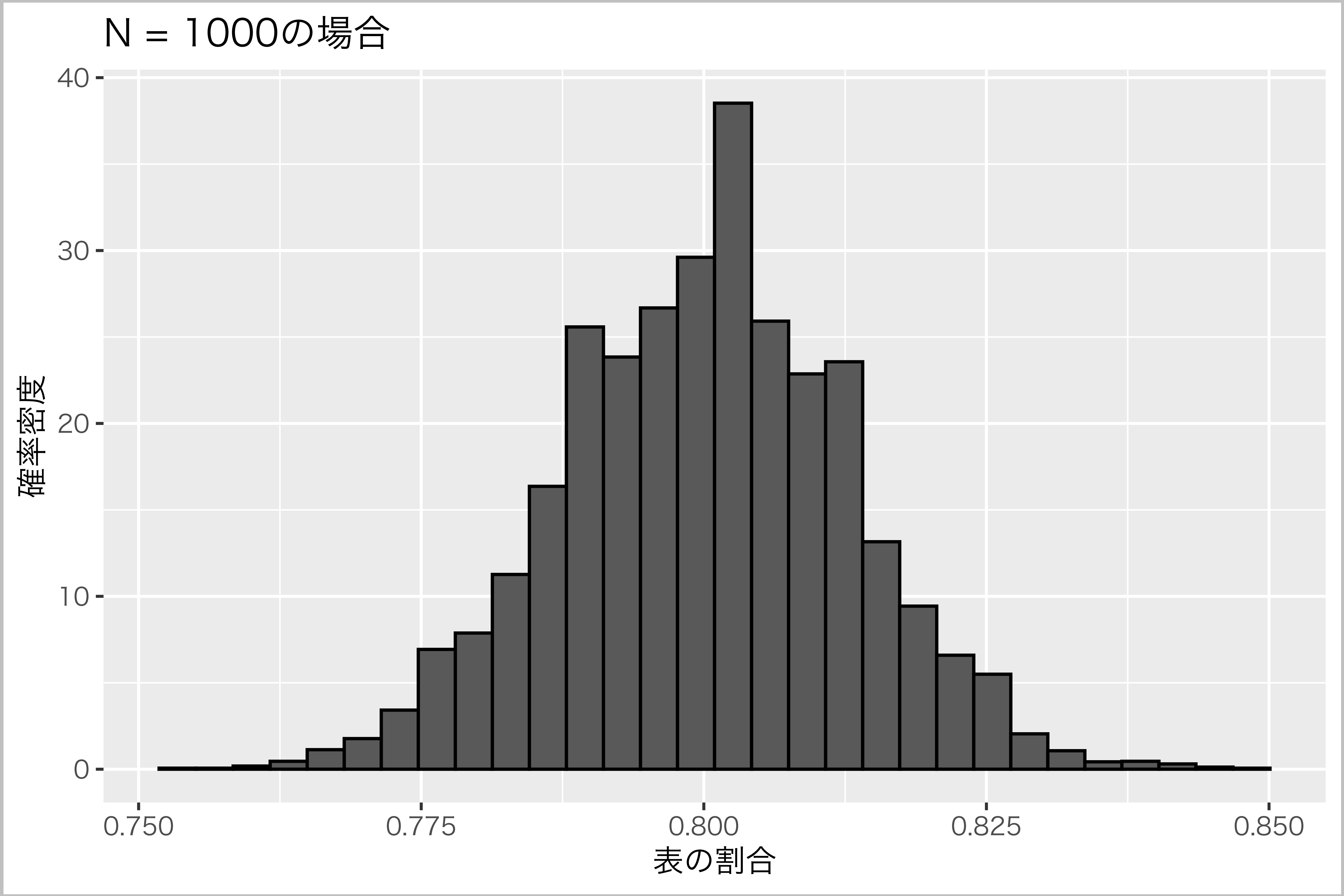

N = 5の場合
10.0
7.5
5.0
2.5
0.0
確率密度
0.00
0.25
0.50
0.75
1.00
表の割合



N = 10の場合
確率密度
0
3
6
9
0.25
0.50
0.75
1.00
表の割合


N = 50の場合
確率密度
7.5
5.0
2.5
0.0
0.5
0.6
0.7
0.8
0.9
1.0
表の割合


N = 100の場合
10.0
7.5
5.0
2.5
0.0
確率密度
0.7
0.8
0.9
表の割合


N = 1000の場合
40
30
20
10
0
確率密度
0.750
0.775
0.800
0.825
0.850
表の割合

このトピックのまとめ

- 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

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
    - https://yukiyanai.github.io/jp/classes/stat2/contents/R/rng-n-clt.html

次回予告

7. 統計的推定と
仮説検定の基礎