

統計学入門

第13項, 第14項

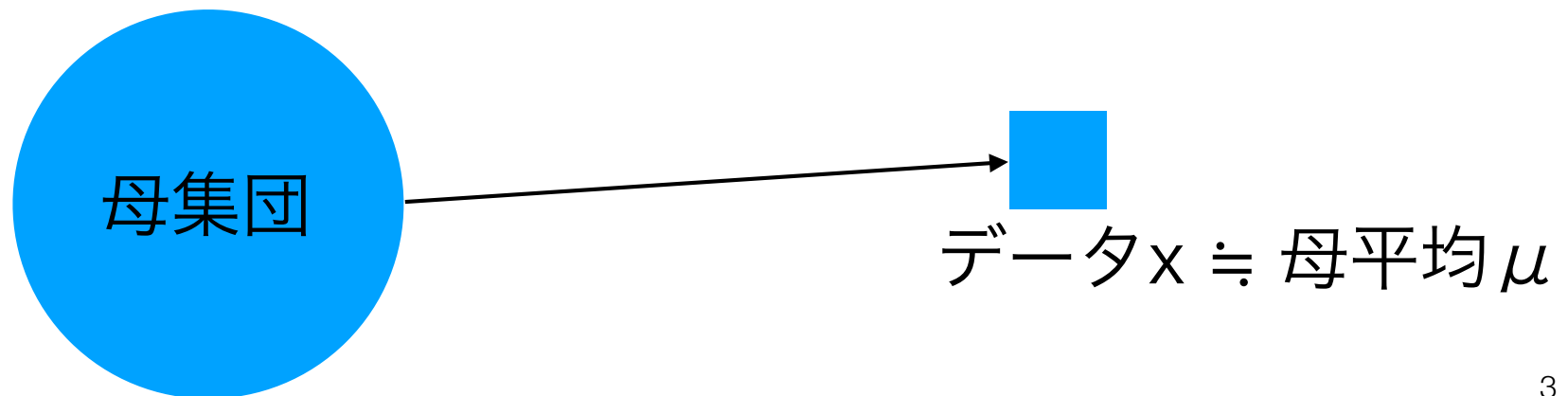
発表者：富田 恭平

第13項

標本平均の考え方

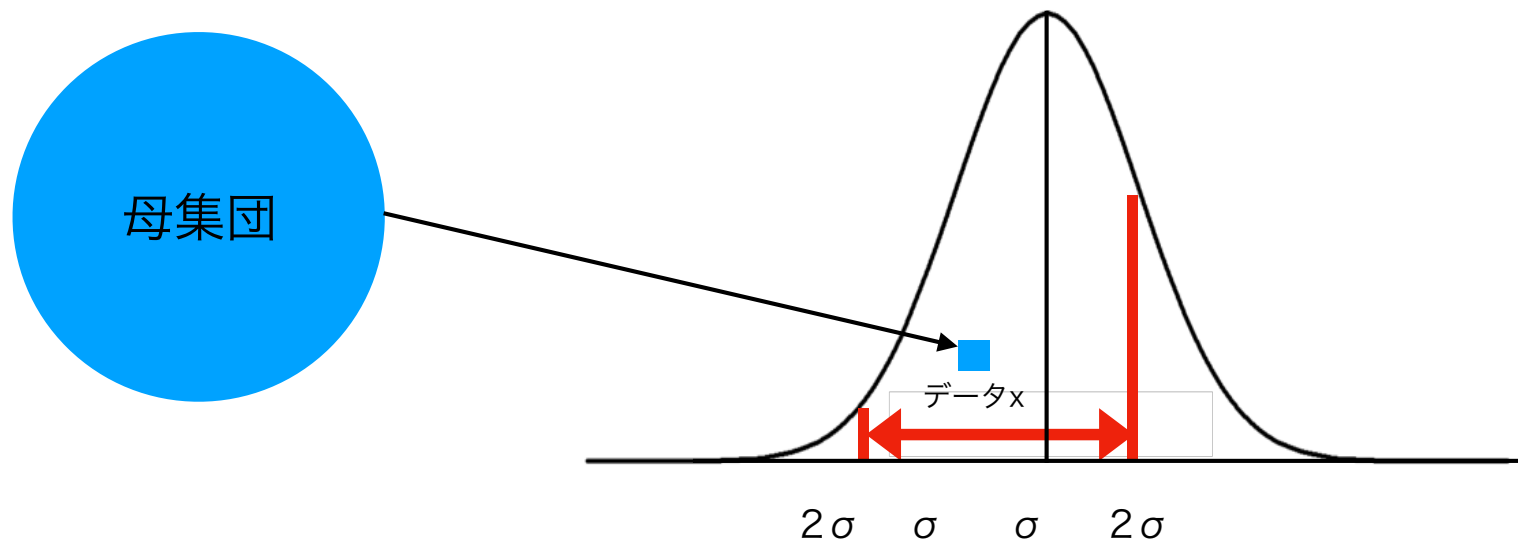
1 個のデータと母集団の関係

- 不確実現象の源である「母集団」について知りたい
- 1 個のデータ x が観測された場合,
母集団についてどのようなことがわかるのか？
- 母平均 μ はデータ x の近くにあるだろう



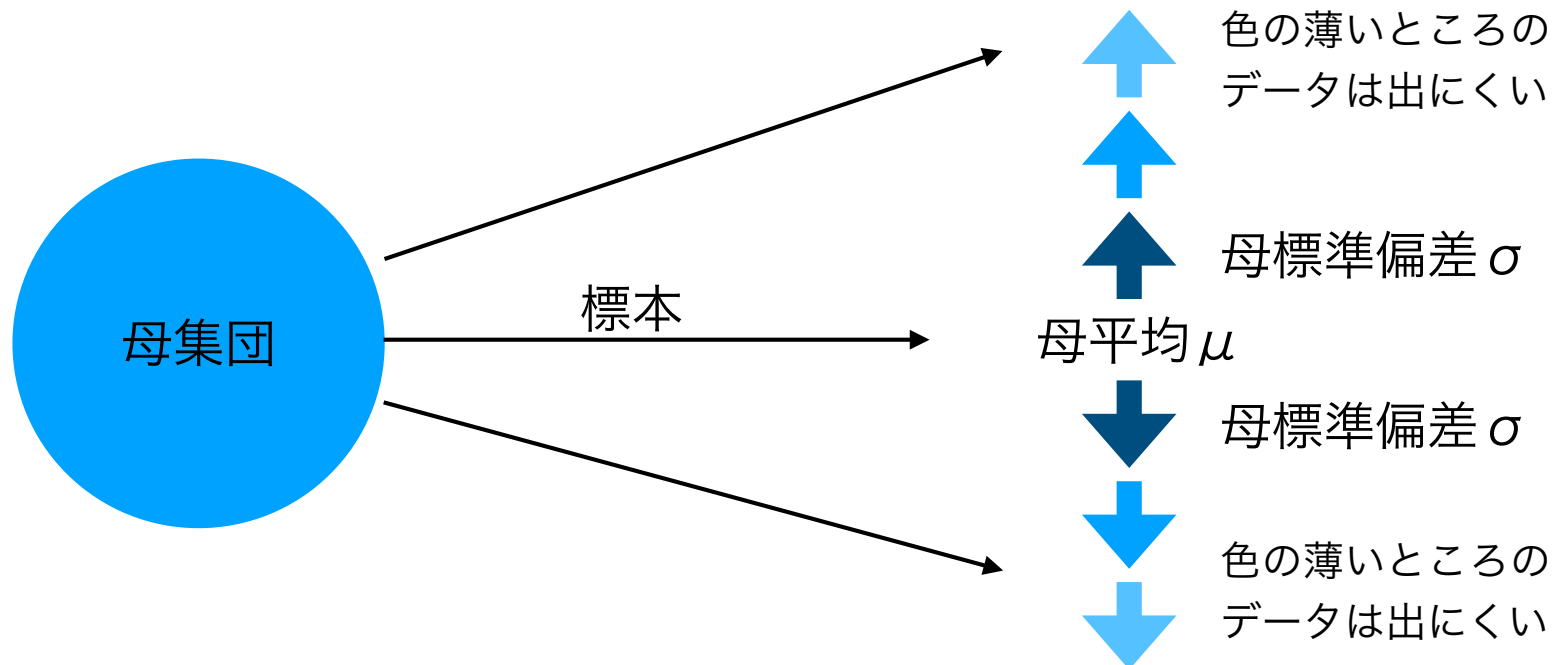
1 個のデータと母集団の関係

- 母集団が正規分布に従い,
母標準偏差 σ がわかっている場合はさらに詳しくわかる
- x から $\sigma * 2$ 程度以内の隔たりで μ が存在するだろう**



母標本偏差が既知のときの推定

- 一般の分布に対しても、 σ^2 を σ^k とすれば推定できる
- チェビシェフの不等式
 μ から σ^k 以上離れるデータは $1/k^2$ の比率しかない

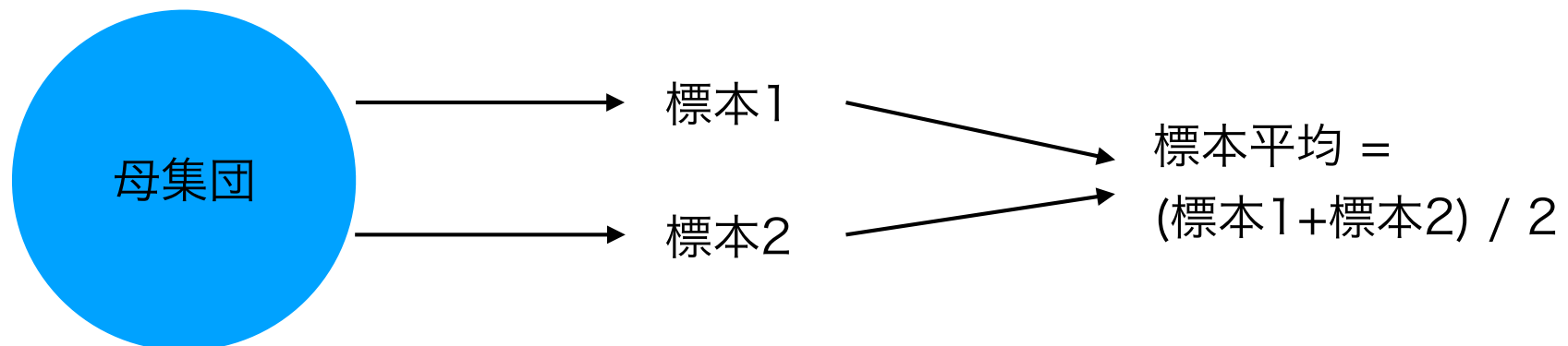


複数の観測データと母集団の関係

複数のデータがあれば, 母平均 μ の推量の精度は上がる

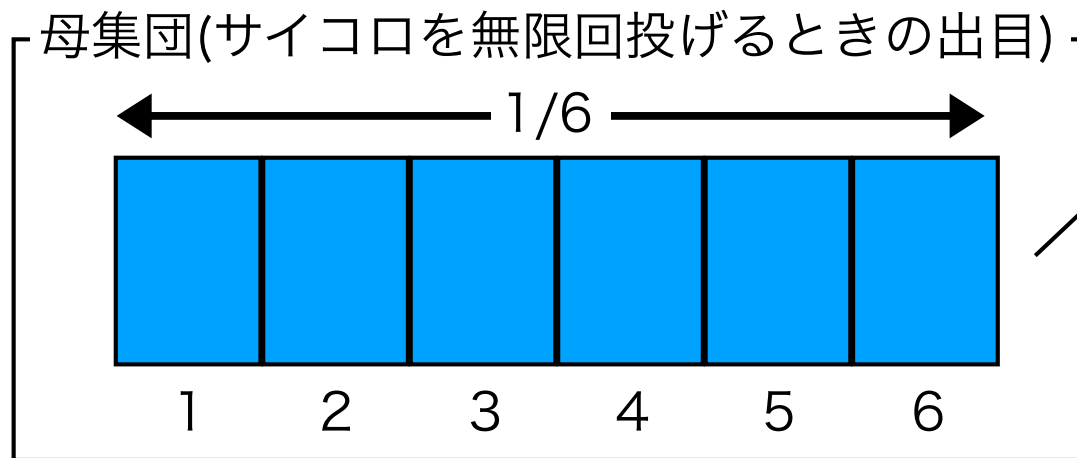
複数の観測データの平均を標本平均という

標本平均 = (観測されたデータの合計) / (観測データ数)



標本平均と統計的推定

サイコロを投げて出た目を無限回記録した母集団



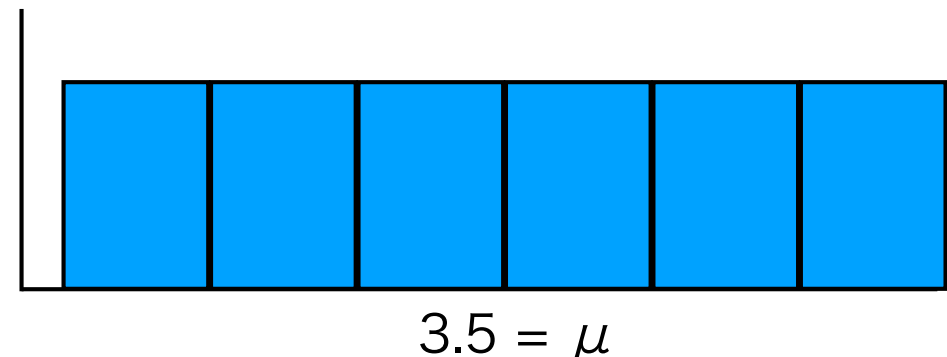
母平均 $\mu = 3.5$

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

標本1

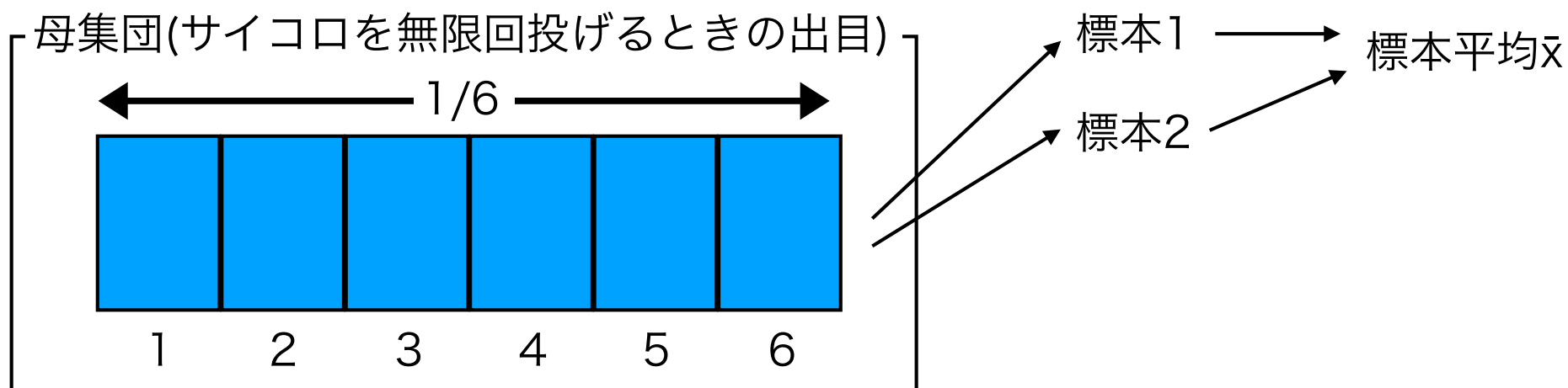
標本1の分布は母集団と同じになる

何回もデータを取り出してヒストグラムを作ると以下となる



標本平均と統計的推定

サイコロを2個投げたときの記録の標本平均



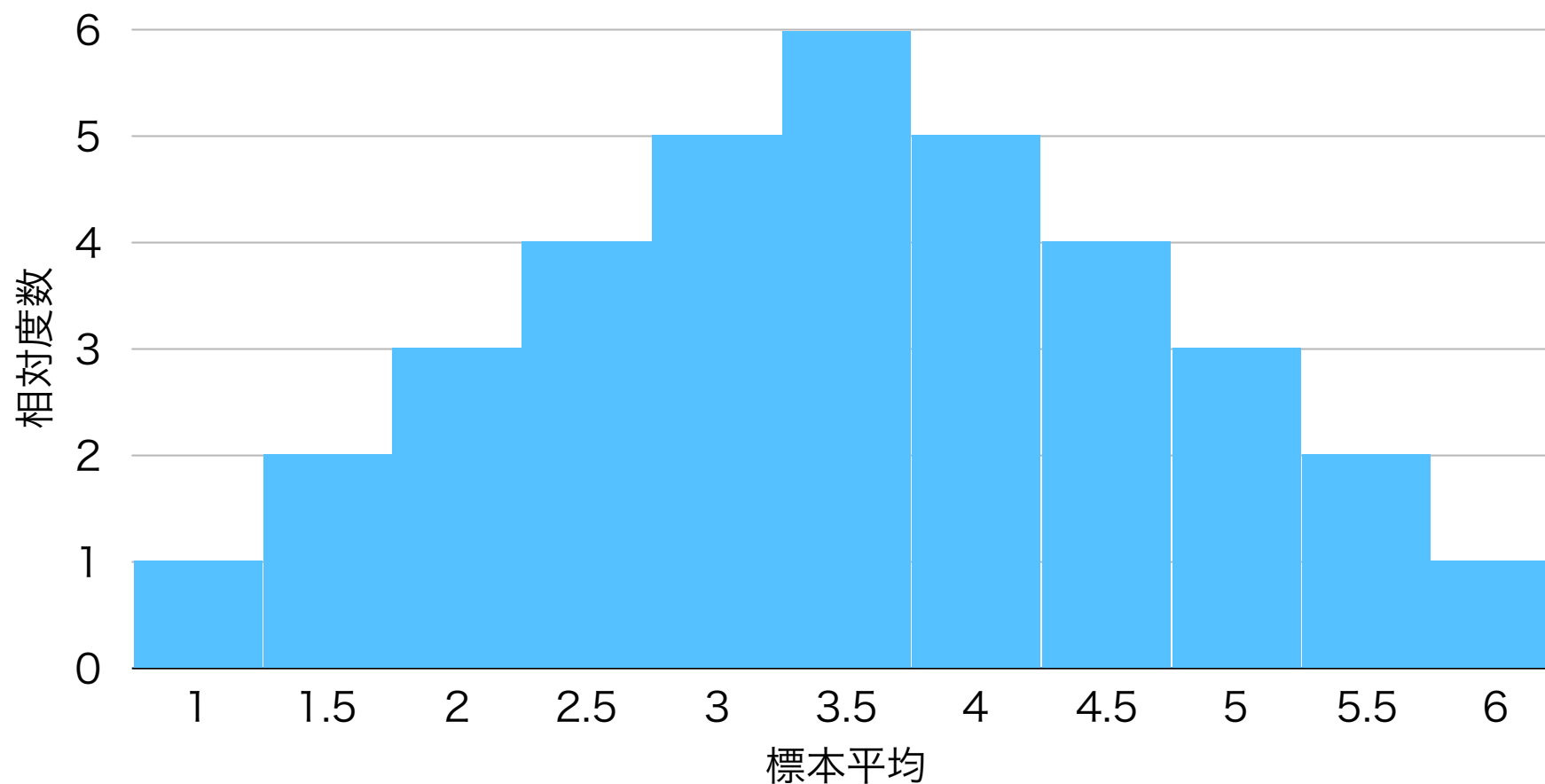
	1	2	3	4	5	6
1	1	1.5	2	2.5	3	3.5
2	1.5	2	2.5	3	3.5	4
3	2	2.5	3	3.5	4	4.5
4	2.5	3	3.5	4	4.5	5
5	3	3.5	4	4.5	5	5.5
6	3.5	4	4.5	5	5.5	6

$\bar{x}$ は表の数値が
その回数に比例して出てくる

たとえば1.5が2/36に対して,
3は5/36出てくる

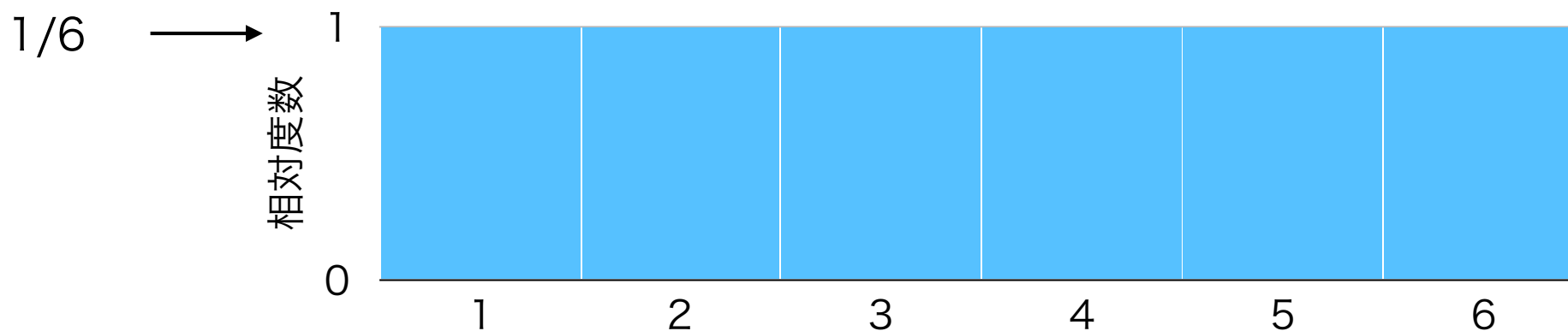
標本平均と統計的推定

サイコロを2個投げたときの標本平均



標本平均と統計的推定

サイコロ投げで出るデータの母集団の分布

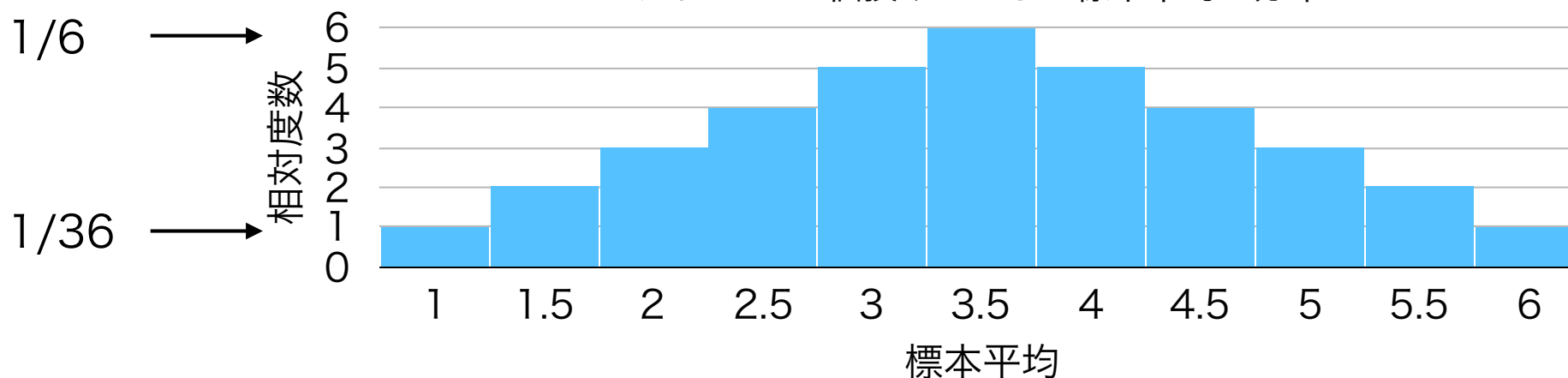


標本平均



母平均の近くはかなり凝縮する

サイコロを2個投げたときの標本平均の分布



複数の観測データと母集団の関係

大数の法則

1つの母集団から, n 個のデータを観測し

その標本平均を $\bar{x}$ を作る

このとき, n が大きければ大きいほど,

標本平均は母平均 μ に近い数値を取る可能性が高くなる

第13項 まとめ

- ・ 観測されたデータは, ある程度母平均に近いものだと考えられる
- ・ 複数のデータを観測し, その平均をとったものを**標本平均 $\bar{x}$** という
- ・ 複数のデータを観測して標本平均を取ると, それは1個のデータより母平均に近い値である.

観測データを増やせば増やすほど, 標本平均は母平均に近くなる

- ・ **大数の法則**

**1つの母集団から n 個のデータを観測しその標本平均を作るとき,
 n が大きければ大きいほど標本平均は母平均 μ に近い数値を取る**

第14項

正規母集団と標本平均

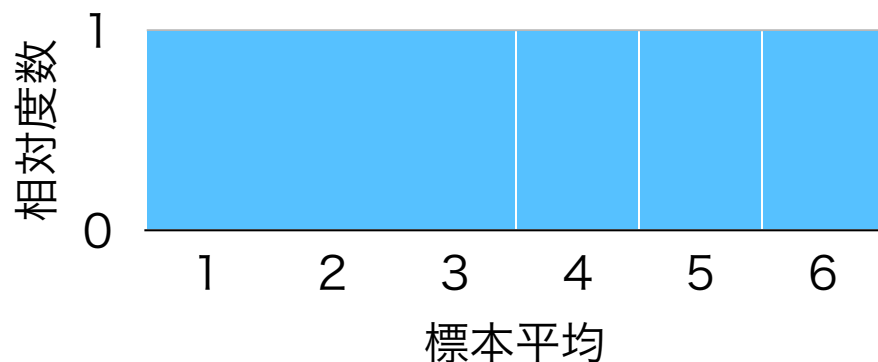
標本平均で区間推定する

- 区間推定に標本平均 $\bar{x}$ を利用すれば、精度の高い推定が可能になる
- しかし、区間推定において、95%信頼区間のように**間違えるリスク**を明確にしなければならない
- 一般の母集団では、母集団全体の分布がわかっているとしても標本平均の分布とは違うため、**間違えるリスク**を一定水準に保てず、推定に使えない

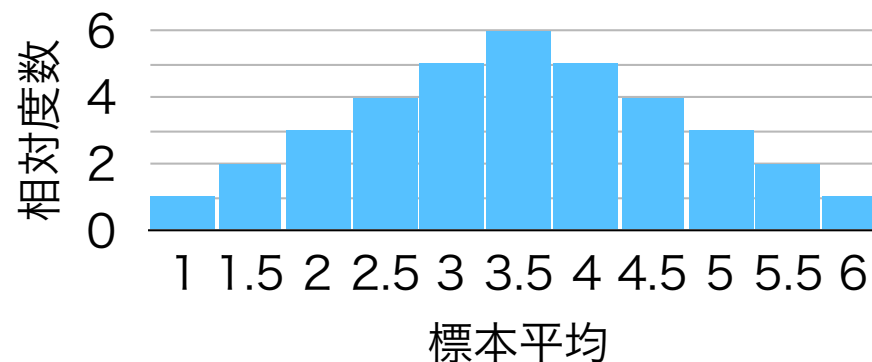
一般の母集団とその標本平均の分布

- サイコロ投げで出る目のデータの母集団は**一様分布**
- 2個の観測データから作る標本平均 $\bar{x}$ の分布は**母集団と違う分布**
- つまり、サイコロ投げの母集団では、
**標本平均の分布が変化していくため、
間違うリスクを一定水準に保つのは難しい**

サイコロ投げで出るデータの母集団の分布



2個の観測データの標本平均の分布



正規分布と正規母集団

- 母集団が正規分布しているものを**正規母集団**という
- 正規母集団では、**標本平均の分布は正規分布のまま**

正規母集団からの標本平均の性質

正規母集団の母平均を μ 、母標準偏差を σ とすると、観測されるデータ x の n 個に対する**標本平均 $\bar{x}$ の分布は正規分布**となる

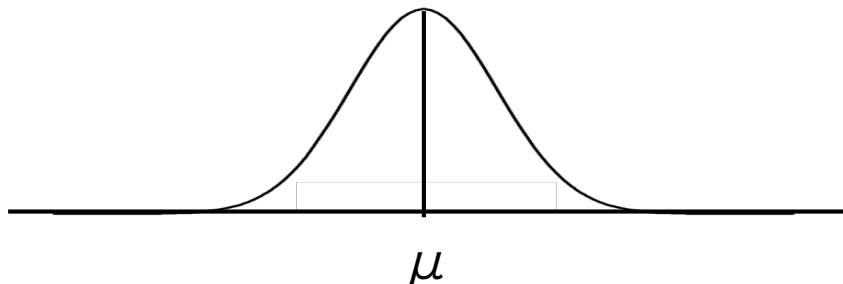
$\bar{x}$ の分布の平均値は μ のままだが、**標準偏差は $\sigma / \sqrt{n}$ となり、母集団にくらべて $\sqrt{n}$ 分の1に縮む**

正規母集団の標本平均の分布

正規母集団の母平均を μ 、母標準偏差を σ とすると、観測されるデータ x の n 個に対する **標本平均 $\bar{x}$ の分布は正規分布となる**

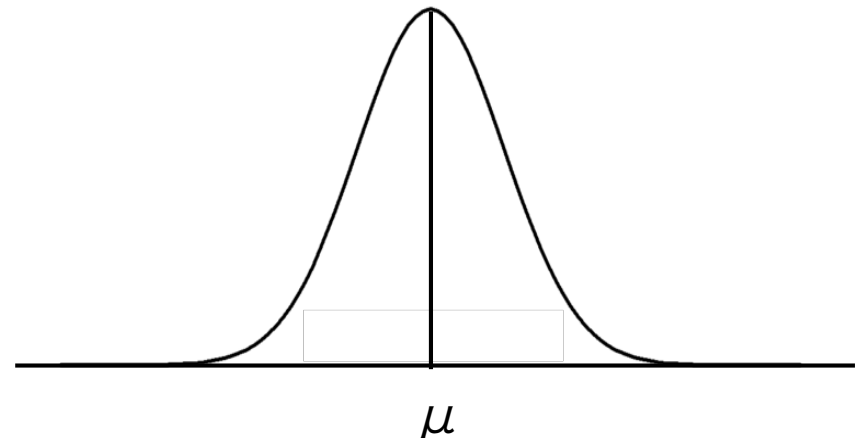
$\bar{x}$ の分布の **平均値は μ のまま**だが、**標準偏差は $\sigma / \sqrt{n}$ となり、母集団にくらべて $\sqrt{n}$ 分の1に縮む**

母集団の分布



標本平均の分布

(標本平均を集めて、それを母集団とみなしたものの分布)



正規母集団の標本平均の分布

- 「**n個**に対する**標本平均 $\bar{x}$** の分布について、
その**平均値が母平均 μ** と一致し、
標準偏差が母標準偏差 σ を $\sqrt{n}$ で割ったものになる」という
性質はどのような分布でも成立する
- ただし、
「**標本平均の分布が母集団の分布と同じ形のままである**」と
いう性質は**正規分布と一部特例を省けばほとんど成立しない**

標本平均と95%予言的中区間

一般正規分布の95%予言的中区間

平均値 μ , S.D.が σ の正規分布の95%予言的中区間は
 $(\mu - 1.96\sigma)$ 以上 $(\mu + 1.96\sigma)$ 以下

- 正規母集団に言い換えると, 「**母平均 μ から母標準偏差 σ の1.96倍の範囲内にあるデータが観測される**」と予言すると, それは「**95%の確率で的中する**」ということ
- 正規母集団から n 個のデータを観測する場合, その**標本平均に対する予言**をするならどの範囲を言えばいいのか?

標本平均と95%予言的中区間

正規母集団からの標本平均の95%予言的中区間

母平均が μ で母標準偏差が σ の正規母集団からのデータ n 個の
標本平均に対する95%予言的中区間は,
 $(\mu - 1.96\sigma/\sqrt{n})$ 以上 $(\mu + 1.96\sigma/\sqrt{n})$ 以下

不等式で表すと,

$$-1.96 \leq (\bar{x} - \mu)/(\sigma/\sqrt{n}) \leq +1.96$$

を解いて得られる範囲

正規分布に従うデータから母平均を引いてS.D.で割ると, 標準正規分布するデータに加工されるという性質を利用したもの

標本平均と95%予言的中区間

正規母集団からの標本平均の分布についての一般法則

正規母集団

→ n 個のデータの標本平均 $\bar{x}$

母集団のデータ

→ 標本平均 $\bar{x}$

$(\mu - 1.96\sigma) \sim (\mu + 1.96\sigma)$
の範囲に95%のデータが入る

→ $(\mu - 1.96\sigma/\sqrt{n}) \sim (\mu + 1.96\sigma/\sqrt{n})$
の範囲に95%の $\bar{x}$ が入る

95%予言的中区間の具体例

—— 正規母集団からの標本平均の95%予言的中区間 ——

母平均が μ で母標準偏差が σ の正規母集団からのデータ n 個の
標本平均に対する95%予言的中区間は,
 $(\mu - 1.96\sigma/\sqrt{n})$ 以上 $(\mu + 1.96\sigma/\sqrt{n})$ 以下

この法則を使えば、次の様な予言が可能

母集団が正規分布, 母平均が200, 母標準偏差が10とする

この母集団から1個だけデータを観測するとき,

- $200 - 1.96 * 10 = 180.4$
- $200 + 1.96 * 10 = 219.6$

を計算し,

「**観測されるデータは, 180.4以上219.6以下の範囲に入っている**」と
予言すれば, **95%**の確率で当たる

95%予言的中区間の具体例

正規母集団からの標本平均の95%予言的中区間

$(\mu - 1.96\sigma/\sqrt{n})$ 以上 $(\mu + 1.96\sigma/\sqrt{n})$ 以下

また, この母集団から4個のデータを観測しその標本平均を作るとき,

- $200 - 1.96 * (10 / \sqrt{4}) = 190.2$
- $200 + 1.96 * (10 / \sqrt{4}) = 209.8$

を計算し,

「観測される**4個のデータの標本平均は,**
190.2以上209.8以下の範囲に入る」と予言すれば, **95%の確率で当たる**

同じく16個の場合, 4個のときと同様に計算し,

「観測される**16個のデータの標本平均は,**
195.1以上204.9以下の範囲に入る」と予言すれば, **95%の確率で当たる**

標本平均と95%予言的中区間まとめ

比較すると、標本平均を作る個数が増えるほど、
予言する区間が狭くてすむ

つまり、

標本平均を作るための観測データの個数を多くすれば
するほど、精度の高い予言ができるようになる

第14項 まとめ

- 正規母集団の母平均を μ , 母標準偏差を σ とする
- 正規母集団からの標本の性質
正規母集団から観測される **データ x の n 個** に対する **標本平均 $\bar{x}$ の** (それらを集めたものを別の母集団ととらえたときの) **分布は正規分布** であり, **$\bar{x}$ の分布の平均値は μ のまま** だが, **標準偏差は $\sigma/\sqrt{n}$** と母集団に比べて $\sqrt{n}$ 分の1に縮む
- 正規母集団からの標本平均の95%予言的中区間
データ n 個の **標本平均 $\bar{x}$** に対する95%予言的中区間は,
 $(\mu - 1.96\sigma/\sqrt{n})$ 以上 $(\mu + 1.96\sigma/\sqrt{n})$ 以下
- 正規母集団からの標本平均の95%予言的中区間：不等式表示
 $-1.96 \leq (\bar{x} - \mu)/(\sigma/\sqrt{n}) \leq +1.96$
を **$\bar{x}$** について解いて得られる範囲