第6章（pp. 70）

中心極限定理

！重要！

中心極限定理

- 標本nの大きさが大きいとき
  標本平均の分布が正規分布に近似する
  - 詳細な説明については後述
  - 正規分布という特定の分布にしたがうならば
    確率の計算が簡単になる

標本平均の分布

（１）確率変数X_i

確率変数X_i

- 標本の観測値x_iは
確率変数X_iとして表すことができる
  - 母集団での比率πがわかれば
確率$\Pr(X = x) = {}_n C_x \pi^x (1 - \pi)^{n-x}$を
計算することができる
  - 母集団からランダムに抽出した標本は
母集団と同じ確率分布をもつと考えると
標本の観測値x_iを確率変数X_iとみなすことができる

確率変数X_i

- 標本の観測値x_iは
  確率変数X_iとして表すことができる
  - 確率変数X_iは確率変数Xと**同じ確率分布にしたがう**
  - それぞれの**X_iは独立**である
    - 母集団から標本を抽出したとしても
      母集団の分布である**確率分布は変化しない**ため

（２）標本平均の分布

標本平均の分布

標本における観測値$x_1, x_2, \ldots, x_n$を確率変数として$X_1, X_2, \ldots, X_n$とする

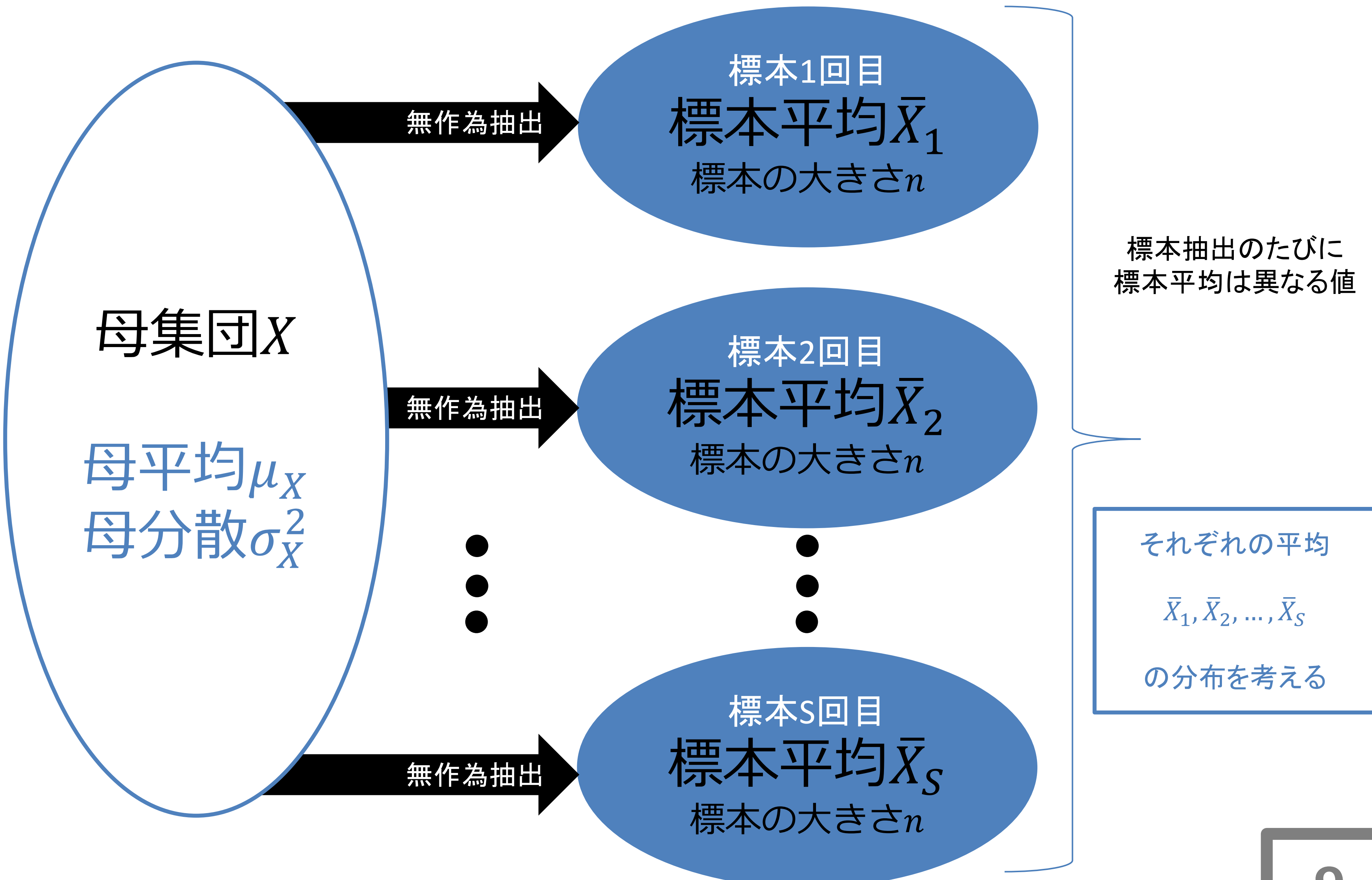
- 関数$g(X_1, X_2, \ldots, X_n)$

- 標本平均

$$\bar{X} = \frac{1}{n}\sum X_i$$

  – 確率変数の関数
    - 標本平均自体が確率変数であるといえる

標本平均の分布



母集団X
母平均μ_X
母分散σ_X^2
無作為抽出
無作為抽出
無作為抽出
標本1回目
標本平均$\bar{X}_1$
標本の大きさn
標本2回目
標本平均$\bar{X}_2$
標本の大きさn
標本S回目
標本平均$\bar{X}_S$
標本の大きさn
標本抽出のたびに
標本平均は異なる値
それぞれの平均
$\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_S$
の分布を考える

（３）標本平均$\bar{X}$の平均値と分散

標本平均$\bar{X}$の平均値と分散

- 母集団X
    - 母平均μ_X
    - 母分散σ_X^2
  - 期待値として表すと
    - 平均$E(X) = \mu_X$
    - 分散$Var(X) = \sigma_X^2$

- この母集団から大きさnの標本を無作為抽出
    - 確率変数X_iは確率変数Xと同じ確率分布にしたがうので
    - $E(X_i) = \mu_X$
    - $Var(X_i) = \sigma_X^2$

標本平均$\bar{X}$の平均値と分散

- 平均値 $E(\bar{X}) = \mu_X$
  - 母平均と同じになる
- 分散 $Var(\bar{X}) = \frac{\sigma_X^2}{n}$
  - 母分散の $\frac{1}{n}$倍

  - 標本平均を確率変数として考えて
    その確率分布の平均値と分散を求める
    （証明については省略(pp. 74)）

標本平均$\bar{X}$の平均値と分散

- 平均値 $E(\bar{X}) = \mu_X$
  - 母平均と同じになる
- 分散 $Var(\bar{X}) = \frac{\sigma_X^2}{n}$
  - 母分散の $\frac{1}{n}$倍

標本の大きさnを
大きくしたときに
標本平均の分散は小さくなる
↓
大数の法則

- 標本平均を確率変数として考えて
その確率分布の平均値と分散を求める
（証明については省略(pp. 74)）

（pp. 76）問題6-1

【難】標本調査の精度

- 無限母集団から
  標本を無作為抽出して調査を行う。
  - 標本の大きさ$n = 200$のとき
    標本平均の標準偏差は理論的には$\frac{\sigma_x}{\sqrt{200}}$
  - 調査の規模を10倍（$\mathrm{n} = 2000$）にしたとき
    調査の精度は10倍になるか？

標本平均の標準偏差を計算する

- 標本平均の標準偏差の大きさを調査精度と考える
- 標本平均の標準偏差は$n = 200$のときの何割程度になるか
- もしも調査の精度が10倍になるなら
  標本平均の標準偏差（ばらつき）は0.1程度になるはず

（pp. 76）問題6-1

標本調査の精度

- 標本の大きさを10倍にしても調査の精度は10倍にはならない

  （**標準偏差は$\frac{1}{10}$にはならない**）

$$\frac{\sigma_X}{\sqrt{200}} : \frac{\sigma_X}{\sqrt{2000}} = 1 : \frac{1}{\sqrt{10}} = 1 : 0.316$$

より、標準偏差は30%程度

  - 標本の大きさを100倍したときに
    $\frac{1}{\sqrt{100}} = \frac{1}{10}$より、標準偏差は10%程度となり、調査精度は10倍になる

- 標本が大きければ大きいほど
  標本は母集団に近づくが**比例するわけではない**ことに注意！

中心極限定理

！超重要！

- 中心極限定理
  - 母集団がどのような分布であっても
    無作為抽出した標本における和の分布は
    標本の大きさnが大きいときに**正規分布**になる

    ↓

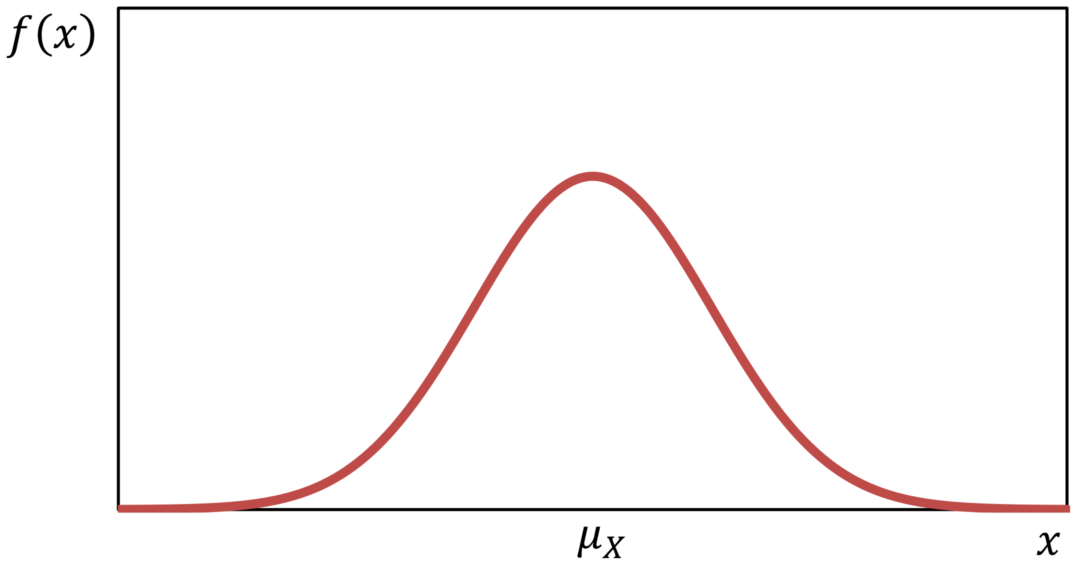
    - 標本平均について言い換えると
  - 母集団がどのような分布であっても
    無作為抽出した標本における**標本$\overline{X}$の分布**は
    標本の大きさnが大きいときに
    平均μ_X　分散$\frac{\sigma_X^2}{n}$　の正規分布になる
    - 標本平均がある範囲内に含まれる確率を、
      正規分布を用いて計算することが可能になる

正規分布（normal distribution）

- 連続確率変数Xの確率密度関数$f(x)$が$f(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right\}$となるときの$X$の確率分布

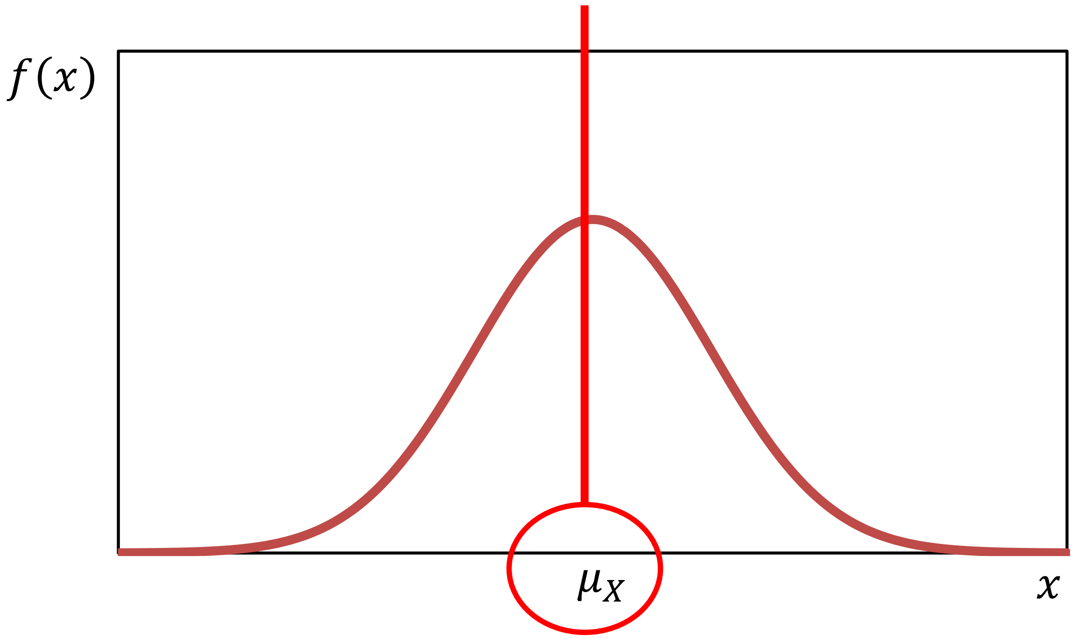
- $X \sim N(\mu_X, \sigma_X^2)$
- 平均を中心として左右対称の連続確率分布

正規分布（normal distribution）



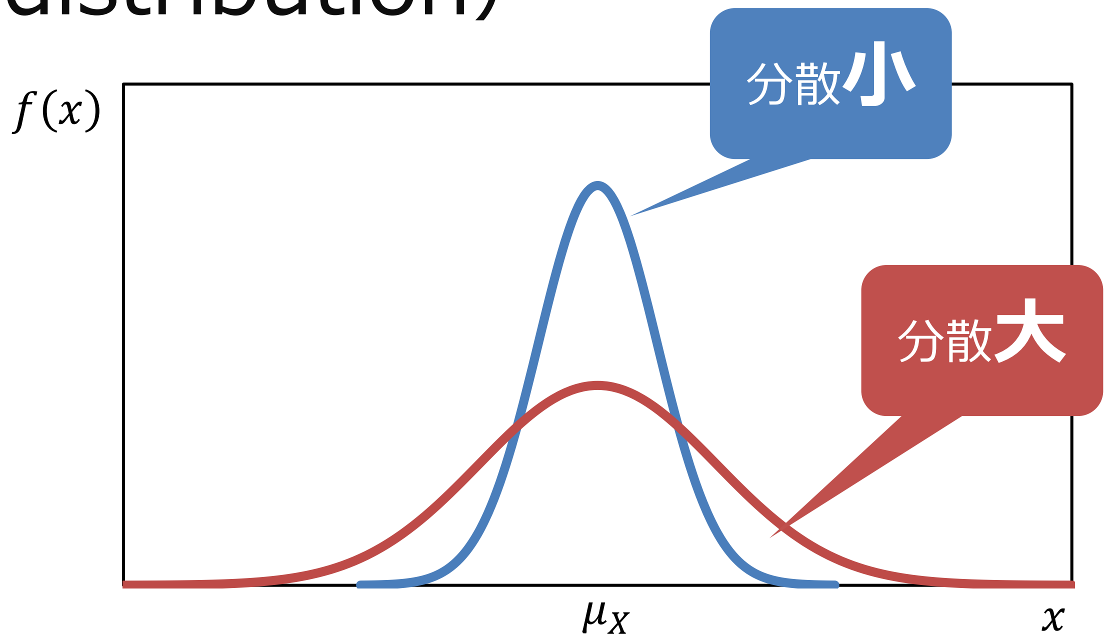
- $X \sim N(\mu_X, \sigma_X^2)$
- 平均を中心として左右対称の連続確率分布

正規分布（normal distribution）



- $X \sim N(\mu_X, \sigma_X^2)$
- 平均を中心として左右対称の連続確率分布

正規分布（normal distribution）



- $X \sim N(\mu_X, \sigma_X^2)$
- 平均を中心として左右対称の連続確率分布

第6章のまとめ

- 標本の観測値x_iを確率変数X_iとして表す
  - 標本平均 $\bar{X}$
  - 標本平均に関する平均値 $E(\bar{X}) = \mu_X$
  - 標本平均に関する分散 $Var(\bar{X}) = \frac{\sigma_X^2}{n}$
- 標本の大きさnを大きくしたとき
  標本平均は母平均に近づく
- 中心極限定理
  - 母集団の分布がどのような分布であっても
    無作為抽出した標本における標本平均は
    標本の大きさnが大きいときに正規分布にしたがう
- 正規分布
  - 中心は平均
  - 左右対称
  - 連続確率分布