統計学 01

早稲田大学政治経済学部
第12回
西郷 浩

本日の目標

- 区間推定
  - 母平均についての区間推定の仕組み（重要）
    - 統計量の標本分布の利用
  - 以下は次回以降に説明する。
    - 母集団比率についての区間推定（重要）
      - 教科書では、二項母集団に関する推定（pp. 229-230）と表現されている。
    - 母分散についての区間推定の仕組み
      - 使用頻度は上記2つに比べると少ない。

標本平均の標本分布(復習)

$X_i \sim N(\mu, \sigma^2)$かつ相互に独立ならば、

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \sigma^2/n\right) \Leftrightarrow \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$$

(標本抽出を何度も繰り返せば、$\bar{X}$が確率変数とみなせる。その確率分布＝標本分布が上のようにあたえられる。)

とくに、

$$P\left(\mu - 1.96\sqrt{\sigma^2/n} \le \bar{X} \le \mu + 1.96\sqrt{\sigma^2/n}\right) = 0.95$$

母平均の区間推定の仕組み(1)

不等式の変形によって、

$$\mu - 1.96\sqrt{\sigma^2/n} \le \bar{X} \le \mu + 1.96\sqrt{\sigma^2/n}$$

$$\Leftrightarrow \bar{X} - 1.96\sqrt{\sigma^2/n} \le \mu \le \bar{X} + 1.96\sqrt{\sigma^2/n}$$

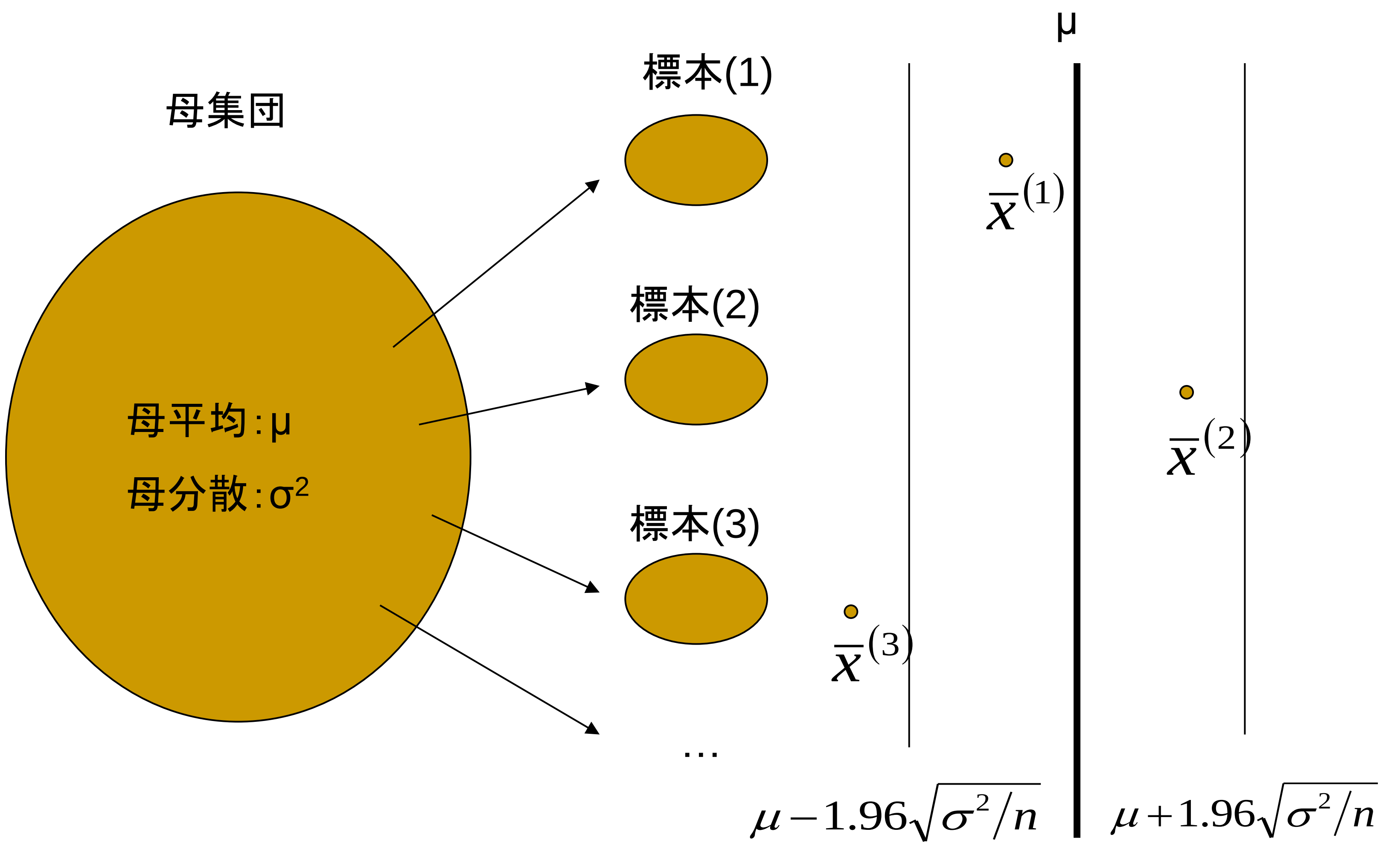
両者は事象として同等。したがって、

$$P\left(\mu - 1.96\sqrt{\sigma^2/n} \le \bar{X} \le \mu + 1.96\sqrt{\sigma^2/n}\right)$$

$$= P\left(\bar{X} - 1.96\sqrt{\sigma^2/n} \le \mu \le \bar{X} + 1.96\sqrt{\sigma^2/n}\right)$$

$$= 0.95$$

母平均の区間推定の仕組み(2)



母平均の区間推定の仕組み(3)

したがって、もしσ^2の値が既知であれば、

$$L = \bar{X} - 1.96\sqrt{\sigma^2/n}$$

$$U = \bar{X} + 1.96\sqrt{\sigma^2/n}$$

とすれば、

$$P\left(L \le \mu \le U\right) = 0.95$$

すなわち、「標本抽出を何度も繰り返せば、区間$[L, U]$が確率0.95で母平均μをふくむ」。

母平均の区間推定の仕組み(4)

- このような区間[L,U] を、信頼係数0.95の信頼区間という。
- ここで使った、「統計量の標本分布の性質を利用して、母数の存在しそうな範囲を指定する」という考え方は、いろいろな場面で利用できる。
- 以下では、
  - 「母分散σ^2が既知である」という条件
  - 「母集団が正規分布にしたがう」という条件

が成り立たない場合を順次考える。

母平均の区間推定の仕組み(5)

- 母分散σ^{2}が未知の場合の母平均の推定

$X_i \sim N(\mu, \sigma^2)$かつ相互に独立であれば、

$$\frac{\bar{X}-\mu}{\sqrt{s^2/n}} \sim t(n-1) \quad ただし、s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

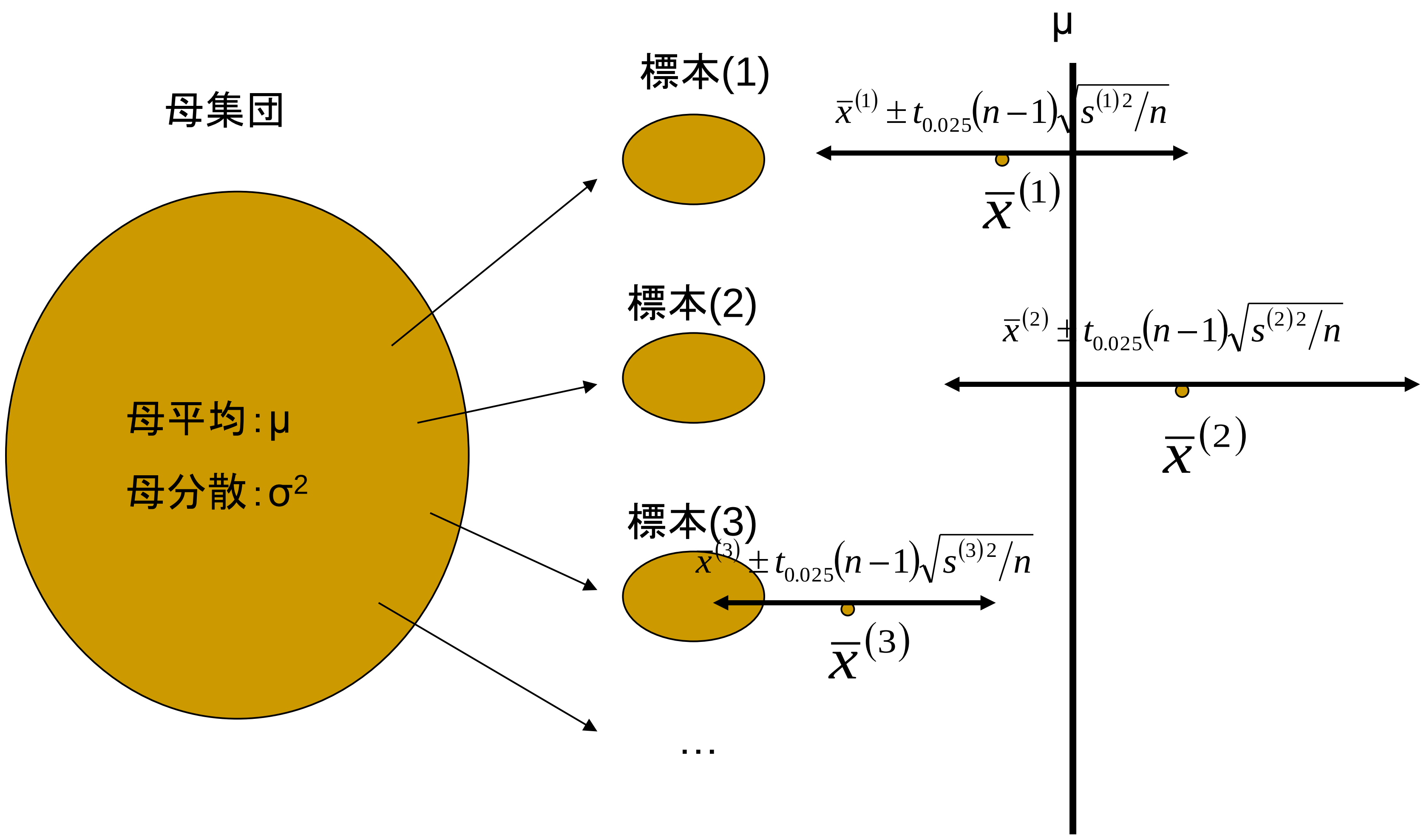
したがって、

$$P\left(-t_{0.025}(n-1) \le \frac{\bar{X}-\mu}{\sqrt{s^2/n}} \le t_{0.025}(n-1)\right)$$

$$= P\left(\bar{X}-t_{0.025}(n-1)\sqrt{s^2/n} \le \mu \le \bar{X}+t_{0.025}(n-1)\sqrt{s^2/n}\right) = 0.95$$

ただし、$t_{0.025}(n-1)$＝自由度$(n-1)$のt分布の上側0.025点

母平均の区間推定の仕組み(6)



母平均の区間推定の仕組み(7)

母平均 μ , 母分散 σ^2 が未知の正規母集団からの

標本にもとづく母平均の区間推定：

信頼係数0.95

信頼区間 $\left[L, U\right]$

ただし、$L = \bar{X} - \underbrace{t_{0.025}(n-1)}_{\text{自由度}n\text{-1 のt分布の上側0.025点}}\sqrt{s^2/n}$

$$U = \bar{X} + t_{0.025}(n-1)\sqrt{s^2/n}$$

(「標本抽出して母平均の信頼区間を構成する」という手続きを繰り返せば、確率0.95で、信頼区間が母平均 μ をその中にふくむ。$P\left(L \le \mu \le U\right) = 0.95$)

母平均の区間推定の仕組み(8)

- 例題：
  - 10人の被験者に対して、睡眠薬を投与したところ、睡眠時間（単位：時間）の増加は以下の通りであった。
    - $x_1 = 1.9$, $x_2 = 0.8$, $x_3 = 1.1$, $x_4 = 0.1$, $x_5 = -0.1$,
    - $x_6 = 4.4$, $x_7 = 2.5$, $x_8 = 1.6$, $x_9 = 2.6$, $x_{10} = 1.4$
  - 睡眠時間の増加 X が正規分布にしたがうと仮定して、この睡眠薬による平均的な睡眠時間の増加分（母平均）の、信頼係数0.95の信頼区間を構成せよ。

母平均の区間推定の仕組み(9)

解答

$$\bar{x}=1.63,\quad s^2=1.75,\quad t_{0.025}(10-1)=2.262$$

$$L=\bar{x}-t_{0.025}(n-1)\sqrt{s^2/n}=1.63-2.262\sqrt{1.75/10}=0.68$$

$$U=\bar{x}+t_{0.025}(n-1)\sqrt{s^2/n}=1.63+2.262\sqrt{1.75/10}=2.58$$

したがって、信頼係数0.95の信頼区間は$[0.68, 2.58]$

母平均の区間推定の仕組み(10)

- 正規分布ではない母集団からの標本にもとづく母平均の区間推定
  - 中心極限定理の利用
    - つまり、「サンプルサイズ n が大きくなるにつれて、標本平均の標本分布が正規分布に近づく」という性質をもちいる。
    - 母分散σ^2 が未知の場合（たいていは未知である）は、標本分散 s^2 で代用する。サンプルサイズが大きくなるにつれて、標本分散 s^2 は母分散σ^2 に近づく（前者は後者の一致推定量）ので、サンプルサイズが大きくなるしたがって、代用の影響は小さくなる。
    - 一概にはいえないものの、$n \geqq 30$ であれば正規近似が使われることがある。

母平均の区間推定の仕組み(11)

- 例題:
  - 在籍者5000人の学部から、100人を無作為抽出して英語のテストをしたところ、平均点が60点、分散が20^2であった。5000人全員にテストをしたとすると、テストの平均点はどれくらいになるか。
  - 解答
    - サンプルサイズが大きい(100)と考えて、標本平均の標本分布が正規分布で近似できると仮定する。

$$\bar{x} = 60,\ s^2 = 20^2$$

$$\bar{x} \pm 1.96\sqrt{s^2/n} = 60 \pm 1.96\sqrt{20^2/100} = 56.08,\ 63.92$$

母平均の区間推定の仕組み(12)

- 母集団が正規分布にしたがうかどうかがわからない場合でも、t分布によって信頼区間を構成することが多い。その方が、信頼区間が広めになるという意味で安全でもある。
- 正規近似を使うのであれば、（所詮近似なのだから）1.96≒2 として、信頼係数0.95の信頼区間を構成することも多い。

問題演習

- 次回までにつぎの問題を解いてくるように。次回の講義で解説します。
  - 11.3, 11.7 i) p. 231.