統計学 01

早稲田大学政治経済学部
第8回
西郷 浩

本日の目標

- 第10章 正規分布からの標本
  - 本章の位置づけ
  - 正規分布の復習
  - 標本平均の標本分布（母分散既知の場合）
  - 標本分散の標本分布
  - 標本平均の標本分布（母分散未知の場合）

第10章の位置づけ(1)

- 第9章のポイント
  - 母数と統計量とは等しくならない。
  - しかし、統計量を「標本抽出にともなう確率変数」とみなし、その確率分布（統計量の標本分布）を考えると、母数と関連づけられる。
    - 第11章以降で明らかになるように、この「統計量の標本分布」の性質を利用して、統計的推測（手許の標本から未知の母数について推測すること）の理論が組み立てられる。
  - したがって、ある統計量の標本分布がどのような確率分布によってあらわされるかが重要になる。

第10章の位置づけ(2)

- ところが、任意の母集団の分布型について、任意の統計量の標本分布を厳密に導くことは難しい。
- 例外的に、正規母集団の場合には、ある種の統計量の標本分布が厳密に導かれている。
  - 例：正規分布、χ^2分布、t分布、F分布
  - さらに、標本の大きさ n がある程度大きければ、母集団の分布型によらず、標本平均の標本分布は正規分布で近似できる（中心極限定理）。
- 母集団の分布として正規分布を仮定することには、理論分析の基準としての意味がある。
- 「統計量の標本分布がきれいに導ける象徴的な例」として本章を学ぶ。

第10章の位置づけ(3)

- 証明
  - 井上先生担当の「数理統計学A」､「数理統計学B」
  - 岩田暁一(1983)『経済分析のための統計的方法』第2版 東洋経済新報社

正規分布の復習(1)

$X \sim N\left(\mu, \sigma^2\right)$のとき、

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \mathrm{d}x$$

$$E(X) = \mu, \quad V(X) = \sigma^2$$

$$P(\mu - 1.645\sigma \leq X) = P(X \leq \mu + 1.645\sigma) = 0.95$$

$$P(\mu - 1.96\sigma \leq X \leq \mu + 1.96\sigma) = 0.95$$

正規分布の復習(2)

$X \sim N(\mu, \sigma^2)$のとき、$Y = aX + b$とすれば、

$Y \sim N(a\mu + b, a^2\sigma^2)$

とくに、$Z = (X - \mu)/\sigma$とすれば、

$Z \sim N(0, 1)$

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$で両者が独立のとき

$X \pm Y \sim N(\mu_1 \pm \mu_2, \sigma_1^2 + \sigma_2^2)$ （複合同順）

正規母集団からの標本平均の標本分布
母分散既知の場合(1)

$X_i \sim N\left(\mu, \sigma^2\right) \quad \left(i = 1, 2, \ldots, n\right)$かつ相互に独立

であれば、

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

(第7.4節参照)

もし、母分散σ^2が既知であれば、この関係式をそのまま母平均μに関する推測に利用できる。

正規母集団からの標本平均の標本分布 母分散既知の場合(2)

- 例題（ホーエル(1981)問題15, p. 135）
  - B大学での過去5年間の男子新入生の体重: 平均=154（ポンド） 標準偏差=20 （ポンド）、ほぼ正規分布で近似できる。
  - 今年の男子新入生から無作為に選んだ100人の体重: 平均=159（ポンド）
  - 今年の男子新入生の平均体重は、例年よりも大きいと考えるべきか？

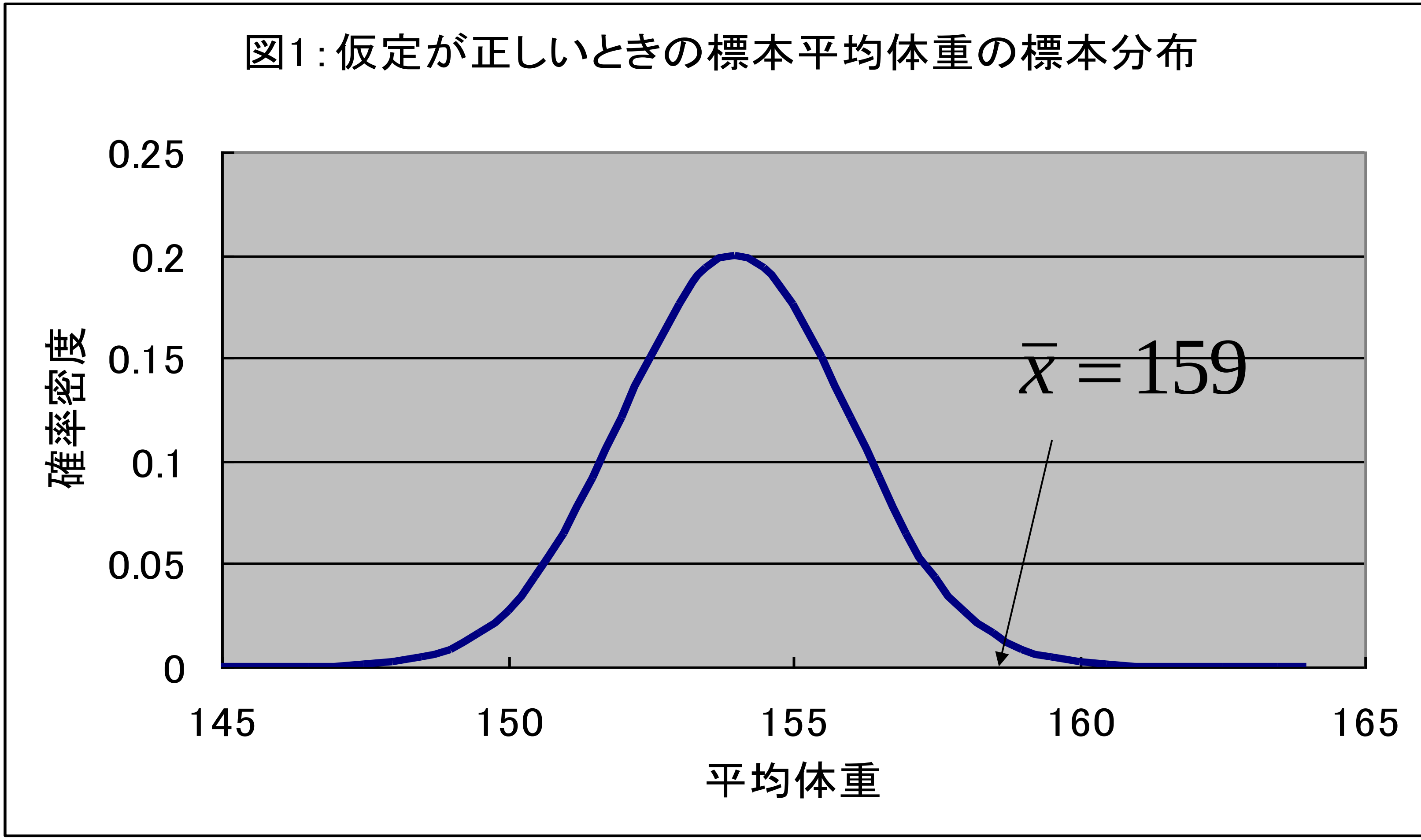
正規母集団からの標本平均の標本分布
母分散既知の場合(3)

- 「今年の男子新入生の体重の分布が例年と変わらない」と仮定する。
- そのとき、無作為抽出された100人の男子新入生の平均体重の分布は、

$$\overline{X} \sim N\left(154, 20^2/100\right)$$

  - ただし、あくまでも、「今年の男子新入生の体重の分布が例年と変わらない」という仮定のもとで成り立つ性質であることに注意。
    - もし、現実と(確率的に)不整合な結果が実際に生じたときには、この仮定がおかしいと考えられる。

正規母集団からの標本平均の標本分布
母分散既知の場合(4)



正規母集団からの標本平均の標本分布
母分散既知の場合(5)

- 仮定が正しいとすると、極めて珍しい現象が起きたことになる。
- 2つの選択肢
  - 「仮定が正しいにもかかわらず、たまたま珍しいことが起きた」と考える。
  - 「仮定が間違っていた」と考える。
- どちらが自然な発想か？
- 注：本来は、統計的仮説検定で解くべき問題。

正規母集団からの標本分散の標本分布(1)

$X_i \sim N\left(\mu, \sigma^2\right) \quad \left(i = 1, 2, \ldots, n\right)$かつ相互に独立のとき、

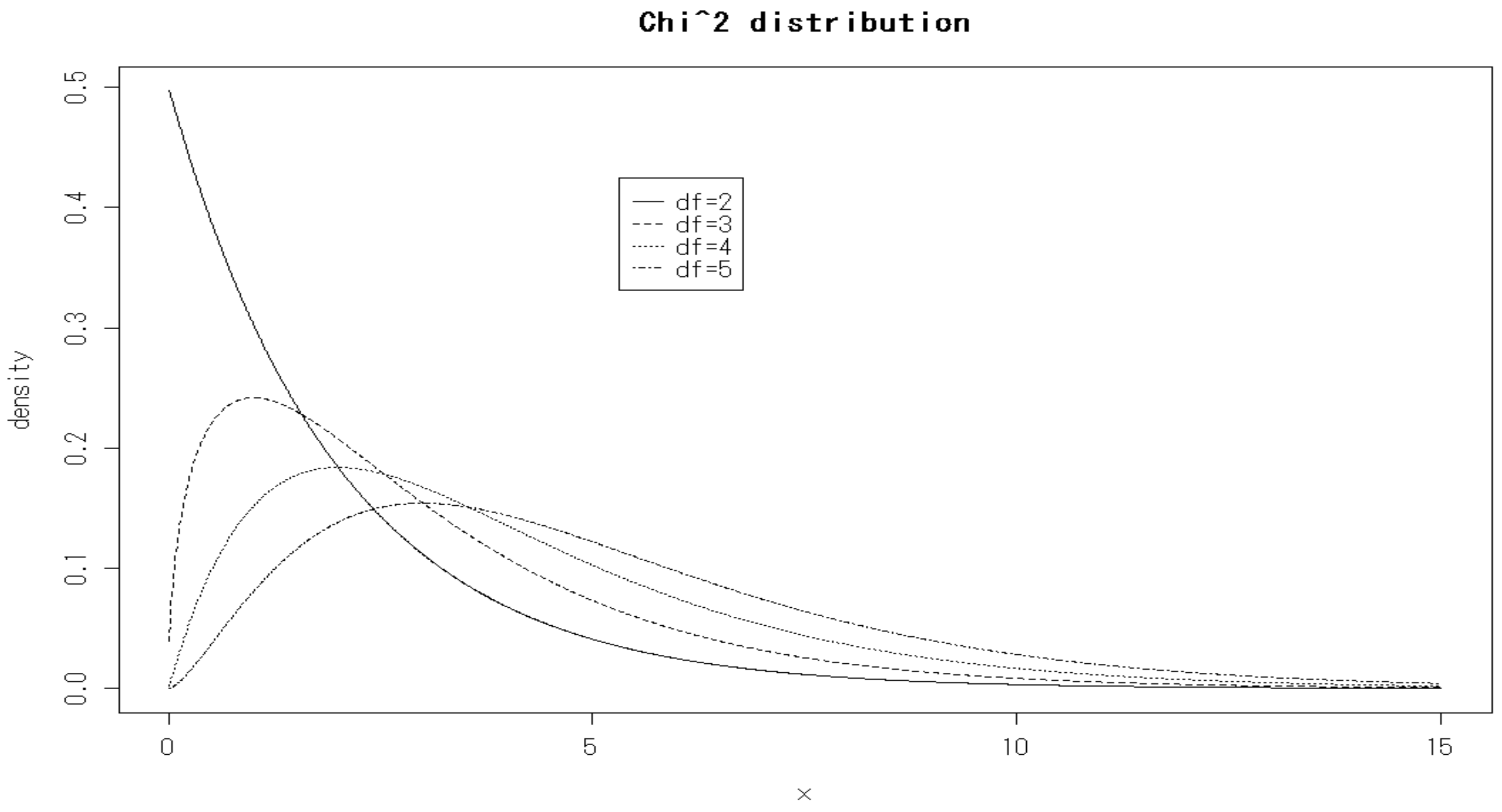
$$\frac{\left(n-1\right)s^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2\left(n-1\right)$$

ここで重要なのは、「$\left(n-1\right)s^2 / \sigma^2$の標本分布が求められる」という事実であって、導出過程は気にしなくてよい。

正規母集団からの標本分散の標本分布(2)

- χ^2分布
  - Z_i ～N(0,1) (i=1,2,…,k)かつ相互に独立であるとき、
    $Y = \sum_i Z_i^2$
    の理論分布が導ける。その分布の形状は k によって異なる。この理論分布をχ^2（カイ二乗）分布とよび、k を自由度という。確率変数Yが自由度 k のχ^2分布にしたがうとき、
    $Y \sim \chi^2(k)$
    とあらわす。

正規母集団からの標本分散の標本分布(3)



正規母集団からの標本分散の標本分布(4)

$X_i \sim N(50, 25) \quad (i = 1, 2, \ldots, 10)$かつ相互に独立であるとき、標本分散$s^2$が50を超える確率はどのくらいか。

答え

$$P(s^2 > 50) = P\left(\frac{(10-1)s^2}{25} > \frac{(10-1)\times 50}{25}\right)$$

$$= P\left(\frac{(10-1)s^2}{25} > 18\right) = 0.038$$

正規母集団からの標本平均の標本分布
母分散未知の場合(1)

$X_i \sim N(\mu, \sigma^2)$ $(i = 1, 2, \ldots, n)$かつ相互に独立のとき、

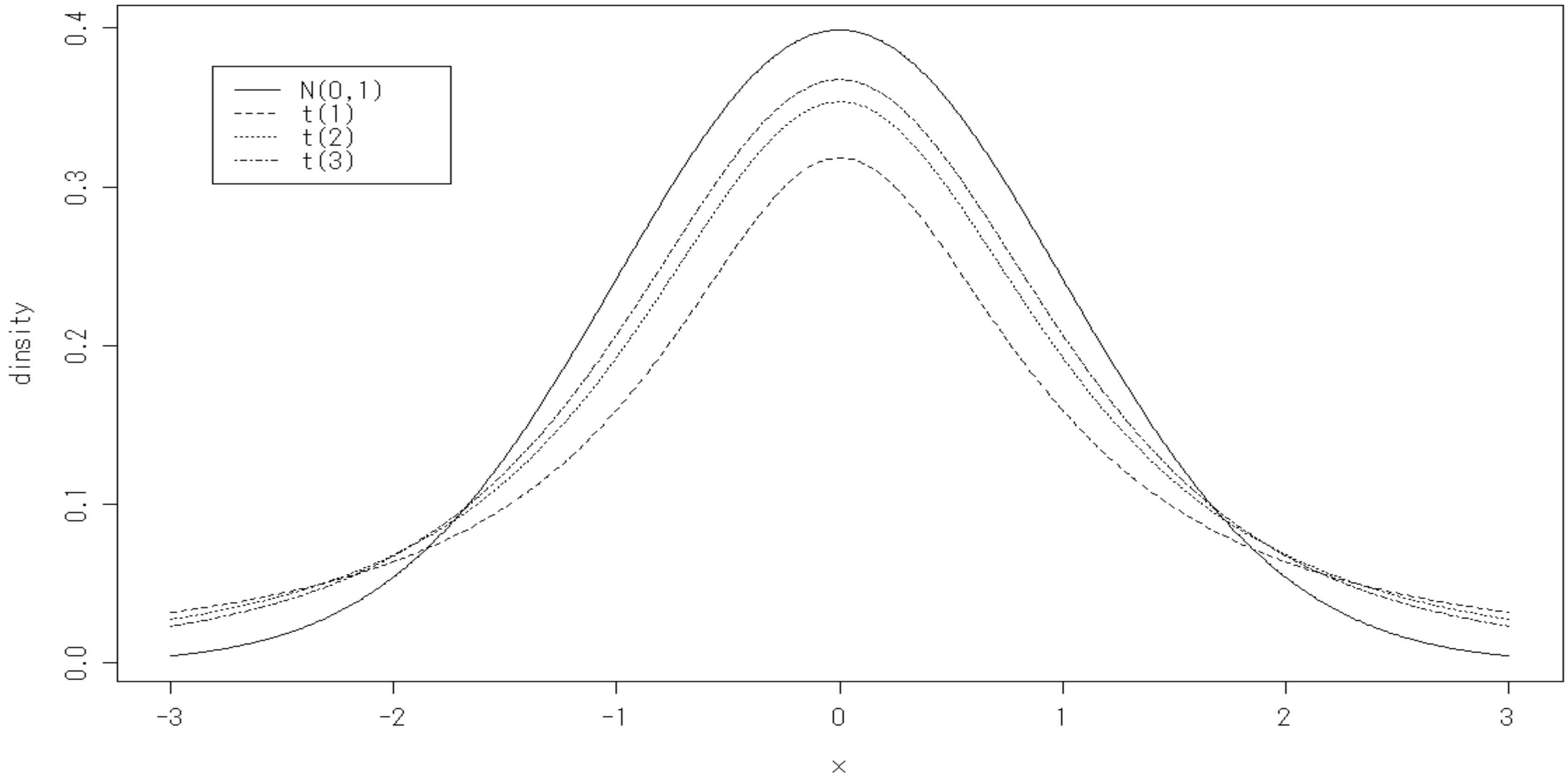
$$Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2 / n}} \sim N(0, 1)$$

未知母数のσ^2を標本分散（既知）s^2で置き換える。

$$T = \frac{\bar{X} - \mu}{\sqrt{s^2 / n}} \sim t(n-1) \quad \text{自由度}(n-1)\text{の}t\text{分布}$$

「σ^2をs^2で置き換えても、（母数に依存しない）理論分布が求められる」という点が大事。

正規母集団からの標本平均の標本分布
母分散未知の場合(2)



正規母集団からの標本平均の標本分布 母分散未知の場合(3)

- t分布の特徴
  - 左右対称
  - 自由度によって形が変わる。
    - 自由度が大きくなる
      → 分布の裾がだんだん薄くなる＋頂点がだんだん高くなる
      → N(0, 1) に近づいていく。
- t分布表(p. 281)の利用
  - 表が複雑なので慣れておくこと。

問題演習

- 教科書 p. 210
  - 10.1
  - 10.2
  - 10.3
  - 10.4