統計学 2

13. t 分布と母平均の推定

矢内　勇生

2019年5月30日

高知工科大学　経済・マネジメント学群

今日の目標

- 標本平均から母平均を推定する方法を理解する
    - もうやったのでは？？？
- t 分布を理解する

母平均の推定

母分散が既知の場合（復習）

- 母平均の点推定：標本平均 = 点推定値
- 母平均の区間推定：
  - 標本平均の分布が正規分布に従うと考え、95%信頼区間を求めた

標準正規分布の特徴を利用して推測する

- 標準正規分布の特徴：[-1.96, 1.96] の区間にデータの95%が収まる

- 正規分布に従う変数を標準化することで、標準正規分布を使える

★**標本サイズ（*n*）が大きくなれば**、誤差の分布は正規分布に近づく（中心極限定理）

標本平均を標準化する

- 標本平均の平均 = 母平均 μ
- 標本平均の標準偏差 = 標準誤差SE

➡ 標本平均の z 値は、

$$z = \frac{\bar{x} - \mu}{\mathrm{SE}} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

z値の95%が[-1.96, 1.96]にある

- 標本平均のz値のうち、95%は 区間[-1.96, 1.96]に収まるはず

- つまり、たくさんある標本の95%について、次の式が成り立つ：

$$-1.96 \leq z \leq 1.96$$

$$-1.96 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$$

95%信頼区間を求める（1）

$$-1.96 \leq \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$$

- 既知のもの：$n, \bar{x}$（σ も知っているとする）
- 推定の対象：μ
- 上の不等式を μ について解けば、μ（母平均）の95%信頼区間が得られる

95%信頼区間を求める（2）

$$-1.96 \leq \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96$$

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

➡ μ の95%信頼区間は

$$\left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

$$= [\bar{x} - 1.96 \cdot \mathrm{SE}, \bar{x} + 1.96 \cdot \mathrm{SE}]$$

母分散が未知だったら？

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

- 母分散を知らない $\rightarrow$ 上の式が統計量にならない（標本から数値を得ることができないから）

‣ どうする？

➡ 区間推定の方法が変わる

★ただし、点推定値にはいつも標本平均を使う

母分散を知らない（かつnが十分大きくない）ときの区間推定

- σ を知らないとき：σ の推定値として u を使う
  - $\dfrac{\bar{x}-\mu}{\frac{u}{\sqrt{n}}}$ は標準正規分布に従わない

➡ 標準正規分布を使って求める信頼区間は使えない

➡ 標準正規分布の代わりに t **分布**を利用する

分散 (variance)

- 分散：データのばらつき（散らばり具合）を表す統計量
- 0以上の値をとる（データの値がすべて等しいとき0）
- データのばらつきが大きいほど、分散も大きくなる

分散の求め方（復習1）

- 標本分散を表す記号：s^2（σ^2は母分散）

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

標本分散の求め方（復習2）

1. 標本平均を求める
2. 偏差（=観測値 - 標本平均）を各観測値について求める
3. それぞれの偏差を2乗する
4. 偏差の2乗をすべて足して、標本サイズ（n）で割る（偏差^{2}の平均値を求める）

標本分散の求め方（復習3）

- 右のデータの分散を求めてみる

$$s^2 = \frac{16+1+0+4+9}{5}$$
$$= \frac{30}{5} = 6$$

	x	偏差	偏差^2
	1	-4	16
	4	-1	1
	5	0	0
	7	2	4
	8	3	9
計	25	0	30

標本分散の偏り

- 標本をいくつも抽出し、それぞれの標本について標本分散を求める：

標本分散の平均値 < 母分散

➡標本分散を母分散の推定に使うと、（小さい方に）偏ってしまう

★偏りがない分散の求め方は？

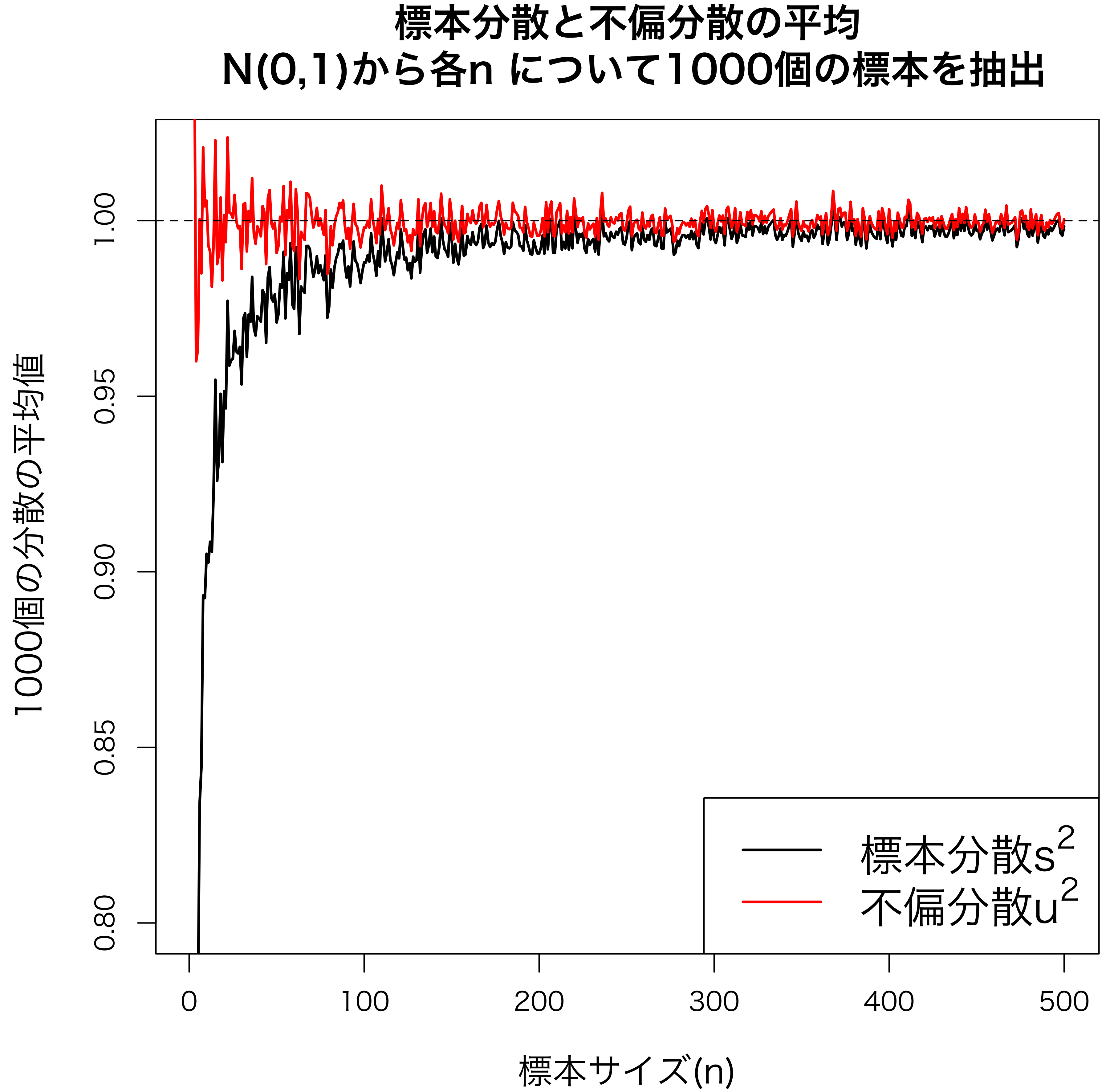
不偏分散（1）（unbiased variance）

- 母分散の偏りのない推定値として使える
- 記号：u^2 で表す
- 分散を求める式の分母を少しだけ小さくする

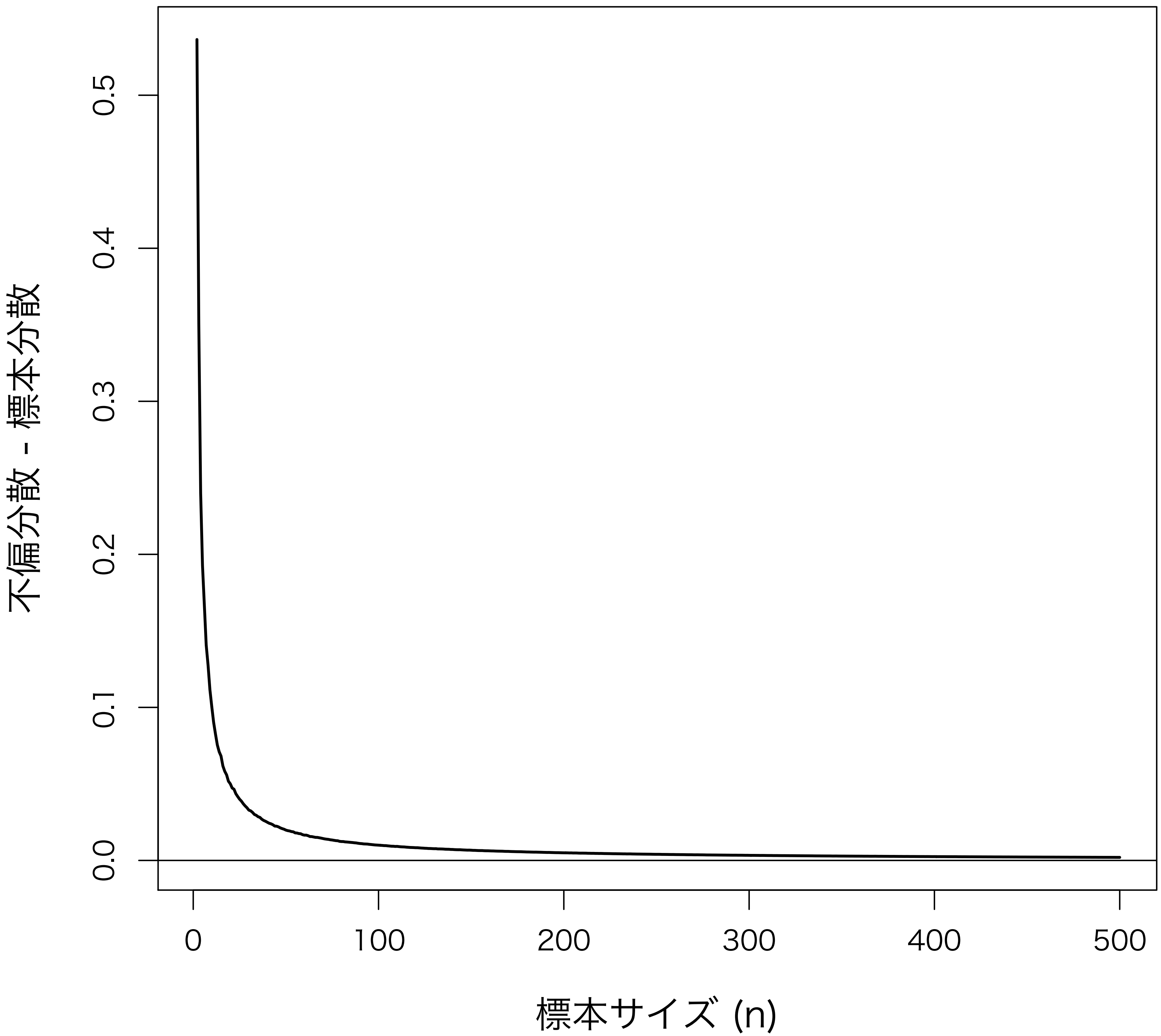
$$u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

不偏分散（2）

- 不偏分散：母分散の点推定値
- 標準偏差の点推定値は：u（不偏分散の平方根）
    - n が十分大きいときは、 n で割ってもよい（結果がほとんど同じなので）
    - 母平均を知っているときは、n で割って分散を求める


標本分散と不偏分散の平均
N(0,1)から各n について1000個の標本を抽出
1000個の分散の平均値
0.80
0.85
0.90
0.95
1.00
0
100
200
300
400
500
標本サイズ(n)
標本分散s^2
不偏分散u^2

不偏分散と標本分散の差



母分散を知らない（かつnが十分大きくない）ときの区間推定

- σ を知らないとき：σ の推定値として u を使う
  - $\dfrac{\bar{x}-\mu}{\frac{u}{\sqrt{n}}}$ は標準正規分布に従わない

➡ 標準正規分布を使って求める信頼区間は使えない

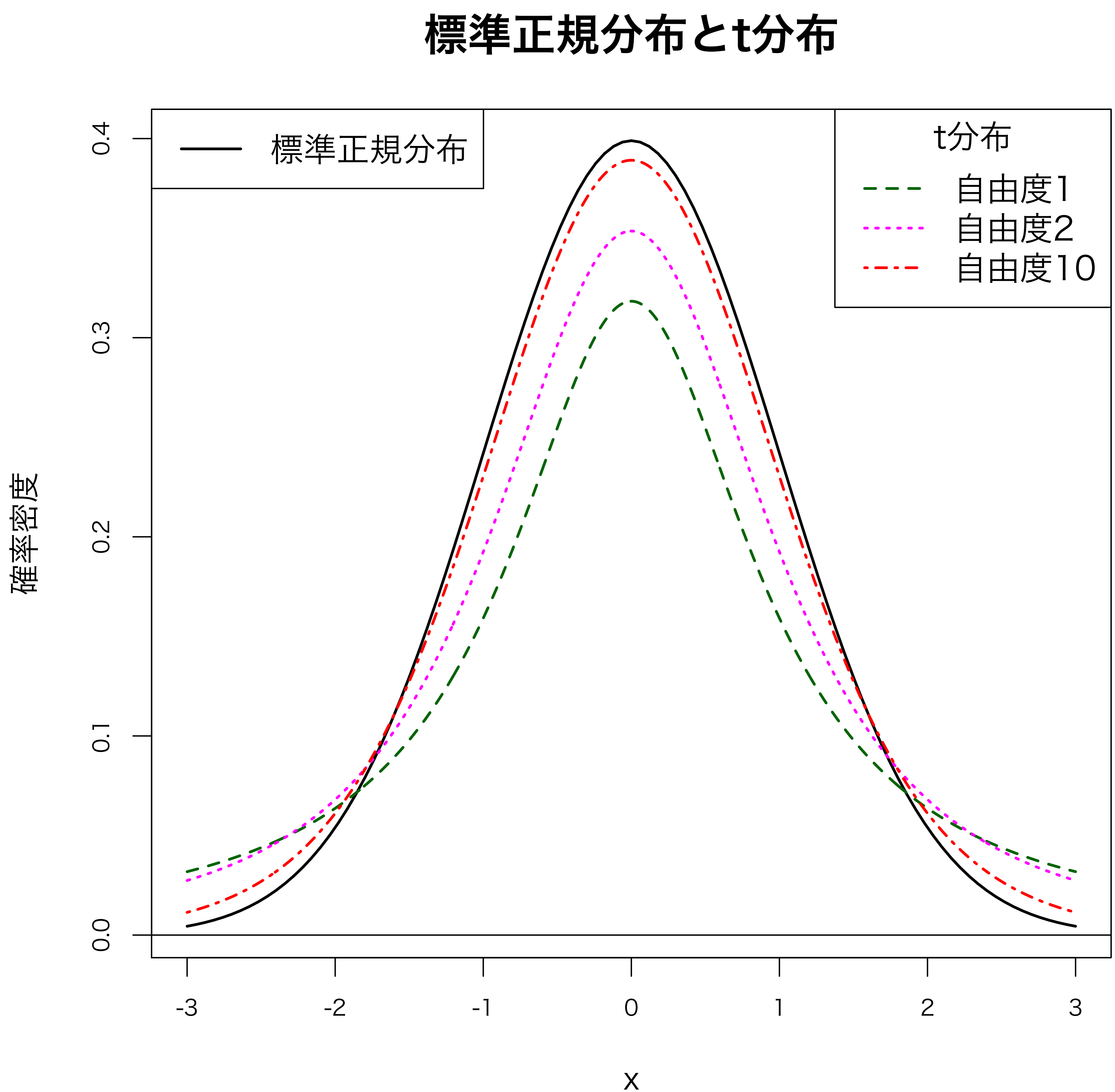
➡ 標準正規分布の代わりに t **分布**を利用する

t 分布 (Student's t distribution)

- スチューデント（Student）の t 分布
- 確率密度関数：

$$f(x;k) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

ただし、kはxの自由度（k>0）、Γ (.) はガンマ関数。


標準正規分布とt分布
標準正規分布
t分布
自由度1
自由度2
自由度10
0.0
0.1
0.2
0.3
0.4
確率密度
-3
-2
-1
0
1
2
3
x

t 分布の特徴

- 自由度によって形が決まる
- 概形は標準正規分布に似ている
  - 分布の中心は0
  - 標準正規分布より山の頂上が低い
  - 標準正規分布より裾が厚い
- 自由度が大きくなるにつれ、標準正規分布に近づく

ウィリアム・ゴセット (William Sealy Gosset：1876-1937)

- イギリスの統計学者・醸造技術者（ギネス社に勤務）
- 推測統計学の確立に貢献
- ペンネーム：Student
- t 分布を発見した

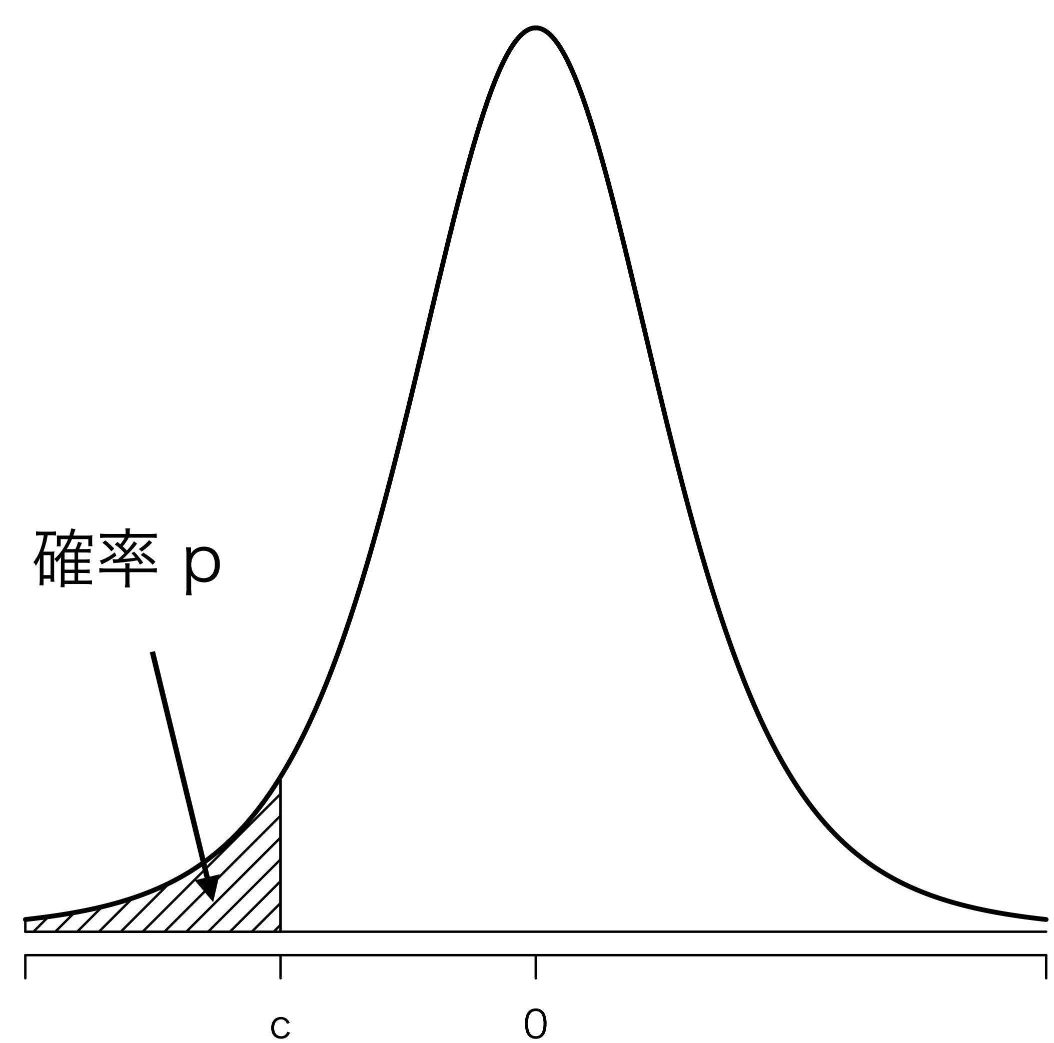
t分布の使い方

- 自分が求めたい確率pと自由度　df の組み合わせを決める

- Rでcを求める

```
qt(p, df)
```

- t 分布は左右対称なので、片側（負の側、左側）の c がわかれば、反対側（正の側、右側）-c もわかる



標本平均とt分布

- $\dfrac{\bar{x}-\mu}{\frac{u}{\sqrt{n}}}$ は、自由度 *n*-1 の t 分布に従う
- 標本サイズ（*n*）が10のとき
  - 自由度 = 9で p = 0.025のとき、c = -2.2622
- 標本の95%について

$$-2.26 \leq \frac{\bar{x}-\mu}{\frac{u}{\sqrt{n}}} \leq 2.26$$

一般的な場合

- 一般的に、自由度 kと p によって決まるt分布の-c を$t_{k,p}$と書くことにすると、標本の100(1-2p)%について、

$$-t_{n-1,p} \leq \frac{\bar{x} - \mu}{\frac{u}{\sqrt{n}}} \leq t_{n-1,p}$$

となる。

- p=0.025 なら、100(1-2α) = 95%について、

$$-t_{n-1,0.025} \leq \frac{\bar{x} - \mu}{\frac{u}{\sqrt{n}}} \leq t_{n-1,0.025}$$

母平均の95%信頼区間

$$-t_{n-1,0.025} \leq \frac{\bar{x}-\mu}{\frac{u}{\sqrt{n}}} \leq t_{n-1,0.025}$$

- 上の式のうち、標本平均 $\bar{x}$、不偏分散の平方根 u、標本サイズ n は知っていて、母平均 μ を推定したい

- 上の不等式を μ について解けばよい

$$\bar{x} - t_{n-1,0.025}\frac{u}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1,0.025}\frac{u}{\sqrt{n}}$$

今日のまとめ

- 母平均の点推定値は標本平均
- 母平均の信頼区間の求め方は、状況によって異なる
  - 母分散が既知~~（または標本サイズ n が十分大きいとき）~~
    - 標準正規分布を使った信頼区間
  - 母分散が未知
    - t 分布を使った信頼区間