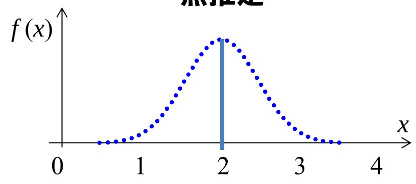
統計学入門 補助資料
～正規母集団の区間推定～

2022年度1学期:　　月曜2限
担当教員:　　　　石垣　司

推定

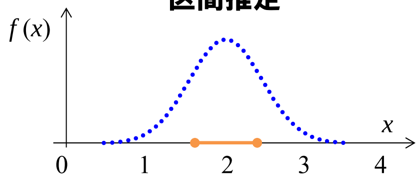
- 標本から母集団の分布のパラメータ（母数）を求める
  - 点推定　未知のパラメータ θ を1点で推定
  - 区間推定　未知のパラメータ θ が存在する区間を推定
  - データ（標本の実現値）を利用して未知 θ の値を推定
    - 例　データ $\{2, 3, 1, 2, 3, 4, \ldots, 2\}$ から母集団の母平均を推定

点推定



$\hat{\theta} = 2$

区間推定



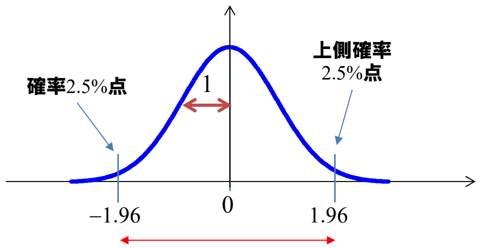
$P(1.7 \leq \hat{\theta} \leq 2.3)$

正規母集団の区間推定

- 正規母集団の母平均・母分散を区間推定
  - 正規母集団以外の母集団は, n が大きい場合, 標本平均を中心極限定理で正規分布に近似できる
  - 信頼区間による推定
- 講義内容の整理（正規母集団に限定）
  - ①母分散が既知の場合の母平均の区間推定
  - ②母分散が未知の場合の母平均の区間推定
    - t 分布、カイ2乗分布
  - ③　正規母集団の母分散の区間推定

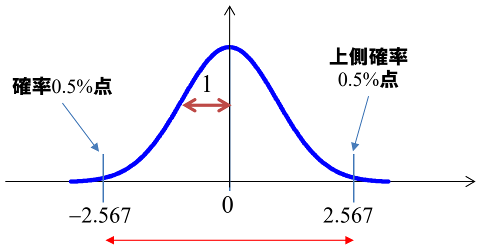
（復習）標準正規分布の上側確率点と信頼区間

- Z_α　標準正規分布の上側確率 $100 \times \alpha\%$点
  - 97.5%点（上側確率2.5%点）　$Z_{0.025} = 1.96$
  - 2.5%点　$Z_{0.975} = -Z_{0.025} = -1.96$
  - 99.5%点（上側確率0.5%点）　$Z_{0.005} = 2.567$
  - 0.5%点　$Z_{0.995} = -Z_{0.005} = -2.567$



この区間の確率（面積）0.95

$P(-1.96 \leq Z \leq 1.96) = 0.95$



この区間の確率（面積）0.99

$P(-2.567 \leq Z \leq 2.567) = 0.99$

①正規母集団の母平均の区間推定（母分散既知）

- **例　母平均 μ の95%信頼区間(復習)**
  - **標準化した標本平均** $\bar{Z} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$

$$P(-1.96 \leq \bar{Z} \leq 1.96) = 0.95$$
$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

  - σ **の値は既知なので信頼区間は計算可能**

- **しかし, 多くの実データの分析において,**
  平均が未知で分散が既知という状況は考えにくい

②正規母集団の母平均の区間推定（母分散未知）

- **"母分散 σ^2 は既知"の条件は現実的ではない**
  ⇒　母分散 σ^2 を標本分散 S^2 で代用

- **t 統計量**
  - **$\bar{Z}$ の中の σ を S で代用した統計量 T**

$$\bar{Z} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \rightarrow T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

  - **T の分布は標準正規分布ではない**
  - **大標本 (サンプルサイズ n が $n \geq 30$ くらい)**
    ⇒ 中心極限定理より T を標準正規分布で近似
  - **小標本 ($n < 30$ くらい) ⇒ t 分布**

小標本の区間推定に必要な確率分布

William Sealy Gosset
1876-1937
ギネスビール社員
"Student" はペンネーム

- Student **の** t **分布** (1908)
  - **t 統計量の従う確率分布**
  - **ビールの品質管理から誕生**
    - **醸造器内の酵母数を標本抽出し顕微鏡で数えて厳密管理**
    - **当初は $\sigma^2 = S^2$ としてビール品質の監視**
    - **しかし, 小標本のとき, S^2 の値がおかしいと気付く**
    - **$\bar{X}$ の信頼性評価に正規分布**(with $\sigma = S/\sqrt{n}$)**は不適の疑念をもつ**
  - **⇒小標本のときの S^2 の正しい見積もりが必要**

- **カイ2乗分布** $\chi^2_{(n)}$
  - **標本分散 S^2 に関連する確率分布**

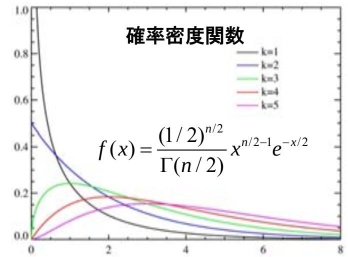
東京大学出版会「統計学入門」p.203コラム、D.サルツブルク「統計学を拓いた異才たち」第3章

χ^2分布

- **自由度 n のカイ2乗分布** $\chi^2_{(n)}$
  - **確率変数 $Z_i \sim i.i.d.N(0,1)$ の2乗和 χ^2 の従う分布**

$$\chi^2 = Z_1^2 +, \ldots, +Z_n^2$$

  - **パラメータ　自由度 n（サンプルサイズ n に対応）**
  - **平均と分散**　$E[\chi^2] = n, V[\chi^2] = 2n$
  - **再生性をもつ**
    - $X_i \sim \chi^2_{(n)}$ **(自由度 n の χ^2分布)**
    - $Y_i \sim \chi^2_{(m)}$ **(自由度 m のχ^2分布)**
    - $X_i + Y_i \sim \chi^2_{(n+m)}$



Wikipedia (public domain)

t 分布と χ^2 分布の関係

- t 統計量は, $\bar{Z}$ と χ^2 分布に従う確率変数の比
  - 標本 $X_i \sim i.i.d. N(\mu, \sigma^2)$ のとき U は自由度 $n-1$ の χ^2 分布に従う

$$U = (n-1)\frac{S^2}{\sigma^2}$$

  - t 統計量

check!

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{\bar{Z}}{\sqrt{U/(n-1)}}$$

← 標準正規分布

← 自由度 $n-1$ の χ^2 分布

    - 標準正規分布に従う確率変数と χ^2 分布に従う確率変数の比
    - 両者の密度関数の比から t 統計量の密度関数を生成

- <u>t 統計量が従う分布を自由度 $n-1$ の t 分布とよぶ</u>

$U \sim \chi^2_{(n)}$ の証明は煩雑なため省略。興味のある人は、「高杉&馬場, 演習統計学キャンパス・ゼミ, マセマ出版社」の p.168 などを参照

t 分布

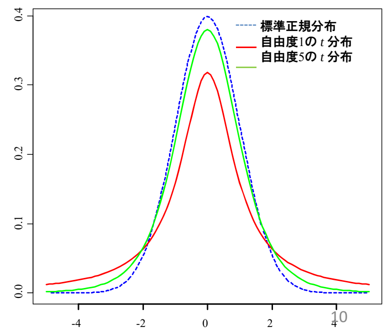
- 自由度 n の t 分布
  - $Z \sim N(0,1)$, $W \sim \chi^2_{(n)}$, Z と W は独立のとき, t が従う分布

$$t = \frac{Z}{\sqrt{W/n}}$$

  - パラメータ　自由度 n
  - 平均と分散
    $E[t] = 0,\ V[t] = \frac{n}{n-2}$ $(n > 2)$
    - $n = 1$ のときコーシー分布
  - 原点に対象
  - 自由度 n が大きい時($n > 30$ くらい)
    正規分布にほとんど一致
    - $n \to \infty$ の時, 正規分布

確率密度関数

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$$

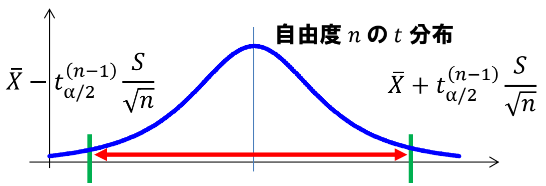


②正規母集団の母平均の区間推定(母分散未知)

- $t_\alpha^{(n-1)}$: 自由度 $n-1$ の t 分布の上側確率 $100 \times \alpha\%$ 点
  - 例 : 自由度 $n-1$ の t 分布の上側 5% 点　$t_{0.05}^{(n-1)}$
  - t 分布は対称　$t_\alpha^{(n-1)} = -t_{1-\alpha}^{(n-1)}$
- $100 \times (1-\alpha)\%$ 信頼区間の計算には $t_{\alpha/2}^{(n-1)}$ を使用

$$P\left(-t_{\alpha/2}^{(n-1)} \le T \le t_{\alpha/2}^{(n-1)}\right) = P\left(\bar{X} - t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}}\right)$$

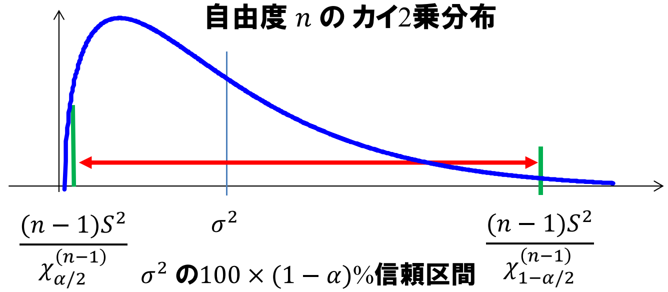
  - 例 95%信頼区間の場合　$\alpha = 0.05,\ 100 \times (1-\alpha) = 95,\ t_{\alpha/2}^{(n-1)} = t_{0.025}^{(n-1)}$



μ の $100 \times (1-\alpha)\%$ 信頼区間

③正規母集団の分散の区間推定

- $U = (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{(n-1)}$
  - $\chi_\alpha^{(n-1)}$: 自由度 $n-1$ の χ^2 分布の上側 $100 \times \alpha\%$ 点
  - $P\left(\chi_{1-\alpha/2}^{(n-1)} \le U \le \chi_{\alpha/2}^{(n-1)}\right) = P\left(\frac{(n-1)S^2}{\chi_{\alpha/2}^{(n-1)}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi_{1-\alpha/2}^{(n-1)}}\right)$



σ^2 の $100 \times (1-\alpha)\%$ 信頼区間

演習問題

- **東証株価指数**(TOPIX)**の日次収益率は正規分布に従うと仮定する**
  $t^{(3)}_{0.025}=3.18$, $t^{(4)}_{0.025}=2.78$, $t^{(3)}_{0.05}=2.35$, $t^{(4)}_{0.05}=2.13$, $Z_{0.025}=1.96$, $Z_{0.05}=1.65$, $\chi^{(3)}_{0.025}$=9.35, $\chi^{(4)}_{0.025}$=11.14, $\chi^{(3)}_{0.05}$=7.18, $\chi^{(4)}_{0.05}$=9.49, $\chi^{(3)}_{0.975}$=0.22, $\chi^{(4)}_{0.975}$=0.48, $\chi^{(3)}_{0.95}$=0.35, $\chi^{(4)}_{0.95}$=0.71

1. **直近の**4**日間の日次収益率は** $\bar{X}=0, S^2=1$ **であった。母平均** μ **の**95%**信頼区間を求めなさい**
2. **このとき**, **母分散** σ^2 **の**95%**信頼区間を求めなさい**
3. **約**1**年間**(256**日**)**の日次収益率は** $\bar{X}=0, S^2=1$**であった。母平均** μ **の**95%**信頼区間を求めなさい**

演習問題

- **東証株価指数**(TOPIX)**の日次収益率は正規分布に従うと仮定する**
  $t^{(8)}_{0.025}=2.31$, $t^{(9)}_{0.025}=2.26$, $t^{(8)}_{0.01}=2.90$, $t^{(9)}_{0.01}=2.82$, $t^{(8)}_{0.005}=3.36$, $t^{(9)}_{0.005}=3.25$, $\chi^{(8)}_{0.01}$= 20.09, $\chi^{(9)}_{0.01}$= 21.67, $\chi^{(8)}_{0.005}$= 21.96, $\chi^{(9)}_{0.005}$= 23.59, $\chi^{(8)}_{0.99}$= 1.65, $\chi^{(9)}_{0.99}$= 2.09, $\chi^{(8)}_{0.995}$= 1.34, $\chi^{(9)}_{0.995}$= 1.74

1. **直近の**9**日間の日次収益率は** $\bar{X}=0, S^2=1$ **であった。母平均** μ **の**99%**信頼区間を求めなさい**
2. **このとき**, **母分散** σ^2 **の**99%**信頼区間を求めなさい**