確率統計基礎講義 A

若木宏文

wakaki@hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/index.shtml

2019.5.29

目 次

4 t-分布と F-分布
4.1. t-分布

μ の信頼区間

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ のとき, μ の信頼区間を考える.
σ^2 が既知の場合：$\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$ とすると, $\bar{X}$ は平均 μ の推定量である. このとき

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1)$$

であるから, $N(0, 1)$ の上側 2.5%点である 1.96 を用いて

$$\mathrm{P}(|\sqrt{n}(\bar{X} - \mu)/\sigma| \leq 1.96) = 0.95$$

であるから, $\bar{X} - \frac{\sigma}{\sqrt{n}}1.96 \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}1.96$ が成り立つ確率が 95% となる.
区間 $\left[\bar{X} - \frac{\sigma}{\sqrt{n}}1.96,\ \bar{X} + \frac{\sigma}{\sqrt{n}}1.96\right]$ は μ の信頼係数 95%の信頼区間となる.

μ の信頼区間 (続き)

σ^2 が未知の場合：σ^2 を，その推定量
$S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ で置き換えた区間を考える．

$$\mathsf{P}\left(\mu \in \left[\bar{X} - \frac{S}{\sqrt{n}}1.96,\ \bar{X} + \frac{S}{\sqrt{n}}1.96\right]\right) = \mathsf{P}\left(\left|\frac{\sqrt{n}(\bar{X}-\mu)}{S}\right| \leq 1.96\right)$$

n を大きくすると，$\sqrt{n}(\bar{X}-\mu)/S$ は $N(0,1)$ に分布収束するので，n が大きければ，信頼係数がほぼ 95%の信頼区間が得られる．

n が大きくない場合に信頼係数が 95% となるような区間を求めるためには，1.96 を $T = \sqrt{n}(\bar{X}-\mu)/S$ の分布の上側 2.5%点で置き換える必要がある．

定義

t-分布

t-分布は 1908 年にギネスビール社で働くゴセット (Gosset, 1876-1936) により発見された．ギネスビール社は社員が学術雑誌に論文を投稿することを禁じていたので，ゴセットはスチューデント (Student) という偽名を使って t-分布に関する論文を発表した．そのため，現在でも t-分布のことをスチューデントの t-分布と呼ぶことがある．

定義 Z と Y が独立で，$Z \sim N(0,1)$，$Y \sim \chi_n^2$ であるとき

$$T = \frac{Z}{\sqrt{Y/n}}$$

の従う分布を**自由度 n の t-分布**といい，$T \sim t_n$ と書く．

t-分布に関する定理 1

定理 4.1.1. $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ のとき

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

t-分布の確率密度関数

$T \sim t_n$ の確率密度関数は

$$f(t;n) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma[n/2]}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \quad (n>0)$$

注 4.1.1) この確率密度関数によって t_n を定義してもよい. この場合, n は自然数でなくてもよい.

Memo

t-分布に関する定理 2

定理 4.1.2.

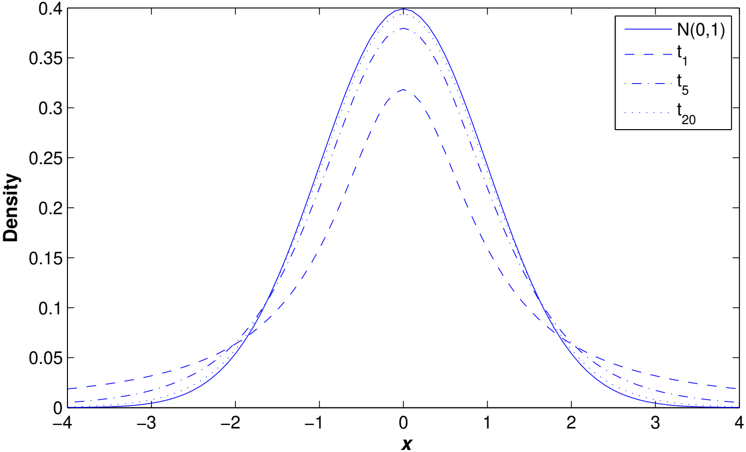
$$^\forall t \in \mathbb{R},\ \lim_{n\to\infty} f(t;n) = \phi(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

$$^\forall A \subset \mathbb{R}\,(\text{可測集合}),\ \lim_{n\to\infty}\int_A f(t;n)dt = \int_A \phi(t)dt$$

Memo

t-分布の特性

1. t_n は 標準正規分布と同様に 0 を中心に左右対称.



分布の特性（続き）

2. $T \sim t_n$ のとき

$$\mathsf{E}[T] = 0 \ (n \geq 2), \quad \mathsf{Var}[T] = \frac{n}{n-2} \ (n > 2)$$

$n > r$ ならば

$$\mathsf{E}[T^r] = \begin{cases} 0 & (r \text{ が奇数 }) \\ \dfrac{1 \times 3 \times \cdots \times (r-1) n^{r/2}}{(n-2)(n-4)\cdots(n-r)} & (r \text{ が偶数 }) \end{cases}$$

自由度が 1 である t-分布を特にコーシー分布と呼ぶ．コーシー分布はモーメントが存在しない分布の代表的なものである．

Memo

分布の特性（続き）

3. $T \sim t_1$ (コーシー分布) の特性関数は

$$\mathcal{C}_T(t) = e^{-|t|}$$

自由度が奇数のときは留数定理で求められるが, 偶数のときは簡単な形に書けない.

Memo