

T分布に関する証明

- Studentの定理
- Studentの定理の証明（あらすじ）
- Studentの定理の証明（詳細）
- 回帰係数とT分布の関係の定式化
- 回帰係数とT分布の関係の証明

この記事ではT分布に関するいくつかの重要事項を証明する。

とくにユニークなのは5章の、単回帰の誤差の標本分散が自由度 $n - 2$ のカイ二乗分布に従うことの証明方法。

自由度についての証明は統計学の入門的教科書ではごまかしてやり過ごされることが多い。また、学部上級生レベルの数理統計学の教科書では特性関数、二次形式、トレース計算が用いられることが多いけど、この記事ではこれらを用いずに証明する方法を与える。

T分布の定義

自由度 n のT分布は以下の確率密度関数として定義される。

$$f_n(t) = \frac{1}{\sqrt{\pi n}} \cdot \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \cdot (1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$$

1. Studentの定理

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本平均を $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i$ と置き、不偏分散を $s_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^2$ と置く。ここで次の変数

$$T = \frac{X - \mu}{\sqrt{s_n^2/n}}$$

を考えると、これは自由度 n のT分布に従う。

2. Studentの定理の証明

以下の4つの補題を用いて証明する。

補題2.3より、補題2.2と補題2.4に示す変数に対して補題2.1を適用することができ、Studentの定理が示される。つまり、

$$T = \frac{\frac{X - \mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)s_n^2/\sigma^2}{n-1}}} = \frac{\bar{X} - \mu}{\sqrt{s_n^2/n}}$$

は自由度 n のT分布に従い、Studentの定理が成立する。

2.1 補題2.1

変数 U が正規分布 $N(0, 1^2)$ に従い、変数 V がカイ二乗分布 $\chi^2(k)$ に従い、これらが互いに独立なとき、次の変数

$$Z = \frac{U}{\sqrt{V/k}}$$

を考えると、これは自由度 k のT分布 $T(k)$ に従う。

2.2 補題2.2

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本平均を $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ と置く。ここで次の変数

$$\frac{\overline{X} - \mu}{\sqrt{\sigma^2/n}}$$

を考えると、これは正規分布 $N(0, 1^2)$ に従う。

2.3 補題2.3

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本平均を $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ 、標本分散を $s_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ と置く。

このとき X と s_n^2 は互いに独立である。

2.4 補題2.4

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本分散を $s_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$ と置く。ここで次の変数

$$\frac{(n-1)s_n^2}{\sigma^2}$$

を考えると、これは自由度 $n-1$ のカイ二乗分布 $\chi^2(n-1)$ に従う。

3. 補題の証明

以下では、前述の4つの補題を証明する。

3.1 補題2.1の証明

補題2.1

変数 U が正規分布 $N(0, 1^2)$ に従い、変数 V がカイ二乗分布 $\chi^2(k)$ に従い、これらが互いに独立なとき、次の変数

$$Z = \frac{U}{\sqrt{V/k}}$$

を考えると、これは自由度 k の T 分布 $T(k)$ に従う。

正規分布とカイ二乗分布の確率密度関数を $\phi(u)$ 、 $\psi(v)$ とする。

定数 a に対して $Z < a$ となる確率は

$$\begin{aligned} P(Z < a) &= P\left(\frac{U}{\sqrt{V/k}} < a\right) = \iint_{\frac{u}{\sqrt{v/k}} < a} \phi(u)\psi(v)dudv \\ &= \iint_{\frac{u}{\sqrt{v/k}} < a} \left(\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\right) \left(\frac{1}{\Gamma(\frac{k}{2})}\left(\frac{1}{2}\right)^{\frac{k}{2}}v^{\frac{k}{2}-1}e^{-\frac{v}{2}}\right)dudv \end{aligned}$$

となる。 $z = \frac{u}{\sqrt{v/k}}$ 、 $v = v$ により変数変換を行い、

$$P(Z < a) = \iint_{z < a, 0 < v < \infty} \frac{1}{\sqrt{2\pi}2^{\frac{k}{2}}\Gamma(\frac{k}{2})} \cdot e^{-\frac{vz^2}{2} - \frac{v}{2}} \cdot v^{\frac{k}{2}-1} \cdot \left| \frac{\sqrt{v/k}}{0} \quad \frac{*}{1} \right| dv dz$$

さらにガンマ関数の形を作るように $s = \frac{vz^2}{2k} + \frac{v}{2}$ 、 $z = z$ と変数変換することで、

$$\begin{aligned} P(Z < a) &= \iint_{z < a, 0 < s < \infty} \frac{1}{\sqrt{2\pi}\sqrt{k}2^{\frac{k}{2}}\Gamma(\frac{k}{2})} \cdot e^{-s} \cdot \left(\frac{s}{\frac{z^2}{2k} + \frac{1}{2}}\right)^{\frac{k-1}{2}} \cdot \left| \frac{\frac{1}{\frac{z^2}{2k} + \frac{1}{2}}}{0} \quad \frac{*}{1} \right| ds dz \\ &= \iint_{z < a, 0 < s < \infty} \frac{1}{\sqrt{k\pi}2^{\frac{k+1}{2}}\Gamma(\frac{k}{2})} \cdot s^{\frac{k+1}{2}-1} \cdot e^{-s} \cdot \frac{1}{\left(\frac{z^2}{2k} + \frac{1}{2}\right)^{\frac{k+1}{2}}} ds dz \\ &= \frac{\Gamma(\frac{k}{2} + 1)}{\sqrt{\pi k}\Gamma(\frac{k}{2})} \int_{z < a} \left(\frac{z^2}{k} + 1\right)^{-\frac{k+1}{2}} dz \end{aligned}$$

よってこのとき Z は自由度 k の T 分布 $T(k)$ に従う。

3.2 補題2.2の証明

補題2.2

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本平均を $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ と置く。ここで次の変数

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$$

を考えると、これは正規分布 $N(0, 1^2)$ に従う。

示すべき命題「 $\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ は正規分布 $N(0, 1^2)$ に従う」は、命題「 $\sum_{i=1}^n X_i$ は正規分布 $N(n\mu, n\sigma^2)$ に従う」と同値である。なぜならば、変数 U と V を $U = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}, V = \sum_{i=1}^n X_i$ と置くと $U = \frac{V - n\mu}{\sqrt{n\sigma^2}}$ であり、

$$P(V < n\mu + a\sqrt{n\sigma^2}) = P(U < a)$$

$$\int_{v < n\mu + a\sqrt{n\sigma^2}} \frac{1}{\sqrt{2\pi}\sqrt{n\sigma^2}} e^{-\frac{(v-n\mu)^2}{2n\sigma^2}} dv = \int_{u < a} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

であるから、上記2つの式の右辺同士が等しいとき左辺同士も等しく、その逆も成立するためである。

後者の命題を証明するには「変数 V と W が正規分布 $N(\mu_1, \sigma_1^2)$ と $N(\mu_2, \sigma_2^2)$ に互いに独立に従うとき、変数 $V+W$ は正規分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ に従う」を示せば十分である。

変数 $Y = V + W$ とし、 V と W の確率密度関数を ϕ_v, ϕ_w とすると

$$P(Y < a) = \int_{y=-\infty}^a \int_{v=-\infty}^{\infty} \phi_v(v) \phi_w(y-v) dv dy = \int_{y=-\infty}^a \int_{v=-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(v-\mu_1)^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-v-\mu_2)^2}{2\sigma_2^2}} dv dy$$

$$= \int_{y=-\infty}^a \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} e^{-\frac{(y-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}} \int_{v=-\infty}^{\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} (v - \frac{\sigma_1^2 y + \sigma_2^2 \mu_1 - \sigma_1^2 \mu_2}{\sigma_1^2 + \sigma_2^2})^2} dv dy$$

$$= \int_{y=-\infty}^a \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} e^{-\frac{(y-(\mu_1+\mu_2))^2}{2(\sigma_1^2+\sigma_2^2)}} \sqrt{2\pi \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}} dy$$

$$= \int_{y=-\infty}^a \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{(y-(\mu_1+\mu_2))^2}{2(\sigma_1^2 + \sigma_2^2)}} dy$$

よって、変数 $V+W$ は正規分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ に従う。

すなわち、 $\sum_{i=1}^n X_i$ は正規分布 $N(n\mu, n\sigma^2)$ に従う。

すなわち、 $\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}$ は正規分布 $N(0, 1^2)$ に従う。

3.3 補題2.3の証明

補題2.3

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本平均を $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ 、標本分散を $s_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ と置く。

このとき X と s_n^2 は互いに独立である。

n 次元ベクトルと n 次元直交行列

$$U = \begin{pmatrix} \frac{X_1 - \mu}{\sigma} \\ \vdots \\ \frac{X_n - \mu}{\sigma} \end{pmatrix}, A = \begin{pmatrix} 1/\sqrt{n} & \dots & 1/\sqrt{n} \\ * & \dots & * \\ \vdots & & \vdots \\ * & \dots & * \end{pmatrix}$$

を考える。(Aのような直交行列が存在することは、グラム・シュミットの方法で実際に書き下せば分かるがここでは省略する。)

UとAにより、 n 次元ベクトル

$$V = \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix} = AU$$

を定義する。

このとき、

$$\sum_{i=1}^n V_i^2 = V^T V = (AU)^T (AU) = U^T (A^T A) U = U^T U = \sum_{i=1}^n U_i^2$$

となることを利用して、下記のように \bar{X} と s_n^2 を表すことができる。

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^n X_i}{n} = \frac{\sum_{i=1}^n (\mu + \sigma U_i)}{n} = \mu + \frac{\sigma}{\sqrt{n}} \sum_{i=1}^n \frac{U_i}{\sqrt{n}} = \mu + \frac{\sigma}{\sqrt{n}} V_1 \\ s_n^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{\sigma}{n-1} \sum_{i=1}^n (U_i - U)^2 = \frac{\sigma}{n-1} \left(\sum_{i=1}^n U_i^2 - nU^2 \right) \\ &= \frac{\sigma}{n-1} \left(\sum_{i=1}^n U_i^2 - \left(\sum_{i=1}^n U_i / \sqrt{n} \right)^2 \right) = \frac{\sigma}{n-1} \left(\sum_{i=1}^n V_i^2 - V_1^2 \right) = \frac{\sigma}{n-1} \sum_{i=2}^n V_i^2\end{aligned}$$

ここで、 $\bar{U} = \sum_{i=1}^n U_i / n$ である。

\bar{X} は V_1 で表され、 s_n^2 は V_2, \dots, V_n で表されるので、 \bar{X} と s_n^2 が独立であることを示すには、 V_i たちが互いに独立であることを示せば十分である。

補題2.2の証明の中で用いた方法と同様にして、 U_i, V_i は正規分布 $N(0, 1^2)$ に従うことが分かる。

任意の領域 $D_1 \subset \mathbb{R}^n$ について、直交行列 A^T を左から掛ける写像によって D_1 が移される領域を D_2 とする。この写像には逆写像があるので

$$U \in D_2 \Leftrightarrow V \in D_1$$

である。このとき、

$$\begin{aligned}P(V \in D_1) &= P(U \in D_2) = \int_{u \in D_2} \frac{1}{\sqrt{2\pi}} e^{u_1^2/2} \dots \frac{1}{\sqrt{2\pi}} e^{u_n^2/2} du_1 \dots du_n \\ &= \int_{u \in D_2} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{\frac{u_1^2 + \dots + u_n^2}{2}} du_1 \dots du_n \\ &= \int_{v \in D_1} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{\frac{v_1^2 + \dots + v_n^2}{2}} |A| dv_1 \dots dv_n \\ &= \int_{v \in D_1} \frac{1}{\sqrt{2\pi}} e^{v_1^2/2} \dots \frac{1}{\sqrt{2\pi}} e^{v_n^2/2} dv_1 \dots dv_n\end{aligned}$$

よって、 V の確率密度関数は各 V_i の確率密度関数の積となるので、 V_i たちは互いに独立である。

すなわち、 \bar{X} と s_n^2 は独立な変数でそれぞれ表されるので、 \bar{X} と s_n^2 は互いに独立である。

3.4 補題2.4の証明

補題2.4

X_1, \dots, X_n が正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数であるとする。また標本分散を $s_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$ と置く。ここで次の変数

$$\frac{(n-1)s_n^2}{\sigma^2}$$

を考えると、これは自由度 $n-1$ のカイ二乗分布 $\chi^2(n-1)$ に従う。

下記(1)(2)(3)より、変数 $\frac{(n-1)s_n^2}{\sigma^2}$ は自由度 $n-1$ のカイ二乗分布 $\chi^2(n-1)$ に従うことが示される。

(1). 変数 X が正規分布 $N(0, 1^2)$ に従うとき、 X^2 はカイ二乗分布 $\chi^2(1)$ に従う。

(2). 変数 W_1, \dots, W_n が互いに独立にカイ二乗分布 $\chi^2(1)$ に従うとき、変数 $W_1 + \dots + W_n$ はカイ二乗分布 $\chi^2(n)$ に従う。

(3). 変数 $\frac{(n-1)s_n^2}{\sigma^2}$ は、正規分布 $N(0, 1^2)$ に従う $n-1$ 個の互いに独立な変数の二乗和である。

以下では(1)(2)(3)について証明する。

(1)の証明：

変数 $W = X^2$ とする。

$$P(W < a) = P(-\sqrt{a} < X < \sqrt{a}) = \int_{-\sqrt{a}}^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2 \int_0^{\sqrt{a}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

\$u = x^2\$として変数変換すると

$$P(W < a) = 2 \int_0^a \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \frac{du}{2\sqrt{u}} = \int_0^a \frac{1}{\sqrt{\pi}} \left(\frac{1}{2}\right)^{-\frac{1}{2}} u^{-\frac{1}{2}} e^{-\frac{u}{2}} du$$

となる。すなわち、 $W = X^2$ はカイ二乗分布 $\chi^2(1)$ に従う。

(2)の証明：

変数 W_1 がカイ二乗分布 $\chi^2(n)$ に従い、変数 W_2 がカイ二乗分布 $\chi^2(m)$ に従うとき、変数 $W_1 + W_2$ はカイ二乗分布 $\chi^2(n + m)$ に従うことを示せば十分である。

変数 $Z = W_1 + W_2$ とし、自由度 k のカイ二乗分布を f_k と表すとすると

$$\begin{aligned} P(Z < a) &= \int_{z=0}^a \int_{w_1=0}^{\infty} f_n(w_1) f_m(z - w_1) dw_1 dz \\ &= \int_{z=0}^a \int_{w_1=0}^{\infty} \left(\frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}\right)^{-\frac{n}{2}} w_1^{\frac{n}{2}-1} e^{-\frac{w_1}{2}} \right) \left(\frac{1}{\Gamma(\frac{m}{2})} \left(\frac{1}{2}\right)^{-\frac{m}{2}} (z - w_1)^{\frac{m}{2}-1} e^{-\frac{z-w_1}{2}} \right) dw_1 dz \end{aligned}$$

となる。 w_1 の積分において $u = w_1/z$ と変数変換すると

$$\begin{aligned} P(Z < a) &= \int_{z=0}^a \int_{u=0}^1 \left(\frac{1}{\Gamma(\frac{n}{2})} \left(\frac{1}{2}\right)^{-\frac{n}{2}} (zu)^{\frac{n}{2}-1} e^{-\frac{zu}{2}} \right) \left(\frac{1}{\Gamma(\frac{m}{2})} \left(\frac{1}{2}\right)^{-\frac{m}{2}} (z - zu)^{\frac{m}{2}-1} e^{-\frac{z-zu}{2}} \right) z du dz \\ &= \int_{z=0}^a \frac{1}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{1}{2}\right)^{-\frac{n+m}{2}} z^{\frac{n+m}{2}-1} e^{-\frac{z}{2}} \int_{u=0}^1 u^{\frac{n}{2}-1} (1-u)^{\frac{m}{2}-1} du dz \end{aligned}$$

となり、 u の積分においてベータ関数の性質を利用して、

$$\begin{aligned} P(Z < a) &= \int_{z=0}^a \frac{1}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} \left(\frac{1}{2}\right)^{-\frac{n+m}{2}} z^{\frac{n+m}{2}-1} e^{-\frac{z}{2}} \frac{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}{\Gamma(\frac{n+m}{2})} dz \\ &= \int_{z=0}^a \frac{1}{\Gamma(\frac{n+m}{2})} \left(\frac{1}{2}\right)^{-\frac{n+m}{2}} z^{\frac{n+m}{2}-1} e^{-\frac{z}{2}} dz \end{aligned}$$

となる。よって、変数 W_1 がカイ二乗分布 $\chi^2(n)$ に従い、変数 W_2 がカイ二乗分布 $\chi^2(m)$ に従うとき、変数 $Z = W_1 + W_2$ はカイ二乗分布 $\chi^2(n + m)$ に従うことを示せた。

すなわち、変数 W_1, \dots, W_n が互いに独立にカイ二乗分布 $\chi^2(1)$ に従うとき、変数 $W_1 + \dots + W_n$ はカイ二乗分布 $\chi^2(n)$ に従う。

(3)の証明：

補題2.3の証明の中で定義した変数 V_1, \dots, V_n について、これらは互いに独立に正規分布 $N(\mu, \sigma^2)$ に従い、また

$$s_n^2 = \frac{\sigma^2}{n-1} \sum_{i=2}^n V_i^2$$

である。すなわち、変数 $\frac{(n-1)s_n^2}{\sigma^2}$ は、正規分布 $N(0, 1^2)$ に従う $n - 1$ 個の互いに独立な変数の二乗和である。

4.回帰係数とT分布の関係

ここではまず回帰問題の定式化を行い、そのあと回帰係数と T 分布の関係について記す。

無作為に選ばれた値とそれに対応する実験によって得た値の組 $(X_1, Y_1), \dots, (X_n, Y_n)$ について、以下のような関係があるとする。

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 - \overline{X} \\ \vdots & \vdots \\ 1 & X_n - \overline{X} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

ここで

- 説明変数 X_i は互いに異なる値をとる。(これらは確率変数ではなく定数である。)

- 説明変数の平均値を $X = \frac{\sum_{i=1}^n X_i}{n}$ とする
- 確率変数 $\epsilon_1, \dots, \epsilon_n$ は真の誤差と呼ばれ、正規分布 $N(0, \sigma^2)$ に従う独立な確率変数である
- 定数 β_0, β_1 は真の回帰係数と呼ばれる

とする。このとき、被説明変数 Y_i は互いに独立に $N(\beta_0 + (X_i - \bar{X})\beta_1, \sigma^2)$ に従う。

回帰係数は最尤法によって推定することとし、

$$L(\beta_0, \beta_1) = \left(\frac{1}{2\pi\sigma^2} \right)^{n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1(X_i - \bar{X}))^2 \right\}$$

について、 β_0, β_1 の推定値を b_0, b_1 とすると

$$\frac{\partial}{\partial \beta_0} \log L(b_0, b_1) = \frac{\partial}{\partial \beta_1} \log L(b_0, b_1) = 0$$

を満たす。すなわち、 $\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n}$ として

$$b_0 = \bar{Y}$$

$$b_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})Y_i}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})\epsilon_i}{\sum_{i=1}^n (X_i - \bar{X})^2} + \beta_1$$

である。

このような状況において、 $\hat{\epsilon}_i$ たちを

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 - \bar{X} \\ \vdots & \vdots \\ 1 & X_n - \bar{X} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

として定める。

以上のような回帰問題において、

$$T_0 = \frac{b_0 - \beta_0}{\sqrt{\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{n(n-2)}}}$$

$$T_1 = \frac{b_1 - \beta_1}{\sqrt{\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{(n-2) \sum_{i=1}^n (X_i - \bar{X})^2}}}$$

を満たす変数 T_0, T_1 は自由度 $n - 2$ の T 分布 $T(n - 2)$ にそれぞれ従う。

5. 回帰係数とT分布の関係の証明

前章の変数 T_0, T_1 が T 分布に従うことを証明する。

ここで、補題2.1を再び利用する。

2.1 補題

変数 U が正規分布 $N(0, 1^2)$ に従い、変数 V がカイ二乗分布 $\chi^2(k)$ に従い、これらが互いに独立なとき、次の変数

$$Z = \frac{U}{\sqrt{V/k}}$$

を考えると、これは自由度 k の T 分布 $T(k)$ に従う。

補題2.1の U, V, k として適切なものを選ぶことで、 T 分布に従う変数 T_0, T_1 を導出する。

4章で b_0, b_1 を表した式より、 b_0 は正規分布 $N(\beta_0, \frac{\sigma^2}{n})$ に従い、 b_1 は正規分布 $N(\beta_1, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2})$ に従うことが分かる。

つまり変数 $\frac{b_0 - \beta_0}{\sqrt{\frac{\sigma^2}{n}}}$ と $\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}}$ は正規分布 $N(0, 1^2)$ に従う（これらは補題2.2の証明前半部分と同様に示すことができる）。

ここで、変数

$$\frac{\sum_1^n \epsilon_i^2}{\sigma^2}$$

を考え、これが前記の $\frac{b_0 - \beta_0}{\sqrt{\frac{\sigma^2}{n}}}$ と $\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_1^n (X_i - \bar{X})^2}}}$ とそれぞれ互いに独立で自由度 $n - 2$ のカイ二乗分布 $\chi^2(n - 2)$ に従うことが分かれば、すなわち、補題2.1より変数

$$T_0 = \frac{\frac{b_0 - \beta_0}{\sqrt{\frac{\sigma^2}{n}}}}{\sqrt{\frac{\sum_1^n \epsilon_i^2 / \sigma^2}{n - 2}}} = \frac{b_0 - \beta_0}{\sqrt{\frac{\sum_1^n \epsilon_i^2}{n(n - 2)}}}$$

$$T_1 = \frac{\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_1^n (X_i - \bar{X})^2}}}}{\sqrt{\frac{\sum_1^n \epsilon_i^2 / \sigma^2}{n - 2}}} = \frac{b_1 - \beta_1}{\sqrt{\frac{\sum_1^n \epsilon_i^2}{(n - 2) \sum_1^n (X_i - \bar{X})^2}}}$$

はそれぞれ自由度 $n - 2$ の T 分布 $T(n - 2)$ に従う。

これらより、以下の補題5.1を証明すればよい。

5.1 補題5.1

前記の回帰問題において、変数

$$\frac{\sum_1^n \epsilon_i^2}{\sigma^2}$$

は、 $\frac{b_0 - \beta_0}{\sqrt{\frac{\sigma^2}{n}}}$ と $\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{\sum_1^n (X_i - \bar{X})^2}}}$ とそれぞれ互いに独立で自由度 $n - 2$ のカイ二乗分布 $\chi^2(n - 2)$ に従う。

補題5.1の証明

まず、次のような式が成り立つ。

$$\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \begin{pmatrix} 1 & X_1 - \bar{X} \\ \vdots & \vdots \\ 1 & X_n - \bar{X} \end{pmatrix} \begin{pmatrix} b_0 - \beta_0 \\ b_1 - \beta_1 \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

$$= (b_0 - \beta_0) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + (b_1 - \beta_1) \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

正確な表現ではないが直感的な解説をしておくと、左辺はカイ二乗分布 $\chi^2(n)$ に従う変数の素(もと)であり、右辺第1項、第2項、第3項はそれぞれカイ二乗分布 $\chi^2(1)$, $\chi^2(1)$, $\chi^2(n - 2)$ に従う変数の素であるということ意識しながら式を変形していく。

b_0, b_1 の定義を代入してさらに式変形すると

$$\begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} = \bar{Y} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \frac{\sum_1^n \epsilon_i (X_i - \bar{X})}{\sum_1^n (X_i - \bar{X})^2} \begin{pmatrix} X_1 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

$$= \sum_1^n \frac{\epsilon_i}{\sqrt{n}} \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix} + \sum_1^n \frac{\epsilon_i (X_i - \bar{X})}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \begin{pmatrix} \frac{X_1 - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \\ \vdots \\ \frac{X_n - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

$$= \left(\frac{1}{\sqrt{n}} \quad \cdots \quad \frac{1}{\sqrt{n}} \right) \cdot \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix} + \left(\frac{X_1 - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \quad \cdots \quad \frac{X_n - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \right) \cdot \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \begin{pmatrix} \frac{X_1 - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \\ \vdots \\ \frac{X_n - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}$$

n 次元ベクトル

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \hat{\epsilon} = \begin{pmatrix} \hat{\epsilon}_1 \\ \vdots \\ \hat{\epsilon}_n \end{pmatrix}, a_1 = \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix}, a_2 = \begin{pmatrix} \frac{X_1 - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \\ \vdots \\ \frac{X_n - \bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2}} \end{pmatrix}$$

と置くと、上式は

$$\epsilon = (a_1^T \cdot \epsilon) a_1 + (a_2^T \cdot \epsilon) a_2 + \hat{\epsilon}$$

と書ける。

a_1 と a_2 は直交する長さ1のn次元ベクトルであるので、 a_1 と a_2 を含む正規直交ベクトル a_1, a_2, \dots, a_n をとることができる（このような直交座標系が存在することは、グラム・シュミットの方法で実際に書き下せば分かるがここでは省略する）。またこのときn行n列行列 $A = (a_1 a_2 \dots a_n)^T$ は直交行列になる。

ここで、新たに z_i を

$$\begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} = A\epsilon = \begin{pmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{X_1 - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} & \cdots & \frac{X_n - \bar{X}}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

と定める。 ϵ_i は互いに独立に正規分布 $N(0, \sigma^2)$ に従うことと、 a_1, a_2, \dots, a_n が正規直交ベクトルであることから、 z_i は正規分布 $N(0, \sigma^2)$ に従う（これは補題2.2の証明と同様に示すことができる）。また z_i は互いに独立である（これは補題2.3で V_i の独立性を証明したのと同様に示すことができる）。

これらより、 ϵ に左から A をかけ、変形すると

$$\begin{aligned} A\epsilon &= A(a_1^T \cdot \epsilon) a_1 + A(a_2^T \cdot \epsilon) a_2 + A\hat{\epsilon} \\ \begin{pmatrix} a_1^T \cdot \epsilon \\ \vdots \\ a_n^T \cdot \epsilon \end{pmatrix} &= (a_1^T \cdot \epsilon) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + (a_2^T \cdot \epsilon) \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + A\hat{\epsilon} \\ \begin{pmatrix} 0 \\ 0 \\ z_3 \\ \vdots \\ z_n \end{pmatrix} &= A\hat{\epsilon} \end{aligned}$$

となる。よって、

$$\sum_{i=1}^n \hat{\epsilon}_i^2 = \|\hat{\epsilon}\|^2 = \hat{\epsilon}^T \hat{\epsilon} = \hat{\epsilon}^T (A^T A) \hat{\epsilon} = (A\hat{\epsilon})^T (A\hat{\epsilon}) = \|A\hat{\epsilon}\|^2 = \sum_{i=3}^n z_i^2$$

が成立する。 z_i/σ は互いに独立に正規分布 $N(0, 1^2)$ に従うので、

$$\frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sigma^2} = \sum_{i=3}^n \left(\frac{z_i}{\sigma}\right)^2$$

は $n-2$ 個の互いに独立に正規分布 $N(0, 1^2)$ に従う変数の二乗和で表されており、補題2.4(1),(2)より自由度 $n-2$ のカイ二乗分布 $\chi^2(n-2)$ に従う。

よって補題5.1は成立する。

すなわち、前記のように定義した T_0, T_1 は自由度 $n-2$ の T 分布 $T(n-2)$ に従う。

参考文献

- Robert V. Hogg, Joseph McKean, Allen T Craig “Introduction to Mathematical Statistics, Fourth Edition”
- 伏見正則「確率と確率過程」
- 松原望「入門統計解析」
- [不偏分散と自由度n-1のカイ二乗分布 | 高校数学の美しい物語](#)
- [正規母集団から得られる標本平均と標本分散は独立であることの証明 | 理工系数学のアラカルト](#)

