t分布

渡邉 俊夫

要点

母平均 μ、母分散 σ^2 の正規分布 $f(x_i) = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ にしたがう母集団からランダムに抽出したサイズ n の標本 $x_i\,(i = 1, 2, \cdots, n)$ に対して、標本平均を $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ 、不偏分散を $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ とすると、$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$ は自由度 $n-1$ のt分布

$$f_{n-1}(t) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \left(1 + \frac{t^2}{n-1}\right)^{-n/2}$$

にしたがう。

t分布の導出

正規分布 $f(x_i) = \frac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ （平均：μ、分散：σ^2）

母平均 μ、母分散 σ^2 の正規分布にしたがう母集団からランダムに抽出したサイズ n の標本 x_i に対して、標本平均を $\bar{x}$、不偏分散を s^2 とすると、

$$X = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

は標準正規分布にしたがい（pp. 24-25）、

標本平均 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

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

$$Y = \frac{(n-1)s^2}{\sigma^2}$$

は自由度 $n-1$ のカイ2乗分布にしたがう（pp. 26-30）。このとき、

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{X}{\sqrt{Y/(n-1)}}$$

t は母標準偏差 σ を含まず、標本標準偏差 s で表される

が自由度 $n-1$ のt分布にしたがうことになる。

t分布の導出

確率変数 X は標準正規分布

$$f(X)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-X^2/2}$$

にしたがい、確率変数 Y は自由度 $m=n-1$ のカイ2乗分布（χ^2分布）

$$f_m(Y)=\chi_m^2(Y)=\frac{Y^{(m/2)-1}\mathrm{e}^{-Y/2}}{2^{m/2}\Gamma(m/2)}$$

にしたがう。確率変数 X,Y は独立であるから、

$$f(X,Y)=f(X)f_m(Y)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-X^2/2}\frac{Y^{(m/2)-1}\mathrm{e}^{-Y/2}}{2^{m/2}\Gamma(m/2)}$$

$$=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{m/2}\Gamma(m/2)}Y^{(m/2)-1}\mathrm{e}^{-(X^2+Y)/2}$$

t分布の導出

$$f(X,Y)=\frac{1}{\sqrt{2\pi}}\frac{1}{2^{m/2}\Gamma(m/2)}Y^{(m/2)-1}\mathrm{e}^{-(X^2+Y)/2}$$

ここで、

$$t=\frac{X}{\sqrt{Y/m}},\quad y=Y$$

$$\therefore X=t\sqrt{\frac{Y}{m}}=t\sqrt{\frac{y}{m}},\quad X^2=\frac{t^2y}{m}$$

と変数変換すると、$\dfrac{\partial X}{\partial t}=\sqrt{\dfrac{y}{m}}$ より $\mathrm{d}X\mathrm{d}Y=\sqrt{\dfrac{y}{m}}\mathrm{d}t\mathrm{d}y$ だから

$$f(t,y)=\frac{1}{\sqrt{2m\pi}}\frac{1}{2^{m/2}\Gamma(m/2)}y^{\frac{m+1}{2}-1}\mathrm{e}^{-\left(1+\frac{t^2}{m}\right)\frac{y}{2}}$$

$$=\frac{1}{\sqrt{m\pi}}\frac{1}{\Gamma(m/2)}\frac{1}{2}\left(\frac{y}{2}\right)^{\frac{m+1}{2}-1}\mathrm{e}^{-\left(1+\frac{t^2}{m}\right)\frac{y}{2}}$$

t分布の導出

$$f(t,y)=\frac{1}{\sqrt{m\pi}}\frac{1}{\Gamma(m/2)}\frac{1}{2}\left(\frac{y}{2}\right)^{\frac{m+1}{2}-1}\mathrm{e}^{-\left(1+\frac{t^2}{m}\right)\frac{y}{2}}$$

これを y で積分して、t の確率密度関数 $f_m(t)$ を求めると

$$f_m(t)=\int_0^\infty f(t,y)\mathrm{d}y=\frac{1}{\sqrt{m\pi}}\frac{1}{\Gamma(m/2)}\frac{1}{2}\int_0^\infty\left(\frac{y}{2}\right)^{\frac{m+1}{2}-1}\mathrm{e}^{-\left(1+\frac{t^2}{m}\right)\frac{y}{2}}\mathrm{d}y$$

ここで、$u=\left(1+\frac{t^2}{m}\right)\frac{y}{2}$ とおくと、$\mathrm{d}u=\left(1+\frac{t^2}{m}\right)\frac{\mathrm{d}y}{2}$ だから

$$f_m(t)=\frac{1}{\sqrt{m\pi}}\frac{1}{\Gamma(m/2)}\frac{1}{(1+t^2/m)^{(m+1)/2}}\int_0^\infty u^{((m+1)/2)-1}\mathrm{e}^{-u}\mathrm{d}u$$

t分布の導出

$$f_m(t) = \frac{1}{\sqrt{m\pi}} \frac{1}{\Gamma(m/2)} \frac{1}{(1+t^2/m)^{(m+1)/2}} \int_0^\infty u^{((m+1)/2)-1} \mathrm{e}^{-u} \mathrm{d}u$$

ガンマ関数の定義 $\Gamma(z) = \int_0^\infty s^{z-1} \mathrm{e}^{-s} \mathrm{d}s$ より

$$f_m(t) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}$$

これが自由度 m のt分布の確率密度関数である。

t分布：確率密度関数

自由度 m のt分布の確率密度関数

$$f_m(t) = \frac{1}{\sqrt{m\pi}}\frac{\Gamma((m+1)/2)}{\Gamma(m/2)}\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}$$

を全範囲 $(-\infty < y < \infty)$ で積分すると

$$\int_{-\infty}^{\infty} f_m(t)\mathrm{d}t = \frac{1}{\sqrt{m\pi}}\frac{\Gamma((m+1)/2)}{\Gamma(m/2)}\int_{-\infty}^{\infty}\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}\mathrm{d}t$$

$$= \frac{2}{\sqrt{m\pi}}\frac{\Gamma((m+1)/2)}{\Gamma(m/2)}\int_{0}^{\infty}\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}\mathrm{d}t$$

t分布：確率密度関数

ここで、$1+\dfrac{t^2}{m}=\dfrac{1}{u}$ とおくと、$t=\sqrt{\dfrac{m(1-u)}{u}}$ であり、$\dfrac{2t\mathrm{d}t}{m}=-\dfrac{\mathrm{d}u}{u^2}$ より

$$\mathrm{d}t=-\frac{m}{2tu^2}\mathrm{d}u=-\frac{m}{2u^2}\sqrt{\frac{u}{m(1-u)}}\mathrm{d}u=-\frac{\sqrt{m}}{2}u^{-\frac{3}{2}}(1-u)^{-\frac{1}{2}}\mathrm{d}u$$

だから

$$\int_0^\infty\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}}\mathrm{d}t=-\int_1^0 u^{\frac{m+1}{2}}\frac{\sqrt{m}}{2}u^{-\frac{3}{2}}(1-u)^{-\frac{1}{2}}\mathrm{d}u$$

$$=\frac{\sqrt{m}}{2}\int_0^1 u^{\frac{m}{2}-1}(1-u)^{-\frac{1}{2}}\mathrm{d}u$$

t分布：確率密度関数

$$\int_0^\infty \left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t = \frac{\sqrt{m}}{2}\int_0^1 u^{\frac{m}{2}-1}(1-u)^{-\frac{1}{2}}\mathrm{d}u$$

$$= \frac{\sqrt{m}}{2}\mathrm{B}\left(\frac{m}{2},\frac{1}{2}\right) = \frac{\sqrt{m}}{2}\frac{\Gamma(m/2)\Gamma(1/2)}{\Gamma((m+1)/2)}$$

$$= \frac{\sqrt{m\pi}}{2}\frac{\Gamma(m/2)}{\Gamma((m+1)/2)}$$

したがって、確率密度関数の全範囲$(-\infty < y < \infty)$での積分は

$$\int_{-\infty}^\infty f_m(t)\mathrm{d}t = \frac{2}{\sqrt{m\pi}}\frac{\Gamma((m+1)/2)}{\Gamma(m/2)}\int_0^\infty \left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t = 1$$

t分布：平均

自由度 m のt分布にしたがう t の平均は

$$\bar{t} = \int_{-\infty}^{\infty} t f_m(t) \mathrm{d}t = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \int_{-\infty}^{\infty} t \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t$$

$$= \frac{2}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \left[-\frac{m}{m-1} \left(1 + \frac{t^2}{m}\right)^{-\frac{m-1}{2}} \right]_{-\infty}^{\infty} = 0$$

である。ただし、$m = 1$ の場合は別に考える必要がある。

t分布$(m=1)$:平均

自由度 1 のt分布にしたがう t の平均は

$$\bar{t}=\int_{-\infty}^{\infty}tf_1(t)\mathrm{d}t=\frac{1}{\sqrt{\pi}}\frac{\Gamma(1)}{\Gamma(1/2)}\int_{-\infty}^{\infty}t(1+t^2)^{-1}\mathrm{d}t$$

$$=\frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\frac{t}{1+t^2}\mathrm{d}t=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{t}{1+t^2}\mathrm{d}t$$

$$=\frac{1}{\pi}\left[\frac{1}{2}\log(1+t^2)\right]_{-\infty}^{\infty}$$

となり、発散してしまう。したがって、自由度 1 のt分布(コーシー分布、ローレンツ分布)の平均は存在しない。

t分布：分散

自由度 m のt分布にしたがう t の2乗平均は

$$\overline{t^2} = \int_{-\infty}^{\infty} t^2 f_m(t)\mathrm{d}t = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t$$

ここで、$m \geq 3$ のとき

$$\int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t = 2\int_{0}^{\infty} t^2 \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t$$

$$= -2\int_{1}^{0} m\frac{1-u}{u} u^{\frac{m+1}{2}} \frac{\sqrt{m}}{2} u^{-\frac{3}{2}} (1-u)^{-\frac{1}{2}} \mathrm{d}t$$

$$= m^{\frac{3}{2}} \int_{0}^{1} u^{\frac{m}{2}-2} (1-u)^{\frac{1}{2}} \mathrm{d}t = m^{\frac{3}{2}} \mathrm{B}\left(\frac{m-2}{2}, \frac{3}{2}\right)$$

$1 + \frac{t^2}{m} = \frac{1}{u}$ とおくと、

$\mathrm{d}t = -\frac{\sqrt{m}}{2} u^{-\frac{3}{2}} (1-u)^{-\frac{1}{2}} \mathrm{d}u$

t分布：分散

自由度 m のt分布にしたがう t の2乗平均は、$m \geq 3$ のとき

$$\overline{t^2} = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \int_{-\infty}^{\infty} t^2 \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \mathrm{d}t$$

$$= \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \cdot m^{\frac{3}{2}} \mathrm{B}\left(\frac{m-2}{2}, \frac{3}{2}\right)$$

$$= m \frac{\Gamma((m+1)/2)}{\sqrt{\pi}\Gamma(m/2)} \frac{\Gamma((m-2)/2)\Gamma(3/2)}{\Gamma((m+1)/2)} = \frac{m}{m-2}$$

$$\Gamma\left(\frac{m}{2}\right) = \frac{m-2}{2}\Gamma\left(\frac{m-2}{2}\right)$$

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}$$

したがって、t の分散は $V(t) = \overline{t^2} - \bar{t}^2 = \dfrac{m}{m-2} - 0 = \dfrac{m}{m-2}$ である。

t分布($m=1$):分散

自由度 1 のt分布にしたがう t の2乗平均は

$$\overline{t^2}=\int_{-\infty}^{\infty}t^2 f_1(t)\mathrm{d}t=\frac{1}{\sqrt{\pi}}\frac{\Gamma(1)}{\Gamma(1/2)}\int_{-\infty}^{\infty}t^2(1+t^2)^{-1}\mathrm{d}t$$

$$=\frac{1}{\sqrt{\pi}}\frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty}\frac{t^2}{1+t^2}\mathrm{d}t=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{t^2}{1+t^2}\mathrm{d}t$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}\left(1-\frac{1}{1+t^2}\right)\mathrm{d}t=\frac{1}{\pi}\left[t-\tan^{-1}t\right]_{-\infty}^{\infty}$$

となり、発散してしまう。したがって、自由度 1 のt分布(コーシー分布、ローレンツ分布)の分散は存在しない。

t分布$(m=2)$:分散

自由度 2 のt分布にしたがう t の2乗平均は

$$\overline{t^2}=\int_{-\infty}^{\infty}t^2 f_2(t)\mathrm{d}t=\frac{1}{\sqrt{2\pi}}\frac{\Gamma(3/2)}{\Gamma(1)}\int_{-\infty}^{\infty}t^2\left(1+\frac{t^2}{2}\right)^{-3/2}\mathrm{d}t$$

$$=\int_{-\infty}^{\infty}\frac{t^2}{\sqrt{(2+t^2)^3}}\mathrm{d}t=\left[\frac{-t}{\sqrt{2+t^2}}\right]_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\frac{1}{\sqrt{2+t^2}}\mathrm{d}t$$

$$=-2+\left[\log\left|t+\sqrt{2+t^2}\right|\right]_{-\infty}^{\infty}$$

となり、発散してしまう。したがって、自由度 2 のt分布の分散は存在しない。

確率密度関数

自由度 $m = 1, 2, 3, 4$ に対して、t分布の確率密度関数は、次のようになる。

$$f_1(t) = \frac{1}{\sqrt{\pi}} \frac{\Gamma(1)}{\Gamma(1/2)} (1 + t^2)^{-1} = \frac{1}{\pi} \frac{1}{1 + t^2}$$

$m = 1$ は、コーシー分布またはローレンツ分布と呼ばれる。

$$f_2(t) = \frac{1}{\sqrt{2\pi}} \frac{\Gamma(3/2)}{\Gamma(1)} \left(1 + \frac{t^2}{2}\right)^{-3/2} = \frac{1}{\sqrt{(2 + t^2)^3}}$$

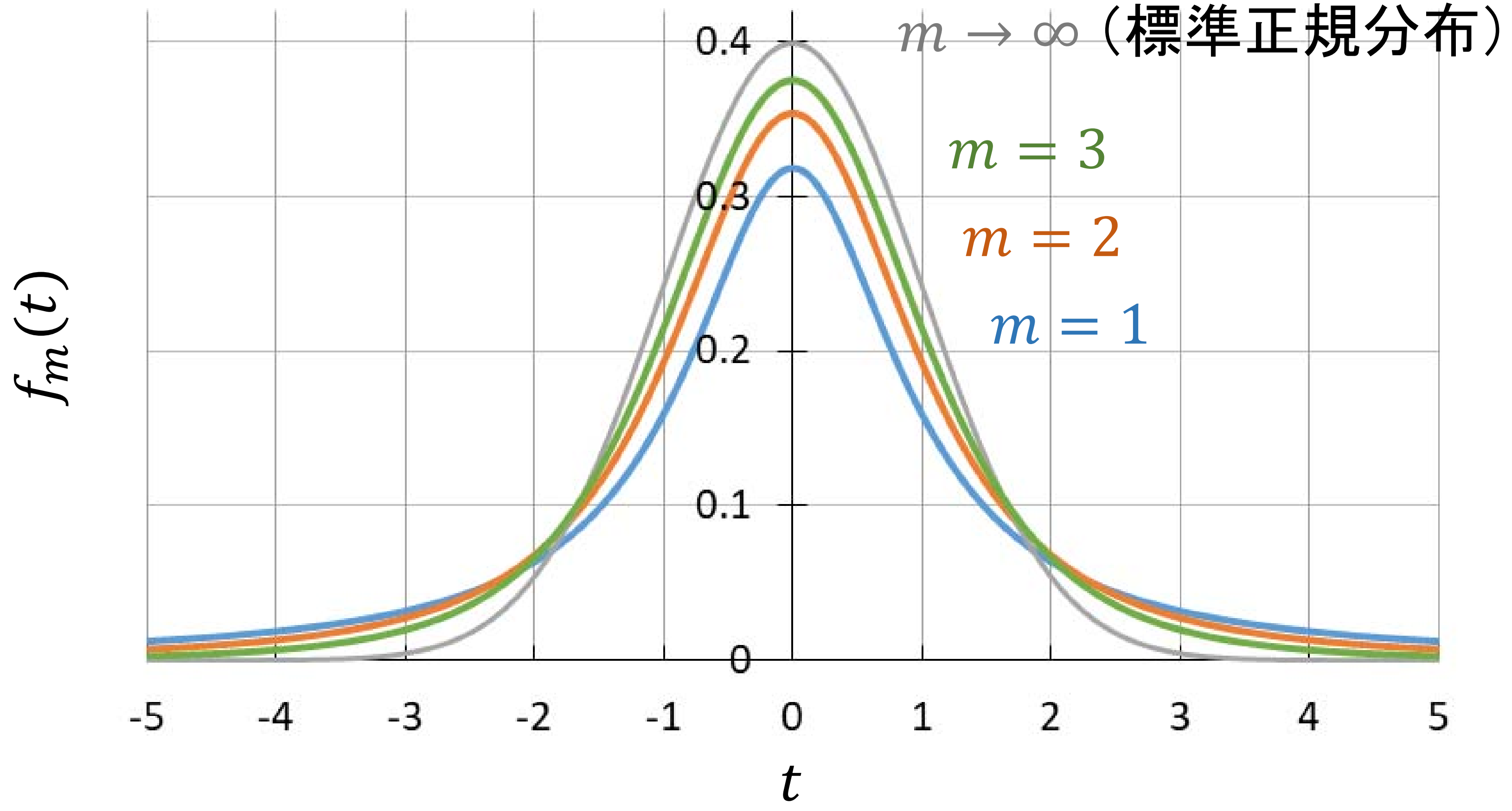
$$f_3(t) = \frac{1}{\sqrt{3\pi}} \frac{\Gamma(2)}{\Gamma(3/2)} \left(1 + \frac{t^2}{3}\right)^{-2} = \frac{6\sqrt{3}}{\pi} \frac{1}{(3 + t^2)^2}$$

$$f_4(t) = \frac{1}{2\sqrt{\pi}} \frac{\Gamma(5/2)}{\Gamma(2)} \left(1 + \frac{t^2}{4}\right)^{-5/2} = \frac{12}{\sqrt{(4 + t^2)^5}}$$

ローレンツ分布はオランダの物理学者H. A. Lorentz（1853-1928）にちなむ。Lorentzはローレンツ力とローレンツ変換にも名を残し、ゼーマン効果に関して1902年にノーベル物理学賞を受賞している。いっぽう、ローレンツ数（熱伝導率と電気伝導率の比）とローレンツゲージは、デンマークの物理学者L. V. Lorenz（1829-1891）にちなむ。名前のつづりが違うことに注意。なお、屈折率と分極率の関係式は、両者にちなんで「Lorentz-Lorenzの式」と呼ばれる。

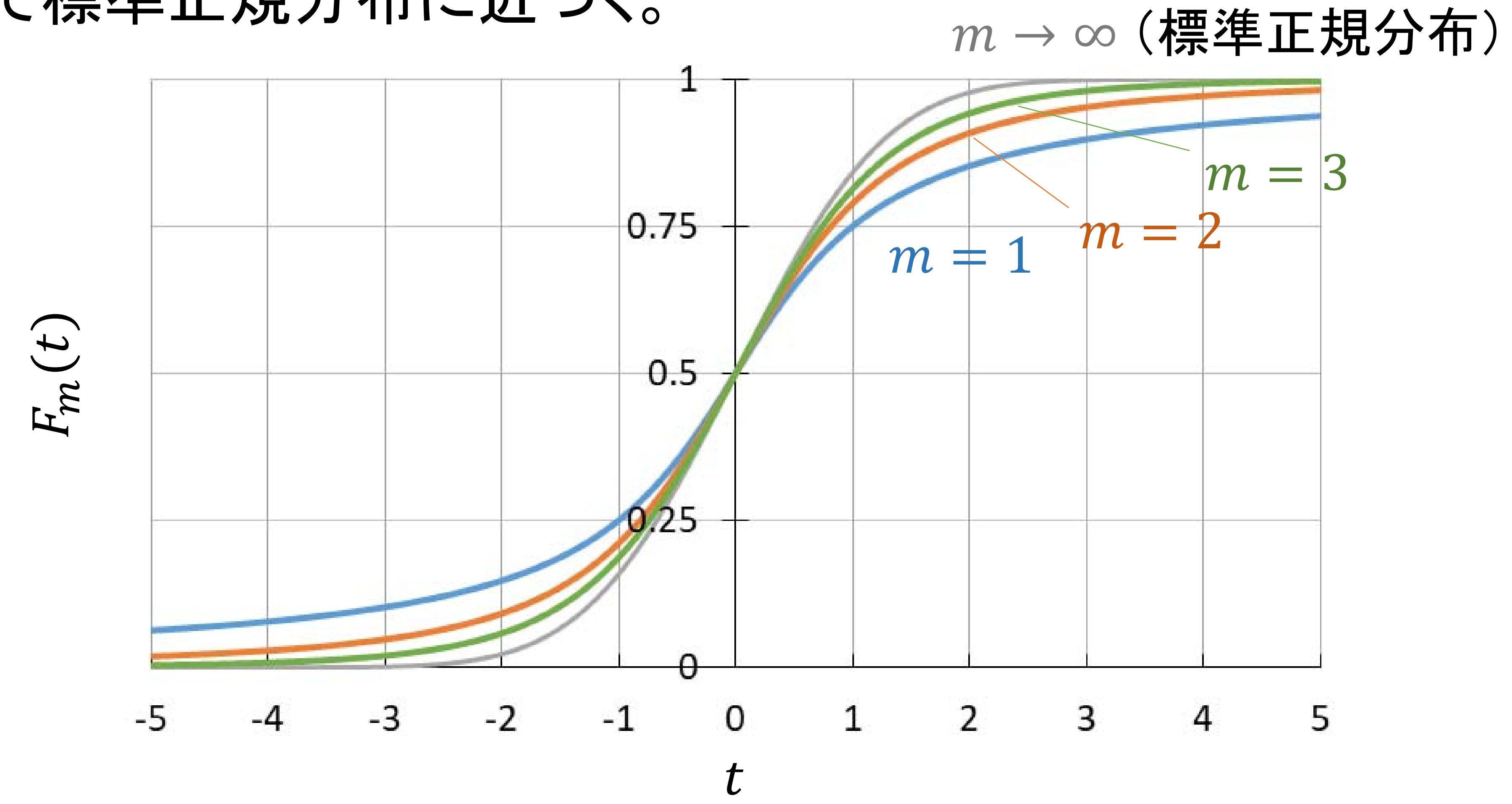
確率密度関数

自由度 $m = 1, 2, 3$ のt分布の確率密度関数のグラフを下図に示す。
$m \to \infty$ で標準正規分布に近づく。



累積分布関数

自由度 $m = 1, 2, 3$ のt分布の累積分布関数のグラフを下図に示す。
$m \to \infty$ で標準正規分布に近づく。



t分布と標準正規分布

t分布の確率密度関数

$$f_m(t) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}}$$

は自由度 $m \to \infty$ のとき、

$$\frac{1}{\sqrt{m}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \to \frac{1}{\sqrt{2}}, \qquad \left(1 + \frac{t^2}{m}\right)^{-\frac{m+1}{2}} \to \mathrm{e}^{-t^2/2}$$

となり、標準正規分布

$$f(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}$$

に近づく。

t分布と標準正規分布

まず、$m \to \infty$ で

$$\frac{1}{\sqrt{m}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \to \frac{1}{\sqrt{2}}$$

となることを示す。

ガンマ関数 $\Gamma(z)$ の定義より

$$\Gamma(z) = \int_0^\infty s^{z-1} \mathrm{e}^{-s} \mathrm{d}s = \lim_{N\to\infty} \int_0^N s^{z-1} \mathrm{e}^{-s} \mathrm{d}s$$

$$= \lim_{N\to\infty} \int_0^N s^{z-1} \left(1 - \frac{s}{N}\right)^N \mathrm{d}s$$

t分布と標準正規分布

$$\Gamma(z)=\int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s=\lim_{N\to\infty}\int_0^N s^{z-1}\left(1-\frac{s}{N}\right)^N\mathrm{d}s$$

$u=\dfrac{s}{N}$ と変数変換すると、$\mathrm{d}u=\dfrac{\mathrm{d}s}{N}$ だから

$$\int_0^n s^{z-1}\left(1-\frac{s}{N}\right)^N\mathrm{d}s=N^z\int_0^1 u^{z-1}(1-u)^N\mathrm{d}u=N^z\mathrm{B}(z,N+1)$$

$$=N^z\frac{\Gamma(z)\Gamma(N+1)}{\Gamma(z+N+1)}$$

ここで、$\mathrm{B}(z,n+1)$ はベータ関数であり、ベータ関数とガンマ関数の関係 $\mathrm{B}(p,q)\Gamma(p+q)=\Gamma(p)\Gamma(q)$ を用いた。

t分布と標準正規分布

$$\Gamma(z) = \int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s = \lim_{N\to\infty} N^z \frac{\Gamma(z)\Gamma(N+1)}{\Gamma(z+N+1)}$$

ここで、$z = \dfrac{1}{2}, N+1 = \dfrac{m}{2}$とすると

$$1 = \lim_{N\to\infty} \sqrt{N}\frac{\Gamma(N+1)}{\Gamma\left(\frac{1}{2}+N+1\right)} \approx \lim_{N\to\infty} \sqrt{N+1}\frac{\Gamma(N+1)}{\Gamma\left(\frac{1}{2}+N+1\right)}$$

$$= \lim_{m\to\infty} \sqrt{\frac{m}{2}}\frac{\Gamma(m/2)}{\Gamma((m+1)/2)}$$

ゆえに、$m \to \infty$ のとき $\dfrac{1}{\sqrt{m}}\dfrac{\Gamma((m+1)/2)}{\Gamma(m/2)} \to \dfrac{1}{\sqrt{2}}$ である。

t分布と標準正規分布

また、$m \to \infty$ で

$$\left(1+\frac{t^2}{m}\right)^{-\frac{m+1}{2}} \approx \left(1+\frac{t^2}{m}\right)^{-\frac{m}{2}} = \left(1+\frac{t^2}{m}\right)^{-\frac{m}{t^2}\cdot\frac{t^2}{2}} \to \mathrm{e}^{-t^2/2}$$

だから、t分布の確率密度関数

$$f_m(t) = \frac{1}{\sqrt{m\pi}} \frac{\Gamma((m+1)/2)}{\Gamma(m/2)} \left(1+\frac{t^2}{m}\right)^{-(m+1)/2}$$

は自由度 $m \to \infty$ のとき、標準正規分布

$$f(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2}$$

に近づく。

正規分布の標準化

母平均 μ、母分散 σ^2 の正規分布 $f(x_i) = \dfrac{1}{\sqrt{2\pi\sigma}} \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ にしたがう

母集団からランダムに抽出したサイズ n の標本 $x_i\,(i = 1, 2, \cdots, n)$ に

対して、$X_i = \dfrac{x_i - \mu}{\sigma}$ は標準正規分布 $f(x_i) = \dfrac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{X_i{}^2}{2}}$ にしたがう

から、その n 個の和 $\displaystyle\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma}$ は平均 0、分散 n の正規分布

にしたがう。よって、$\displaystyle\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma}$ は平均 0、分散 1 の標準

正規分布にしたがう。

正規分布の標準化

母平均 μ、母分散 σ^2 の正規分布にしたがう母集団からランダムに抽出したサイズ n の標本 x_i に対して、標本平均を $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ とすると、

$$\frac{\bar{x}-\mu}{\sigma} = \frac{1}{\sigma}\left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) - \mu\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

である。ここで、$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ は標準正規分布にしたがうから、

$X = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ は標準正規分布にしたがう。

正規分布の標準化

母平均 μ、母分散 σ^2 の正規分布 $f(x_i) = \dfrac{1}{\sqrt{2\pi\sigma}}\mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ にしたがう

母集団からランダムに抽出したサイズ n の標本 $x_i\,(i = 1, 2, \cdots, n)$ に

対して、$X_i = \dfrac{x_i - \mu}{\sigma}$ は標準正規分布 $f(x_i) = \dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{{X_i}^2}{2}}$ にしたがう

から、その2乗和 $y = S_n = \displaystyle\sum_{i=1}^{n} {X_i}^2 = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2$ は、自由度 n の

カイ2乗分布 $f_n(y) = \chi_n^2(y) = \dfrac{y^{(n/2)-1}\mathrm{e}^{-y/2}}{2^{n/2}\Gamma(n/2)}$ にしたがう。

偏差の2乗和と母分散

$X_i = \dfrac{x_i - \mu}{\sigma}$ の母平均 μ を、標本平均 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ に置き換えて、

標本平均 $\bar{x}$ からの偏差の2乗和と母分散 σ^2 との比を

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

で表す。ここで、

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

は、標本の不偏分散である。

偏差の2乗和

標本平均 $\bar{x}$ からの偏差の2乗和について、次式が成り立つ。

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= \sum_{i=1}^{n}\{(x_i-\mu)-(\bar{x}-\mu)\}^2 \\ &= \sum_{i=1}^{n}(x_i-\mu)^2-2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)+\sum_{i=1}^{n}(\bar{x}-\mu)^2 \\ &= \sum_{i=1}^{n}(x_i-\mu)^2-2n(\bar{x}-\mu)^2+n(\bar{x}-\mu)^2 \\ &= \sum_{i=1}^{n}(x_i-\mu)^2-n(\bar{x}-\mu)^2\end{aligned}$$

偏差の2乗和

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

の右辺第2項において

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

だから

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

偏差の2乗和と母分散の比

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

より

$$Y=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}{X_i}^2-\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right)^2$$

X_i は標準正規分布にしたがうから、n 個の X_i の和は平均 0、分散 n の正規分布にしたがう。それを $\sqrt{n}$ で割ったものは標準正規分布にしたがう。

この第1項は自由度 n のカイ2乗分布に、第2項は自由度 1 のカイ2乗分布にしたがうから、Y は自由度 $n-1$ のカイ2乗分布にしたがう。

ガンマ関数

ガンマ関数 $\Gamma(z)$ は

$$\Gamma(z) = \int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s$$

で定義される。

$$\Gamma(z) = \int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s = \left[-s^{z-1}\mathrm{e}^{-s}\right]_0^\infty + (z-1)\int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s$$

$$= (z-1)\Gamma(z-1)$$

$$\Gamma(1) = \int_0^\infty \mathrm{e}^{-s}\mathrm{d}s = \left[-\mathrm{e}^{-s}\right]_0^\infty = 1$$

ガンマ関数は、階乗を非整数に拡張したもの

より、正の整数 n に対して、$\Gamma(n) = (n-1)!$ である。

ガンマ関数

$$\Gamma\left(\frac{1}{2}\right)=\int_0^\infty s^{-1/2}\mathrm{e}^{-s}\mathrm{d}s=\int_0^\infty \frac{\mathrm{e}^{-s}}{\sqrt{s}}\mathrm{d}s$$

$s=u^2$ とおくと

$$\Gamma\left(\frac{1}{2}\right)=\int_0^\infty \frac{\mathrm{e}^{-s}}{\sqrt{s}}\mathrm{d}s=\int_0^\infty \frac{\mathrm{e}^{-u^2}}{u}2u\mathrm{d}u=2\int_0^\infty \mathrm{e}^{-u^2}\mathrm{d}u$$

$$=\int_{-\infty}^\infty \mathrm{e}^{-u^2}\mathrm{d}u=\sqrt{\pi}$$

これより、正の整数 m に対して

$$\Gamma\left(m+\frac{1}{2}\right)=\frac{2m-1}{2}\Gamma\left(m-\frac{1}{2}\right)=\frac{(2m-1)!!}{2^m}\sqrt{\pi}$$

二重階乗

$(2m-1)!!=(2m-1)\cdot(2m-3)\cdot\cdots\cdot3\cdot1$

ベータ関数

ベータ関数 $\mathrm{B}(p,q)$ は

$$\mathrm{B}(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t$$

で定義される。

$u = 1 - t$ とおくと

$$\mathrm{B}(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t = \int_1^0 (1-u)^{p-1}u^{q-1}(-\mathrm{d}u)$$

$$= \int_0^1 u^{q-1}(1-u)^{p-1}\mathrm{d}u = \mathrm{B}(q,p)$$

だから、$\mathrm{B}(p,q) = \mathrm{B}(q,p)$ である。

ベータ関数とガンマ関数

$$\mathrm{B}(p,q)=\int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t\,,\quad \Gamma(z)=\int_0^\infty s^{z-1}\mathrm{e}^{-s}\mathrm{d}s$$

より

$$\mathrm{B}(p,q)\Gamma(p+q)=\int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t\int_0^\infty s^{p+q-1}\mathrm{e}^{-s}\mathrm{d}s$$

ここで、$u=st, v=s(1-t)$ と変数変換すると

$$\begin{vmatrix}\partial u/\partial s & \partial u/\partial t\\ \partial v/\partial s & \partial v/\partial t\end{vmatrix}=\begin{vmatrix}t & s\\ 1-t & -s\end{vmatrix}=-st-s(1-t)=-s$$

より

$$\mathrm{d}u\mathrm{d}v=|-s|\mathrm{d}s\mathrm{d}t=s\mathrm{d}s\mathrm{d}t$$

ベータ関数とガンマ関数

$$\mathrm{B}(p,q)\Gamma(p+q) = \int_0^1 t^{p-1}(1-t)^{q-1}\mathrm{d}t \int_0^\infty s^{p+q-1}\mathrm{e}^{-s}\mathrm{d}s$$

$u = st, v = s(1-t)$ と変数変換すると、$\mathrm{d}u\mathrm{d}v = s\mathrm{d}s\mathrm{d}t$ より

$$\mathrm{B}(p,q)\Gamma(p+q) = \int_0^\infty \int_0^\infty t^{p-1}(1-t)^{q-1}s^{p+q-2}\mathrm{e}^{-s}\mathrm{d}u\,\mathrm{d}v$$

$$= \int_0^\infty \int_0^\infty u^{p-1}v^{q-1}\mathrm{e}^{-(u+v)}\mathrm{d}u\,\mathrm{d}v$$

$$= \int_0^\infty u^{p-1}\mathrm{e}^{-u}\mathrm{d}u \int_0^\infty v^{q-1}\mathrm{e}^{-v}\mathrm{d}v = \Gamma(p)\Gamma(q)$$

$\therefore\ \mathrm{B}(p,q)\Gamma(p+q) = \Gamma(p)\Gamma(q)$

参考文献

- 小寺平治「新統計入門」裳華房、1996.
- 水本久夫「統計の基礎」培風館、1994.
- 薩摩順吉「確率・統計」（理工系の数学入門コース 7）岩波書店、1989.
- 柴田文明「確率・統計」（理工系の基礎数学 7）岩波書店、1996.
- 大鑄史男「工科のための確率・統計」（工科のための数理）数理工学社、2005.
- 岩佐学、薩摩順吉、林利治「確率・統計」（理工系の数理）裳華房、2018.
- 東京大学教養部統計学教室（編）「統計学入門」（基礎統計学Ｉ）東京大学出版会、1991.
- 竹内淳「高校数学でわかる統計学」（ブルーバックス）講談社、2012.