F 分布 (F-distribution)

x_1，x_2 が互いに独立で、それぞれ自由度 n_1，n_2 の χ^2 分布にしたがっているものとする。それらの確率密度関数は、

$$T_{n_i}(x_i) = \frac{1}{2^{n_i/2}\Gamma(n_i/2)} x_i^{n_i/2-1} e^{-x_i/2}$$

で与えられる。ここで、i は、1 または 2 の値をとるものとする。このとき、x_1，x_2 を自由度で割った比

$$x = \frac{x_1/n_1}{x_2/n_2}$$

の分布について調べる。この分布を $f_{n_1,\,n_2}(x)$ と書くことにすると、

$$f_{n_1,\,n_2}(x) = \int_0^\infty \delta\left(x - \frac{x_1/n_1}{x_2/n_2}\right) T_{n_1}(x_1) T_{n_2}(x_2)\, dx_1 dx_2$$

となる。ここで、x_1 についての積分をおこなうが、この前に、$y = n_2 x_1/n_1 x_2$ とおいて変数変換をおこなう。

$$f_{n_1,\,n_2}(x) = \int_0^\infty \frac{n_1}{n_2} x_2 \delta\left(x - y\right) T_{n_1}(n_1 x_2 y/n_2) T_{n_2}(x_2)\, dy dx_2$$

今度は、y についての積分は、単に y を x で置き換えるだけになって、

$$f_{n_1,\,n_2}(x) = \int_0^\infty \frac{n_1}{n_2} x_2 T_{n_1}(n_1 x_2 x/n_2) T_{n_2}(x_2)\, dx_2$$

となる。ここで、カイ 2 乗分布の確率密度関数の具体的な形を代入すると、

$$f_{n_1,\,n_2}(x) = \frac{x^{(n_1-2)/2}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} \int_0^\infty x_2^{(n_1+n_2)/2-1} e^{-(1+n_1 x/n_2)x_2/2}\, dx_2$$

となる。さらに、積分変数として、$t = (1 + n_1 x/n_2)x_2/2$ を用いることにすると、

$$f_{n_1,\,n_2}(x) = \frac{x^{(n_1-2)/2}}{\Gamma(n_1/2)\Gamma(n_2/2)(1 + n_1 x/n_2)^{(n_1+n_2)/2}} \left(\frac{n_1}{n_2}\right)^{n_1/2} \int_0^\infty t^{(n_1+n_2)/2-1} e^{-t}\, dt$$

最後の積分は、ガンマ関数をもちいて、

$$\Gamma\left((n_1+n_2)/2\right) = \int_0^\infty t^{(n_1+n_2)/2-1} e^{-t}\, dt$$

と書けるので、結局。

$$f_{n_1,\,n_2}(x) = \frac{\Gamma\left((n_1+n_2)/2\right) x^{(n_1-2)/2}}{\Gamma(n_1/2)\Gamma(n_2/2)(1 + n_1 x/n_2)^{(n_1+n_2)/2}} \left(\frac{n_1}{n_2}\right)^{n_1/2}$$

を得る。これを、自由度 $(n_1,\, n_2)$ の F 分布という。これは、また、ベータ関数とガンマ関数の関係式

$$B(\alpha,\, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

を用いて、

$$f_{n_1,\,n_2}(x) = \frac{1}{B(n_1/2,\, n_2/2)} \left(\frac{n_1}{n_2}\right)^{n_1/2} \frac{x^{(n_1-2)/2}}{(1 + n_1 x/n_2)^{(n_1+n_2)/2}}$$

と表すこともできる。

F 分布曲線を図１に示す。

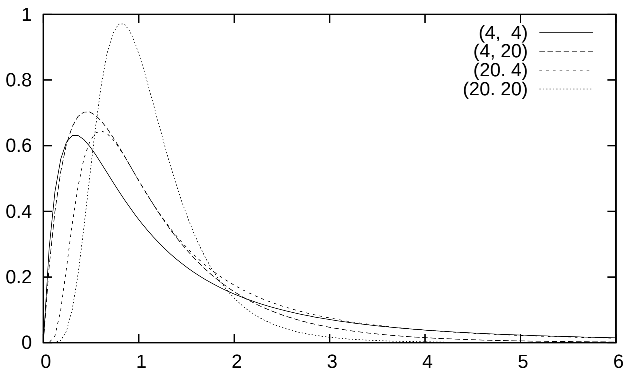

図 1: F 分布曲線。 自由度 (n_1, n_2) が (4, 4), (4, 20), (20, 4), (20, 20) の場合を示す。

F 分布の適用

２つの正規母集団 $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$ があって、それぞれから大きさが、n_1, n_2 の標本 $(x_{1,1}, x_{1,2}, x_{1,3}, \cdots, x_{1,n_1})$，$(x_{2,1}, x_{2,2}, x_{2,3}, \cdots, x_{2n_2})$ を選び出したとする。このとき、各々の標本平均 $\bar{x}_1$, $\bar{x}_2$ を

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{1,i}, \qquad \bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_1} x_{2,i},$$

標本不偏分散 s_1^2, s_2^2 を、

$$\begin{aligned} s_1^2 &= \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(x_{1,i} - \bar{x}_1)^2 \\ s_2^2 &= \frac{1}{n_2 - 1}\sum_{i=1}^{n_1}(x_{2,i} - \bar{x}_2)^2 \end{aligned}$$

とする。このとき、

$$\begin{aligned} \chi_1^2 &= \frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(x_{1,i} - \bar{x}_1)^2 = \frac{(n_1 - 1)s_1^2}{\sigma_1^2} \\ \chi_2^2 &= \frac{1}{\sigma_2^2}\sum_{i=1}^{n_2}(x_{2,i} - \bar{x}_2)^2 = \frac{(n_2 - 1)s_2^2}{\sigma_2^2} \end{aligned}$$

は、それぞれ、自由度 $(n_1 - 1)$, $(n_2 - 1)$ の χ^2 分布に従う。ゆえに、

$$\frac{\chi_1^2/(n_1 - 1)}{\chi_2^2/(n_2 - 1)} = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

は、自由度 $(n_1 - 1)$, $(n_2 - 1)$ の F 分布に従う。そこで、F 分布は、2 つの母集団の分散が等しいかどうかのの推定や検定に使われる。

$N(\mu, \sigma^2)$ に従う母集団から n コの標本を抽出する。$\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$ から標本平均を作る。$\bar{x}$ の分散 $(\bar{x} - \mu)^2$ の期待値は、σ^2/n である。そこで、確率変数 $\bar{x}$ を標準化してつくっ

た分散 $n(\bar{x}-\mu)^2/\sigma^2$ は、自由度 1 の χ^2 分布に従う。また、標本分散を標準化した $(n-1)s^2/\sigma^2$ は、自由度 $(n-1)$ の χ^2 分布に従う。両者を各々の自由度で割ったものの比、

$$X = \frac{n(\bar{x}-\mu)^2/\sigma^2}{s^2/\sigma^2} = \frac{n(\bar{x}-\mu)^2}{s^2}$$

は、自由度 $(1,\, n-1)$ の　F 分布に従う。

F 分布の性質

自由度 $(n_1,\, n_2)$ の F 分布曲線 $f_{n_1,\,n_2}(x)$ の変数 x を $z = \dfrac{1}{x}$ に置き換える。ただし、ただ単に、x を $\dfrac{1}{z}$ で置き換えるのではなく、デルタ関数を用いて以下のように行う。

$$\int f_{n_1,\,n_2}(x)\delta\left(z - \frac{1}{x}\right)\,dx \tag{1}$$

ここで、$a(x) = z - \dfrac{1}{x}$ を $x = \dfrac{1}{z}$ のまわりで、テイラー展開すると、

$$a(x) = z^2\left(x - \frac{1}{z}\right) + \cdots$$

だから、

$$\delta\left(a(x)\right) = \delta\left(z^2\left(x - \frac{1}{z}\right)\right) = z^{-2}\delta\left(x - \frac{1}{z}\right)$$

これを (1) 式に代入して計算すると、

$$\begin{aligned}
\int f_{n_1,\,n_2}(x)\delta\left(z - \frac{1}{x}\right)\,dx &= z^{-2}\int f_{n_1,\,n_2}(x)\delta\left(x - \frac{1}{z}\right)\,dx \\
&= z^{-2} f_{n_1,\,n_2}\left(\frac{1}{z}\right) \\
&= f_{n_2,\,n_1}(z)
\end{aligned}$$

を得る。