χ^2 分布

確率変数 u が標準正規分布 $N(0,1)$ に従うとき、$x = u^2$ の従う分布を $T_1(x)$ と書くことにしよう。これは、標準正規分布の確率密度関数 $f(u) = \dfrac{1}{\sqrt{2\pi}} e^{-u^2/2}$ を用いて表すと、

$$T_1(x) = \int \delta(u^2 - x) f(u)\, du \tag{1}$$

となる。ところでデルタ関数の基本的な関係式によれば、

$$\delta(u^2 - x) = \frac{1}{2\sqrt{x}} \big(\delta(u - \sqrt{x}) + \delta(u + \sqrt{x})\big)$$

となる。これは、x の値が与えられた場合に、対応する u の値は、$u = \pm\sqrt{x}$ の 2 通りがあることを意味している。上記の関係式を (1) 式に代入すると、

$$\begin{aligned} T_1(x) &= \frac{1}{2\sqrt{x}} \Big(f(\sqrt{x}) + f(-\sqrt{x})\Big) \\ &= \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2} \qquad (x > 0) \end{aligned} \tag{2}$$

となる。

次に、$x = u_1^2 + u_2^2 = x_1 + x_2$ の従う分布 $T_2(x)$ を求める。

$$\begin{aligned} T_2(x) &= \int \delta(x_1 + x_2 - x) T_1(x_1) T_1(x_2)\ dx_1 dx_2 \\ &= \int_0^x T_1(x - x_2) T_1(x_2)\ dx_2 \\ &= \frac{1}{2\pi} e^{-x/2} \int_0^x (x - x_2)^{-1/2} x_2^{-1/2}\ dx_2 \end{aligned}$$

ここで、$t = x_2/x$ とおくと、

$$T_2(x) = \frac{1}{2\pi} e^{-x/2} \int_0^1 (1-t)^{-1/2} t^{-1/2}\, dt$$

となるが、最後の積分は、$t = \sin^2\theta$ とおくと、

$$\int_0^1 (1-t)^{-1/2} t^{-1/2}\, dt = 2 \int_0^{\pi/2} d\theta = \pi$$

となるので、結局、

$$T_2(x) = \frac{1}{2} e^{-x/2} \qquad (x > 0) \tag{3}$$

を得る。

一般に、n コの変数 $u_1, u_2, \cdots, u_n$ が互いに独立で、それぞれ $N(0,1)$ に従うとき、$x = u_1^2 + u_2^2 + \cdots + u_n^2$ の分布 $T_n(x)$ は、

$$T_n(x) = \frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \qquad (x > 0) \tag{4}$$

となる。この分布を自由度 n のカイ 2 乗 (χ^2) 分布という。

証明

証明は数学的帰納法による。すなわち、$x_{n-1} = u_1^2 + u_2^2 + \cdots + u_{n-1}^2$ の分布が $T_{n-1}(x)$ と表されたとすると、$T_n(x)$ は、

$$
\begin{aligned}
T_n(x) &= \int \delta(x_{n-1} + x_1 - x) T_{n-1}(x_{n-1}) T_1(x_1) \ dx_{n-1} dx_1 \\
&= \int_0^x T_{n-1}(x - x_1) T_1(x_1) \ dx_1 \\
&= \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma((n-1)/2) \Gamma(1/2)} \int_0^1 (1-t)^{\frac{n-3}{2}} t^{-\frac{1}{2}} \ dt \\
&= \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma((n-1)/2) \Gamma(1/2)} \frac{\Gamma((n-1)/2)\Gamma(1/2)}{\Gamma(n/2)} \\
&= \frac{1}{2\Gamma(n/2)} \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}
\end{aligned}
$$

となる。これは、(4) に一致する。

χ^2 分布の平均と分散

$$
\begin{aligned}
\text{平均} &= \int_0^\infty x T_n(x) \ dx = \frac{1}{2\Gamma(n/2)} \int_0^\infty x \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \ dx \\
&= \frac{2}{\Gamma(n/2)} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{n}{2}} e^{-\frac{x}{2}} \ d\left(\frac{x}{2}\right) = \frac{2}{\Gamma(n/2)} \int_0^\infty t^{(\frac{n}{2}+1)-1} e^{-t} \ dt \\
&= \frac{2\Gamma(n/2+1)}{\Gamma(n/2)} = \frac{2(n/2)\Gamma(n/2)}{\Gamma(n/2)} = n
\end{aligned}
$$

$$
\begin{aligned}
\text{分散} &= \int_0^\infty x^2 T_n(x) \ dx - n^2 = \frac{1}{2\Gamma(n/2)} \int_0^\infty x^2 \left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}} \ dx - n^2 \\
&= \frac{4}{\Gamma(n/2)} \int_0^\infty \left(\frac{x}{2}\right)^{\frac{n}{2}+1} e^{-\frac{x}{2}} \ d\left(\frac{x}{2}\right) - n^2 = \frac{4}{\Gamma(n/2)} \int_0^\infty t^{(\frac{n}{2}+2)-1} e^{-t} \ dt - n^2 \\
&= \frac{4\Gamma(n/2+2)}{\Gamma(n/2)} - n^2 = 4\left(\frac{n}{2}+1\right)\frac{n}{2} - n^2 = 2n
\end{aligned}
$$

標本分散の標本分布

確率変数 u_1, u_2 が互いに独立で、それぞれ標準正規分布 $N(0,1)$ に従うときに、標本平均

$$\bar{u} = \frac{1}{2}(u_1 + u_2) \tag{5}$$

をもちいて、

$$x = (u_1 - \bar{u})^2 + (u_2 - \bar{u})^2 \tag{6}$$

を作る。ここで、x の従う分布を求めたい。ところが、$(u_1 - \bar{u})^2$ と $(u_2 - \bar{u})^2$ は、条件 (5) があるために、互いに独立ではない。したがって、x の従う分布は自由度 2 の χ^2 分布とはならない。(5) を (6) に代入して整理すると、

$$x = \frac{1}{2}(u_1 - u_2)^2$$

となる。ここで、$t=(u_1-u_2)/\sqrt{2}$ と書くことにすると、変数 t は、正規分布の一次結合の平均と分散の関係式から、

$$
\begin{aligned}
\text{平均}\quad \mu_t &= \frac{1}{\sqrt{2}}\times 0-\frac{1}{\sqrt{2}}\times 0=0\\
\text{分散}\quad \sigma_t^2 &= \left(\frac{1}{\sqrt{2}}\right)\times 1+\left(-\frac{1}{\sqrt{2}}\right)\times 1=1
\end{aligned}
$$

となり、平均 0、分散 1 の正規分布に従うことがわかる。そこで、t を 2 乗した $x=\frac{1}{2}(u_1-u_2)^2$ は、自由度 1 の χ^2 分布に従うことがわかる。

もう少し別の見方をしてみよう。新しい変数

$$v_1=u_1-\bar{u},\quad v_2=u_2-\bar{u}$$

を導入すると、

$$x=v_1^2+v_2^2$$

となる。ところが、

$$v_1+v_2=(1,\ 1)\begin{pmatrix}v_1\\ v_2\end{pmatrix}=0$$

の関係がある。これは、ベクトル $(v_1,\ v_2)$ の $(1,\ 1)$ 方向成分が 0 であることを表している。そこで、v_1、v_2 を直交変換した、

$$\begin{pmatrix}v_1'\\ v_2'\end{pmatrix}=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}\begin{pmatrix}v_1\\ v_2\end{pmatrix}$$

を用いると、

$$x=v_1'^2+v_2'^2$$

となる。ここで、$v_2'=(v_1+v_2)/\sqrt{2}=0$ なので、

$$x=v_1'^2=\left(\frac{v_1-v_2}{\sqrt{2}}\right)^2=\frac{(u_1-u_2)^2}{2}$$

を得る。変数 v_1' は標準正規分布に従うので、変数 x は自由度 1 の χ^2 分布に従う。

この方法は、もっと一般の場合にも使えそうである。今度は、三つの互いに独立な変数 u_1,u_2,u_3 の標本平均

$$\bar{u}=\frac{1}{3}(u_1+u_2+u_3) \tag{7}$$

をもちいて表される分散

$$x=(u_1-\bar{u})^2+(u_2-\bar{u})^2+(u_3-\bar{u})^2 \tag{8}$$

の従う分布関数を求めてみよう。

ここで、

$$v_1=u_1-\bar{u},\quad v_2=u_2-\bar{u},\quad v_3=u_3-\bar{u}$$

とおくと、(8) 式は、

$$x=v_1^2+v_2^2+v_3^2$$

となり、 条件 (7) は、

$$v_1+v_2+v_3=0$$

と書き直すことができる。そこで、(v_1, v_2, v_3) 空間で、基底ベクトル $(0,0,1)$ が $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$ となるような直交変換、例えば、

$$\begin{pmatrix} v_1' \\ v_2' \\ v_3' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

を考える。すると、

$$x = v_1'^2 + v_2'^2 + v_3'^2$$

となるが、$v_3' = (v_1 + v_2 + v_3)/\sqrt{3} = 0$ なので、

$$x = v_1'^2 + v_2'^2 \quad = \frac{(u_1 - u_2)^2}{2} + \frac{(u_1 + u_2 - 2u_3)^2}{6}$$

である。ゆえに、x は自由度 2 の χ^2 分布にしたがう。

同様にして、一般の n の場合にも証明できる。標本平均

$$\bar{u} = \frac{1}{n}(u_1 + u_2 + \cdots + u_n)$$

から

$$v_1 = u_1 - \bar{u},\ v_2 = u_2 - \bar{u},\ \cdots\ ,\ v_k = u_k - \bar{u},\ \cdots,\ v_n = u_n - \bar{u}$$

をつくると、これらの間には、

$$v_1 + v_2 + v_3 + \cdots + v_n = 0$$

なる関係がある。ここで、

$$x = v_1^2 + v_2^2 + \cdots + v_n^2$$

とおく。n 次元 $(v_1, v_2, \cdots, v_n)$ 空間で、基底ベクトル $(0, 0, \cdots, 1)$ が、$(1/\sqrt{n}, 1/\sqrt{n}, \cdots, 1/\sqrt{n})$ となるような直交変換

$$\begin{pmatrix} v_1' \\ v_2' \\ \vdots \\ v_{n-1}' \\ v_n' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ v_n \end{pmatrix}$$

をおこなうと、$v_n' = 0$ なので、

$$\begin{aligned} x &= v_1'^2 + v_2'^2 + \cdots + v_{n-1}'^2 \\ &= \frac{(u_1 - u_2)^2}{2} + \frac{(u_1 + u_2 - 2u_3)^2}{6} + \cdots + \frac{(u_1 + u_2 + \cdots + u_{n-1} - (n-1)u_n)^2}{n(n-1)} \end{aligned}$$

を得る。

以上のことから、$N(0,1)$ に従う正規母集団から大きさ n の標本 $u_1,\ u_2,\ \cdots\ , u_n$ を無作為抽出し、標本平均 $\bar{u} = \dfrac{1}{n}(u_1 + u_2 + \cdots + u_n)$ から、

$$x = (u_1 - \bar{u})^2 + (u_2 - \bar{u})^2 + \cdots + (u_n - \bar{u})^2$$

を作ると、 x は自由度 $n-1$ の χ^2 分布に従うことが分かる。

$N(\mu,\sigma^2)$ に従う正規母集団から大きさ n の標本 $x_1,\ x_2,\ \cdots,x_n$ を無作為抽出し、標本平均 $\bar{x}=\dfrac{1}{n}(x_1+x_2+\cdots+x_n)$ を作る。標本 x_i に標準化変換 $u_i=(x_i-\mu)/\sigma$ をほどこすと、

$$\begin{aligned}\bar{u} &= \frac{1}{n}(u_1+u_2+\cdots+u_n)\\ &= \frac{1}{n\sigma}(x_1+x_2+\cdots+x_n-n\mu)\\ &= \frac{1}{\sigma}(\bar{x}-\mu)\end{aligned}$$

だから

$$\begin{aligned}x &= (u_1-\bar{u})^2+(u_2-\bar{u})^2+\cdots+(u_n-\bar{u})^2\\ &= \frac{1}{\sigma^2}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\right]\\ &= \frac{(n-1)s^2}{\sigma^2}\end{aligned}$$

は、自由度 $n-1$ の χ^2 分布に従う。

x の期待値 $E(x)$ と分散 $V(x)$ は、それぞれ、

$$\begin{aligned}E(x) &= n-1\\ V(x) &= 2(n-1)\end{aligned}$$

である。

図 (1) に χ^2 分布の様子を示す。

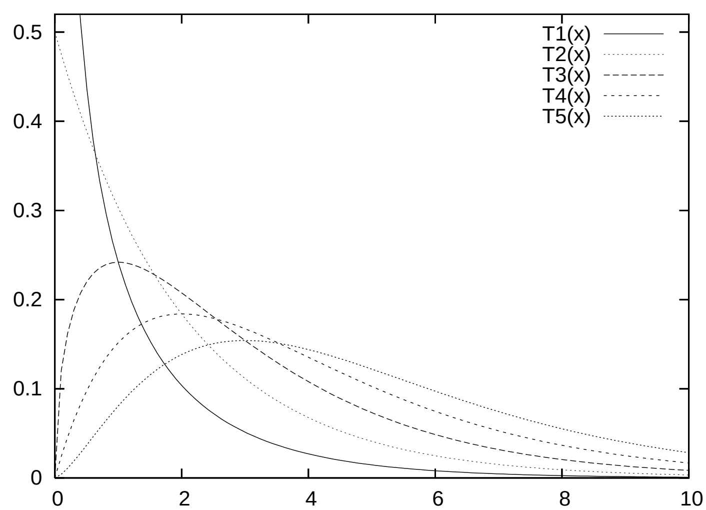

図 1: χ^2 分布