

カイ2乗分布 (χ^2 分布)

渡邊 俊夫

要点

確率変数 $X_i (i = 1, 2, \dots, n)$ が標準正規分布

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

にしたがうとき、その n 個の2乗和

$$y = S_n = \sum_{i=1}^n X_i^2$$

は、自由度 n のカイ2乗分布 (χ^2 分布)

$$f_n(y) = \chi_n^2(y) = \frac{y^{(n/2)-1} e^{-y/2}}{2^{n/2} \Gamma(n/2)} = \frac{1}{2\Gamma\left(\frac{n}{2}\right)} \left(\frac{y}{2}\right)^{\frac{n}{2}-1} e^{-\frac{y}{2}}$$

にしたがう。ここで、 $\Gamma(z)$ はガンマ関数である。

カイ2乗分布: 確率密度関数

自由度 n のカイ2乗分布の確率密度関数

$$f_n(y) = \chi_n^2(y) = \frac{y^{(n/2)-1} e^{-y/2}}{2^{n/2} \Gamma(n/2)}$$

の y についての関数部分は、ガンマ関数

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

の被積分関数と同じである。一般に、確率密度関数が

$$f_k(y) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}$$

で表される確率分布をガンマ分布という。自由度 n のカイ2乗分布は $k = n/2, \lambda = 1/2$ のガンマ分布である。

カイ2乗分布: 確率密度関数

自由度 n のカイ2乗分布の確率密度関数

$$f_n(y) = \chi_n^2(y) = \frac{y^{(n/2)-1} e^{-y/2}}{2^{n/2} \Gamma(n/2)}$$

を全範囲 ($0 \leq y < \infty$) で積分すると1になる。

$$\begin{aligned} \int_0^{\infty} f_n(y) dy &= \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\infty} y^{(n/2)-1} e^{-y/2} dy \\ &= \frac{1}{\Gamma(n/2)} \int_0^{\infty} \left(\frac{y}{2}\right)^{(n/2)-1} e^{-y/2} \frac{dy}{2} \\ &= \frac{1}{\Gamma(n/2)} \int_0^{\infty} s^{(n/2)-1} e^{-s} ds = \frac{1}{\Gamma(n/2)} \cdot \Gamma(n/2) = 1 \end{aligned}$$

カイ2乗分布：平均

自由度 n のカイ2乗分布にしたがう y の平均は

$$\begin{aligned}\bar{y} &= \int_0^{\infty} y f_n(y) dy = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\infty} y^{n/2} e^{-y/2} dy \\ &= \frac{2}{\Gamma(n/2)} \int_0^{\infty} \left(\frac{y}{2}\right)^{n/2} e^{-y/2} \frac{dy}{2} \\ &= \frac{2}{\Gamma(n/2)} \Gamma\left(\frac{n}{2} + 1\right) = \frac{2}{\Gamma(n/2)} \cdot \frac{n}{2} \Gamma(n/2) \\ &= n\end{aligned}$$

である。

確率変数 X_i は標準正規分布にしたがい、平均 0、分散 1 で、2乗平均は 1 であるから、

2乗和 $y = \sum_{i=1}^n X_i^2$ の平均は n になる。

カイ2乗分布: 分散

自由度 n のカイ2乗分布にしたがう y の2乗平均は

$$\begin{aligned}\overline{y^2} &= \int_0^{\infty} y^2 f_n(y) dy = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\infty} y^{(n/2)+1} e^{-y/2} dy \\ &= \frac{2^2}{\Gamma(n/2)} \int_0^{\infty} \left(\frac{y}{2}\right)^{(n/2)+1} e^{-y/2} \frac{dy}{2} \\ &= \frac{2^2}{\Gamma(n/2)} \Gamma\left(\frac{n}{2} + 2\right) = \frac{2^2}{\Gamma(n/2)} \cdot \left(\frac{n}{2} + 1\right) \frac{n}{2} \Gamma(n/2) \\ &= (n+2)n\end{aligned}$$

だから、分散は

$$V(y) = \overline{y^2} - \bar{y}^2 = (n+2)n - n^2 = 2n$$

カイ2乗分布: 最頻値

自由度 n のカイ2乗分布にしたがう y の最頻値(モード)は

$$\begin{aligned}\frac{df_n(y)}{dy} &\propto \frac{d}{dy} (y^{(n/2)-1} e^{-y/2}) \\ &= \left(\frac{n}{2} - 1\right) y^{(n/2)-2} e^{-y/2} - \frac{1}{2} y^{(n/2)-1} e^{-y/2} \\ &= \frac{y^{(n/2)-2} e^{-y/2}}{2} \{(n-2) - y\} = 0\end{aligned}$$

より、 $y = n - 2$ である。

ただし、 $f_n(y)$ が極大値をもつのは $n \geq 3$ のときで、 $n = 2$ のときは下限 $y = 0$ で最大となり、 $n = 1$ のときは $y \rightarrow +0$ で発散する。

確率密度関数

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

$$f_1(y) = \chi_1^2(y) = \frac{y^{-1/2} e^{-y/2}}{2^{1/2} \Gamma(1/2)} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}$$

$$f_2(y) = \chi_2^2(y) = \frac{y^0 e^{-y/2}}{2\Gamma(1)} = \frac{1}{2} e^{-y/2}$$

$n = 2$ のときは指数分布

$$f_3(y) = \chi_3^2(y) = \frac{y^{1/2} e^{-y/2}}{2^{3/2} \Gamma(3/2)} = \sqrt{\frac{y}{2\pi}} e^{-y/2}$$

$$f_4(y) = \chi_4^2(y) = \frac{y^1 e^{-y/2}}{2^2 \Gamma(2)} = \frac{1}{4} y e^{-y/2}$$

確率密度関数

自由度 $n = 5, 6, 7, 8$ に対して、カイ2乗分布の確率密度関数は、次のようになる。

$$f_5(y) = \chi_5^2(y) = \frac{y^{3/2}e^{-y/2}}{2^{5/2}\Gamma(5/2)} = \frac{1}{3\sqrt{2\pi}}\sqrt{y^3}e^{-y/2}$$

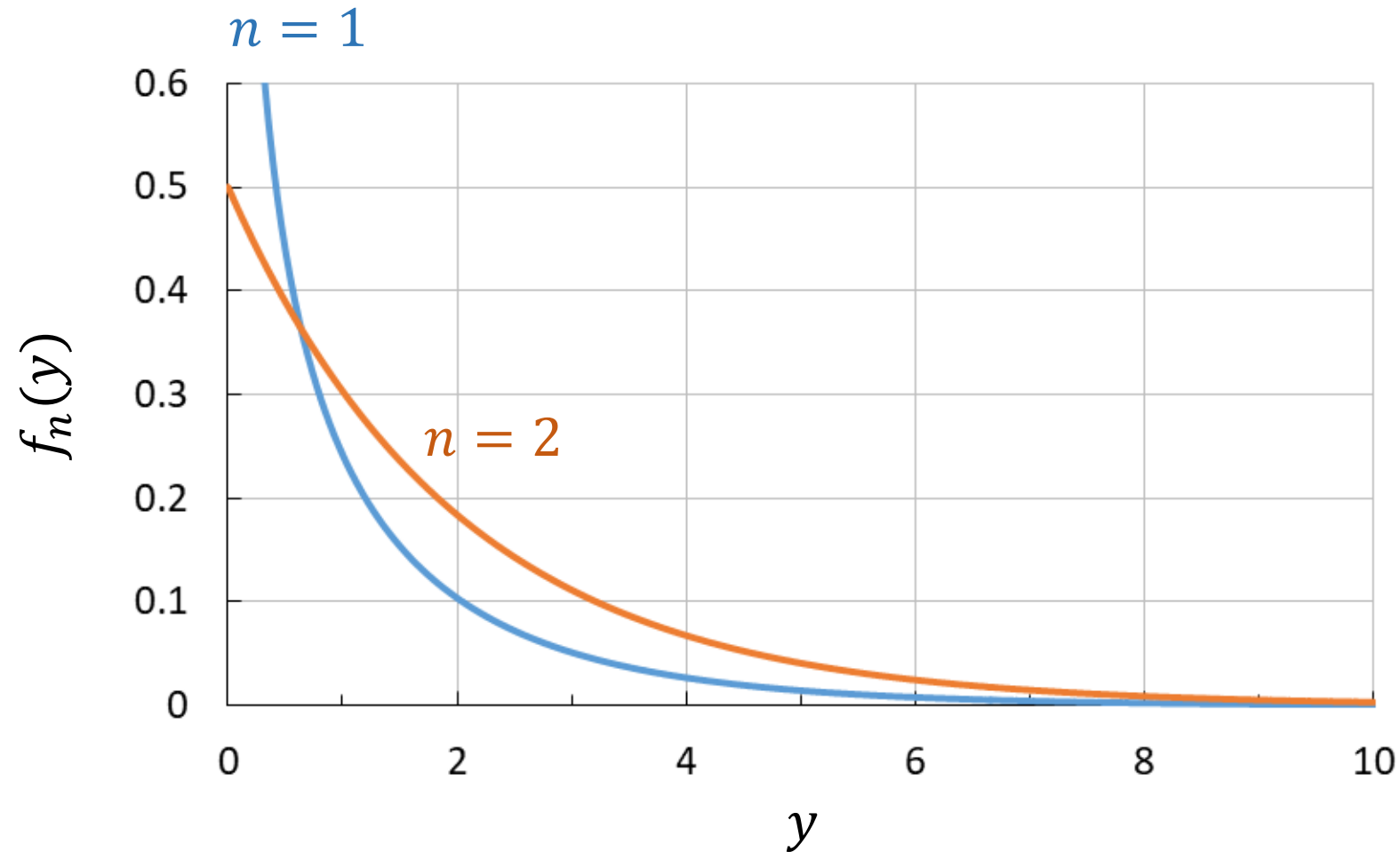
$$f_6(y) = \chi_6^2(y) = \frac{y^2e^{-y/2}}{2^3\Gamma(3)} = \frac{1}{16}y^2e^{-y/2}$$

$$f_7(y) = \chi_7^2(y) = \frac{y^{5/2}e^{-y/2}}{2^{7/2}\Gamma(7/2)} = \frac{1}{15\sqrt{2\pi}}\sqrt{y^5}e^{-y/2}$$

$$f_8(y) = \chi_8^2(y) = \frac{y^3e^{-y/2}}{2^4\Gamma(4)} = \frac{1}{96}y^3e^{-y/2}$$

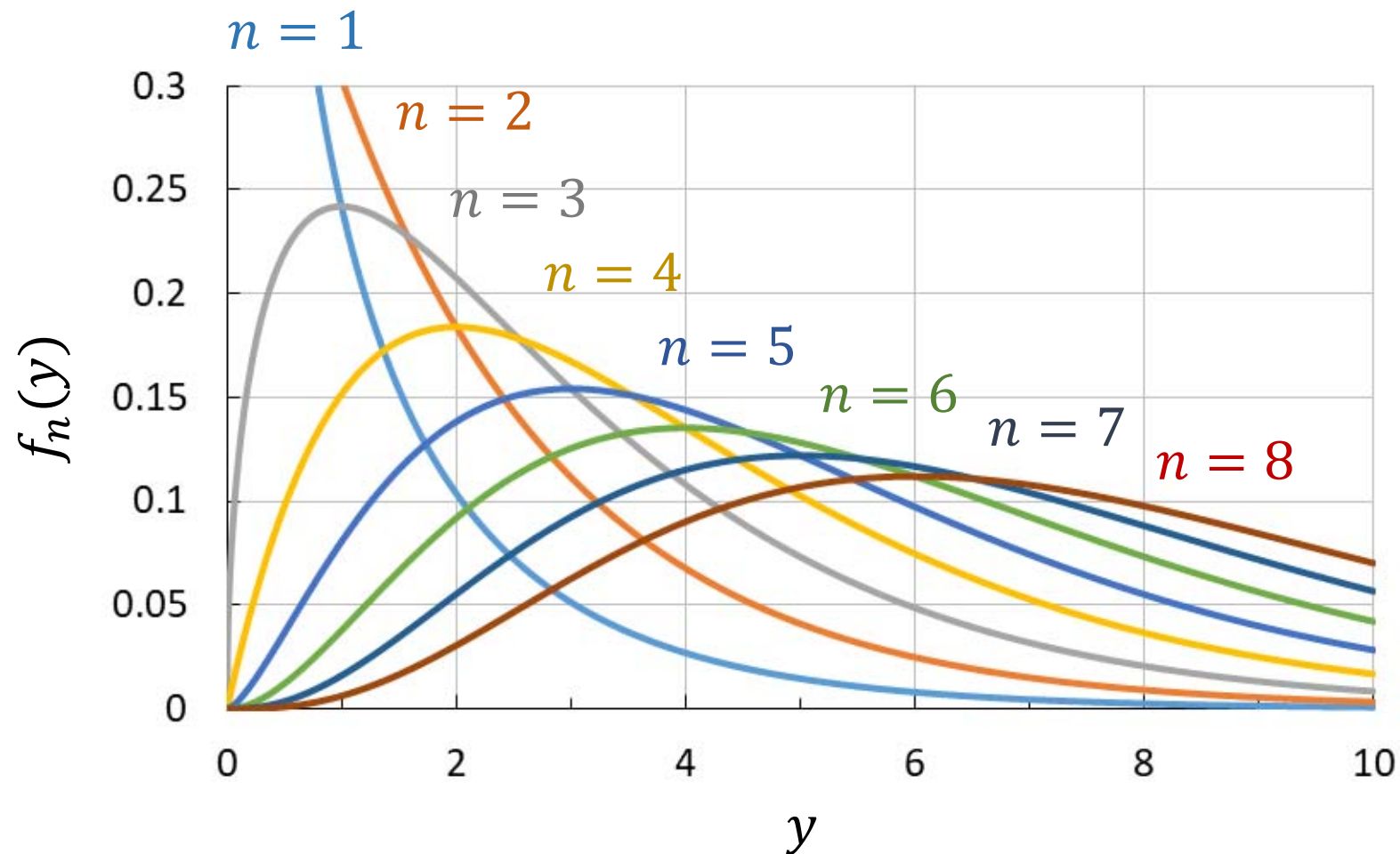
確率密度関数

自由度 $n = 1, 2$ のカイ2乗分布の確率密度関数のグラフを下図に示す。



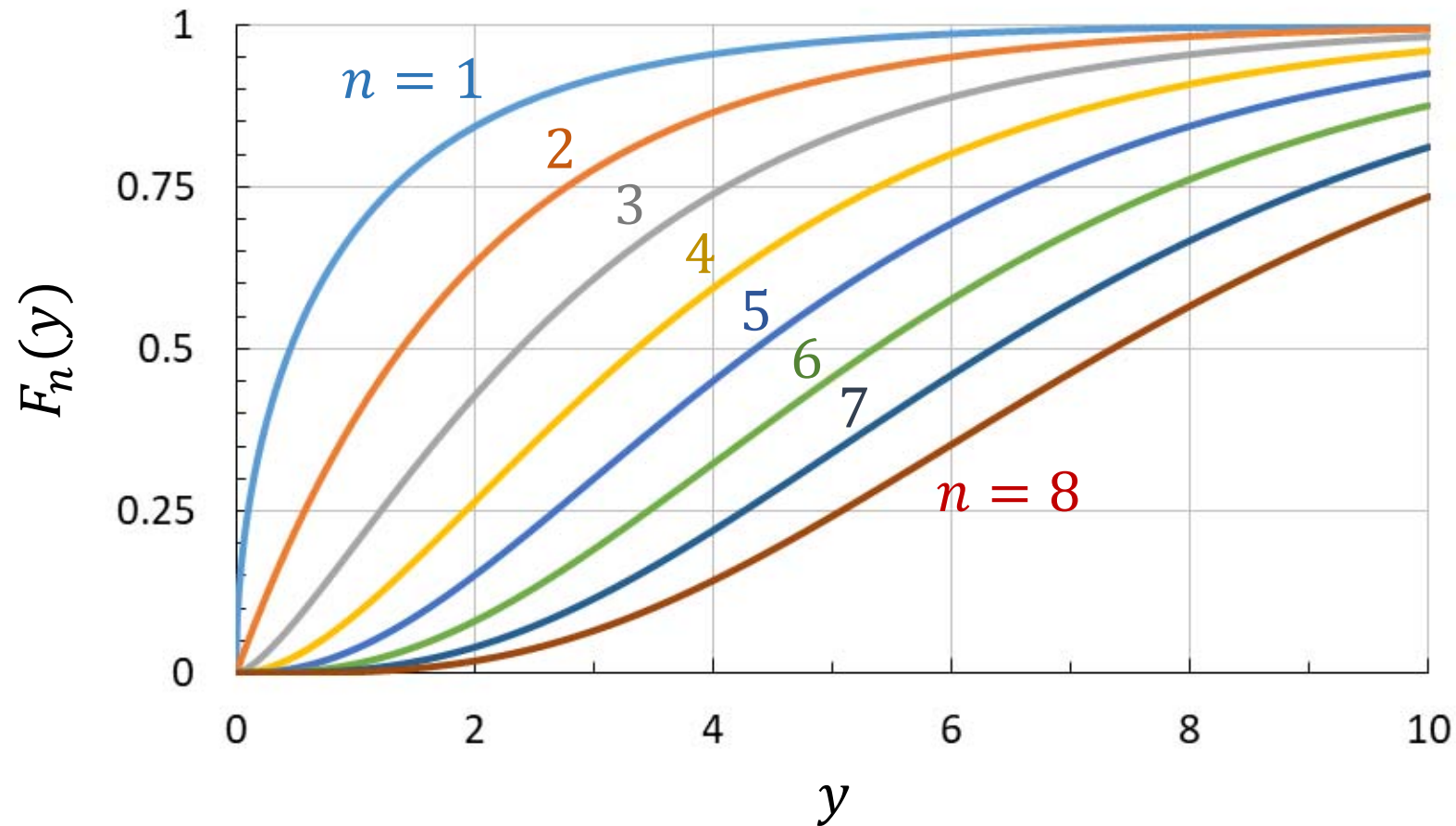
確率密度関数

$n = 1, 2, \dots, 8$ のカイ2乗分布の確率密度関数のグラフを下図に示す。



累積分布関数

$n = 1, 2, \dots, 8$ のカイ2乗分布の累積分布関数のグラフを下図に示す。



確率密度関数 ($n = 1$) の導出

確率変数 X が標準正規分布 $f(X) = \frac{1}{\sqrt{2\pi}} e^{-X^2/2}$ にしたがうとき、

その2乗 $y = X^2$ の確率密度関数を $f_1(y)$ とすると

$$\begin{aligned}\int_{\alpha}^{\beta} f_1(y) dy &= \int_{\sqrt{\alpha}}^{\sqrt{\beta}} f(X) dX + \int_{-\sqrt{\beta}}^{-\sqrt{\alpha}} f(X) dX \\ &= \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\alpha}}^{\sqrt{\beta}} e^{-X^2/2} dX + \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{\beta}}^{-\sqrt{\alpha}} e^{-X^2/2} dX \\ &= \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\alpha}}^{\sqrt{\beta}} e^{-X^2/2} dX\end{aligned}$$

確率密度関数($n = 1$)の導出

標準正規分布にしたがう確率変数 X の2乗 $y = X^2$ に対して

$$\begin{aligned}\int_{\alpha}^{\beta} f_1(y) dy &= \frac{2}{\sqrt{2\pi}} \int_{\sqrt{\alpha}}^{\sqrt{\beta}} e^{-X^2/2} dX \\ &= \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-y/2} \frac{dX}{dy} dy = \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-y/2} \frac{d}{dy} (\sqrt{y}) dy \\ &= \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-y/2} \frac{1}{2\sqrt{y}} dy = \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} \frac{e^{-y/2}}{\sqrt{y}} dy \\ \therefore f_1(y) &= \frac{e^{-y/2}}{\sqrt{2\pi y}}\end{aligned}$$

確率密度関数 ($n = 2$) の導出

確率変数 X_i が標準正規分布にしたがい、その2乗 $y = X_i^2$ が確率

密度関数 $f_1(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}}$ にしたがうとき、2個の2乗和 $y = X_1^2 + X_2^2$

の確率密度関数 $f_2(y)$ は、 $f_1(y)$ の畳込みだから

$$\begin{aligned} f_2(y) &= \int_0^y f_1(y-t)f_1(t)dt \\ &= \int_0^y \frac{e^{-(y-t)/2}}{\sqrt{2\pi(y-t)}} \frac{e^{-t/2}}{\sqrt{2\pi t}} dt \\ &= \frac{e^{-y/2}}{2\pi} \int_0^y \frac{dt}{\sqrt{(y-t)t}} \end{aligned}$$

$X_2^2 = t$ のとき、 $y = X_1^2 + X_2^2$ になるのは $X_1^2 = y - t$ の場合だから、 $f_1(y-t)f_1(t)$ を $t = 0 \sim y$ で積分すれば $f_2(y)$ が求められる。

確率密度関数 ($n = 2$) の導出

2個の2乗和 $y = X_1^2 + X_2^2$ に対して

$$\begin{aligned} f_2(y) &= \frac{e^{-y/2}}{2\pi} \int_0^y \frac{dt}{\sqrt{(y-t)t}} \\ &= \frac{e^{-y/2}}{2\pi} \int_0^y \frac{dt}{\sqrt{\frac{y^2}{4} - \left(t - \frac{y}{2}\right)^2}} = \frac{e^{-y/2}}{2\pi} \left[\sin^{-1} \left(\frac{t - \frac{y}{2}}{\frac{y}{2}} \right) \right]_{t=0}^{t=y} \\ &= \frac{e^{-y/2}}{2\pi} (\sin^{-1} 1 + \sin^{-1} 1) = \frac{e^{-y/2}}{2} \end{aligned}$$

確率密度関数 ($n = 3$) の導出

$X_3^2 = t$ のとき、 $y = X_1^2 + X_2^2 + X_3^2$ になるのは $X_1^2 + X_2^2 = y - t$ の場合だから、 $f_2(y - t)f_1(t)$ を $t = 0 \sim y$ で積分すれば $f_3(y)$ が求められる。

確率変数 x_i が標準正規分布にしたがい、その2乗 $y = X_i^2$ が確率

密度関数 $f_1(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}}$ に、2個の2乗和 $y = X_1^2 + X_2^2$ が確率密度

関数 $f_2(y) = \frac{e^{-y/2}}{2}$ にしたがうとき、3個の2乗和 $y = X_1^2 + X_2^2 + X_3^2$

の確率密度関数 $f_3(y)$ は、 $f_2(y)$ と $f_1(y)$ の畳込みだから

$$\begin{aligned} f_3(y) &= \int_0^y f_2(y-t)f_1(t)dt = \int_0^y \frac{e^{-(y-t)/2}}{2} \frac{e^{-t/2}}{\sqrt{2\pi t}} dt \\ &= \frac{e^{-y/2}}{2\sqrt{2\pi}} \int_0^y \frac{dt}{\sqrt{t}} = \frac{e^{-y/2}}{2\sqrt{2\pi}} [2\sqrt{t}]_0^y = \frac{e^{-y/2}}{2\sqrt{2\pi}} 2\sqrt{y} = \sqrt{\frac{y}{2\pi}} e^{-y/2} \end{aligned}$$

確率密度関数 ($n = k$) の証明

確率変数 X_i が標準正規分布にしたがい、その2乗 $y = X_i^2$ が確率

密度関数 $f_1(y) = \frac{e^{-y/2}}{\sqrt{2\pi y}}$ に、 $(k - 1)$ 個の2乗和 $y = \sum_{i=1}^{k-1} X_i^2$ が

確率密度関数 $f_{k-1}(y) = \frac{y^{((k-1)/2)-1} e^{-y/2}}{2^{(k-1)/2} \Gamma((k-1)/2)}$ にしたがうとき、

k 個の2乗和 $y = \sum_{i=1}^k X_i^2$ の確率密度関数 $f_k(y)$ は、 $f_{k-1}(y)$ と $f_1(y)$

の畳込みである。

$X_k^2 = t$ のとき、 $y = \sum_{i=1}^k X_i^2$ になるのは $\sum_{i=1}^{k-1} X_i^2 = y - t$ の場合だから、 $f_{k-1}(y - t)f_1(t)$ を $t = 0 \sim y$ で積分すれば $f_k(y)$ が求められる。

確率密度関数 ($n = k$) の証明

k 個の2乗和 $y = \sum_{i=1}^k X_i^2$ の確率密度関数

$f_k(y)$ は $f_{k-1}(y)$ と $f_1(y)$ の畳込みだから

$$f_k(y) = \int_0^y f_{k-1}(y-t)f_1(t)dt$$

$$= \int_0^y \frac{(y-t)^{((k-1)/2)-1} e^{-(y-t)/2}}{2^{(k-1)/2} \Gamma((k-1)/2)} \frac{e^{-t/2}}{\sqrt{2\pi t}} dt$$

$$= \frac{e^{-y/2}}{2^{k/2} \Gamma((k-1)/2) \sqrt{\pi}} \int_0^y (y-t)^{((k-1)/2)-1} t^{-1/2} dt$$

$X_k^2 = t$ のとき、

$y = \sum_{i=1}^k X_i^2$ になるのは

$\sum_{i=1}^{k-1} X_i^2 = y - t$ の場合だから、

$f_{k-1}(y-t)f_1(t)$ を $t = 0 \sim y$ で積分すれば $f_k(y)$ が求められる。

確率密度関数 ($n = k$) の証明

ここで、 $t = yu$ とおくと

$$\begin{aligned} & \int_0^y (y - t)^{((k-1)/2)-1} t^{-1/2} dt \\ &= y^{((k-1)/2)-1/2} \int_0^1 (1 - u)^{((k-1)/2)-1} u^{-1/2} du \\ &= y^{(k/2)-1} \int_0^1 (1 - u)^{((k-1)/2)-1} u^{(1/2)-1} du \\ &= y^{(k/2)-1} B((k-1)/2, 1/2) \\ &= y^{(k/2)-1} \frac{\Gamma((k-1)/2) \Gamma(1/2)}{\Gamma(k/2)} \end{aligned}$$

ベータ関数

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

ベータ関数とガンマ関数の関係

$$B(p, q) \Gamma(p + q) = \Gamma(p) \Gamma(q)$$

確率密度関数($n = k$)の証明

$f_k(y)$ が $f_{k-1}(y)$ と $f_1(y)$ の畳込みであることより

$$\begin{aligned} f_k(y) &= \frac{e^{-y/2}}{2^{k/2}\Gamma((k-1)/2)\sqrt{\pi}} \int_0^y (y-t)^{((k-1)/2)-1} t^{-1/2} dt \\ &= \frac{y^{(k/2)-1} e^{-y/2}}{2^{k/2}\Gamma((k-1)/2)\sqrt{\pi}} \int_0^1 (1-u)^{((k-1)/2)-1} u^{(1/2)-1} du \\ &= \frac{y^{(k/2)-1} e^{-y/2}}{2^{k/2}\Gamma((k-1)/2)\sqrt{\pi}} \frac{\Gamma((k-1)/2)\Gamma(1/2)}{\Gamma(k/2)} \\ &= \frac{y^{(k/2)-1} e^{-y/2}}{2^{k/2}\Gamma(k/2)} \end{aligned}$$

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

数学的帰納法により、任意の n に対して y はカイ2乗分布にしたがう。

正規分布の標準化

母平均 μ 、母分散 σ^2 の正規分布 $f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ 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 $x_i (i = 1, 2, \dots, n)$ に対して、 $X_i = \frac{x_i - \mu}{\sigma}$ は標準正規分布 $f(X_i) = \frac{1}{\sqrt{2\pi}} e^{-\frac{X_i^2}{2}}$ にしたがう

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

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

偏差の2乗和と母分散

$X_i = \frac{x_i - \mu}{\sigma}$ の母平均 μ を、標本平均 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 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

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

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。

標本標準偏差の期待値

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

は自由度 $n - 1$ のカイ2乗分布にしたがうから、標本標準偏差

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

の期待値は

$$\bar{s} = \int_0^{\infty} s f_{n-1}(Y) dY = \frac{\sigma / \sqrt{n-1}}{2^{(n-1)/2} \Gamma((n-1)/2)} \int_0^{\infty} Y^{(n-1)/2-1/2} e^{-Y/2} dY$$

標本標準偏差の期待値

標本標準偏差 s の期待値は

$$\begin{aligned}\bar{s} &= \int_0^{\infty} s f_{n-1}(Y) dY = \frac{\sigma / \sqrt{n-1}}{2^{(n-1)/2} \Gamma((n-1)/2)} \int_0^{\infty} Y^{(n-1)/2-1/2} e^{-Y/2} dY \\&= \frac{\sigma / \sqrt{n-1}}{2^{(n-1)/2} \Gamma((n-1)/2)} \int_0^{\infty} Y^{(n/2)-1} e^{-Y/2} dY \\&= \frac{\sigma / \sqrt{n-1}}{2^{(n-1)/2} \Gamma((n-1)/2)} 2^{n/2} \Gamma(n/2) \\&= \sigma \sqrt{\frac{2}{n-1} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}}\end{aligned}$$

標本標準偏差の期待値

標本標準偏差 s の期待値は

$$\bar{s} = \sigma \sqrt{\frac{2}{n-1} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}} = c_4 \sigma$$

であり、母標準偏差 σ に一致しない。すなわち、標本標準偏差(不偏分散の正の平方根) s は母標準偏差 σ の不偏推定量ではない。

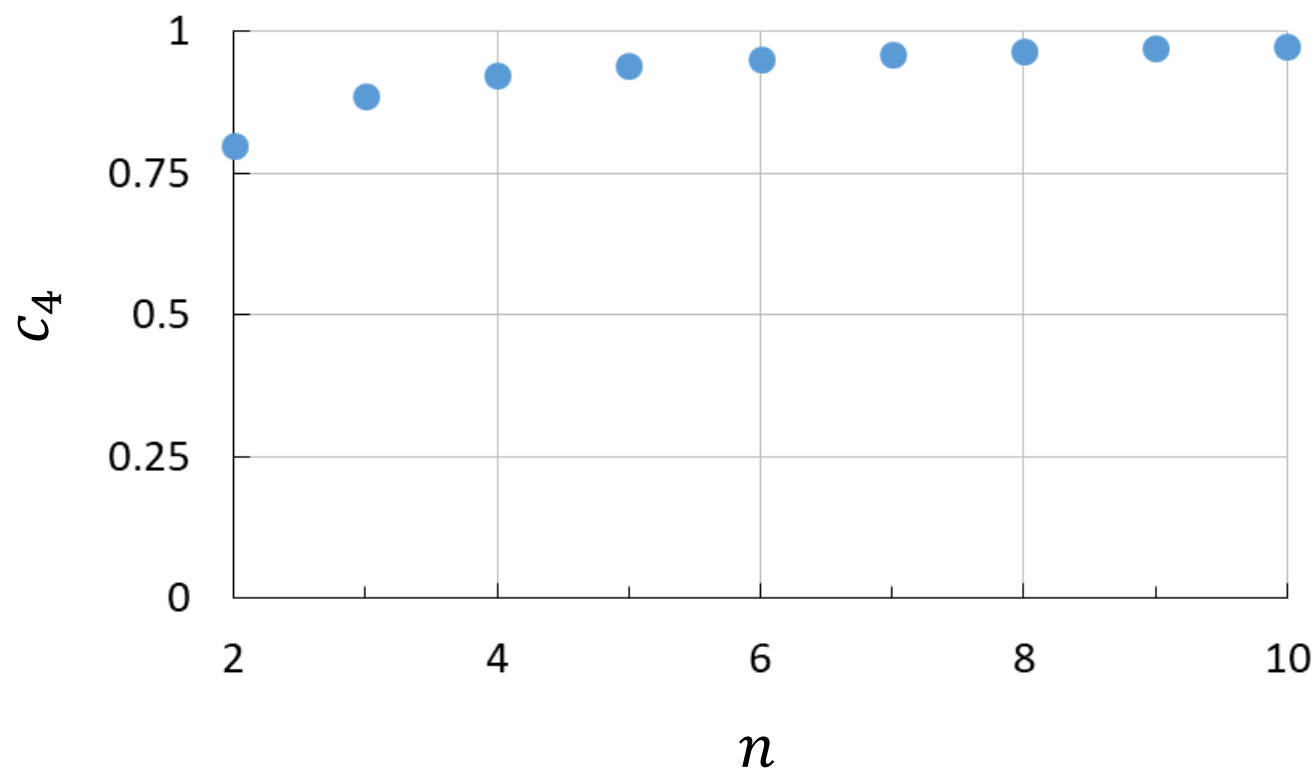
$$c_4 = \frac{\bar{s}}{\sigma} = \sqrt{\frac{2}{n-1} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}}$$

は標本標準偏差 s と母標準偏差 σ の比を表し、標本のサイズ n が大きくなると 1 に近づく。 c_4 は、品質管理で用いられる $\bar{X} - s$ 管理図で、 s の管理限界を定める際に用いられる。

標本標準偏差と母標準偏差の比

標本標準偏差 s と母標準偏差 σ の比 c_4 の値を下に示す。

n	c_4
2	0.797885
3	0.886227
4	0.921318
5	0.939986
6	0.951533
7	0.959369
8	0.965030
9	0.969311
10	0.972659



標本標準偏差と母標準偏差の比

標本のサイズが $n \rightarrow \infty$ のとき、

$$c_4 = \frac{\bar{s}}{\sigma} = \sqrt{\frac{2}{n-1} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}} \rightarrow 1$$

となることを以下に示す。

ガンマ関数 $\Gamma(z)$ の定義より

$$\begin{aligned}\Gamma(z) &= \int_0^{\infty} s^{z-1} e^{-s} ds = \lim_{N \rightarrow \infty} \int_0^N s^{z-1} e^{-s} ds \\ &= \lim_{N \rightarrow \infty} \int_0^N s^{z-1} \left(1 - \frac{s}{N}\right)^N ds\end{aligned}$$

標本標準偏差と母標準偏差の比

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

$u = \frac{s}{N}$ と変数変換すると、 $du = \frac{ds}{N}$ だから

$$\begin{aligned} \int_0^N s^{z-1} \left(1 - \frac{s}{N}\right)^N ds &= N^z \int_0^1 u^{z-1} (1-u)^N du = N^z B(z, N+1) \\ &= N^z \frac{\Gamma(z)\Gamma(N+1)}{\Gamma(z+N+1)} \end{aligned}$$

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

標本標準偏差と母標準偏差の比

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

ここで、 $z = \frac{1}{2}$, $N+1 = \frac{n-1}{2}$ とすると

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

$$= \lim_{n \rightarrow \infty} \sqrt{\frac{n-1}{2} \frac{\Gamma((n-1)/2)}{\Gamma(n/2)}} = \lim_{n \rightarrow \infty} \frac{1}{c_4}$$

ゆえに、 $n \rightarrow \infty$ のとき $c_4 = \frac{\bar{s}}{\sigma} \rightarrow 1$ である。

ガンマ関数

ガンマ関数 $\Gamma(z)$ は

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

で定義される。

$$\begin{aligned}\Gamma(z) &= \int_0^{\infty} s^{z-1} e^{-s} ds = [-s^{z-1} e^{-s}]_0^{\infty} + (z-1) \int_0^{\infty} s^{z-1} e^{-s} ds \\ &= (z-1) \Gamma(z-1)\end{aligned}$$

$$\Gamma(1) = \int_0^{\infty} e^{-s} ds = [-e^{-s}]_0^{\infty} = 1$$

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

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

$s = u^2$ とおくと

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right) &= \int_0^{\infty} \frac{e^{-s}}{\sqrt{s}} ds = \int_0^{\infty} \frac{e^{-u^2}}{u} 2u du = 2 \int_0^{\infty} e^{-u^2} du \\ &= \int_{-\infty}^{\infty} e^{-u^2} du = \sqrt{\pi}\end{aligned}$$

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

$$(2m-1)!! = (2m-1) \cdot (2m-3) \cdot \dots \cdot 3 \cdot 1$$

ベータ関数

ベータ関数 $B(p, q)$ は

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

で定義される。

$u = 1 - t$ とおくと

$$\begin{aligned} B(p, q) &= \int_0^1 t^{p-1} (1-t)^{q-1} dt = \int_1^0 (1-u)^{p-1} u^{q-1} (-du) \\ &= \int_0^1 u^{q-1} (1-u)^{p-1} du = B(q, p) \end{aligned}$$

だから、 $B(p, q) = B(q, p)$ である。

ベータ関数とガンマ関数

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

より

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

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

$$dudv = |-s| ds dt = s ds dt$$

ベータ関数とガンマ関数

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

$u = st, v = s(1 - t)$ と変数変換すると、 $du dv = s ds dt$ より

$$\begin{aligned} B(p, q)\Gamma(p + q) &= \int_0^\infty \int_0^\infty t^{p-1}(1 - t)^{q-1} s^{p+q-2} e^{-s} du dv \\ &= \int_0^\infty \int_0^\infty u^{p-1} v^{q-1} e^{-(u+v)} du dv \\ &= \int_0^\infty u^{p-1} e^{-u} du \int_0^\infty v^{q-1} e^{-v} dv = \Gamma(p)\Gamma(q) \end{aligned}$$

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

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.