2022 年 08 月 02 日

カイ自乗分布の密度関数

新潟工科大学 基礎教育・教養系 竹野茂治

1 はじめに

本稿では、講義で証明を省略した、自由度 n のカイ自乗分布 $\chi^2(n)$ の密度関数が

$$g(v)=\begin{cases}\dfrac{2^{-n/2}}{\Gamma(n/2)}v^{n/2-1}e^{-v/2} & (v>0)\\ 0 & (v<0)\end{cases} \tag{1}$$

となることを示す。ここで $\Gamma(p)$ はガンマ関数

$$\Gamma(p)=\int_0^\infty x^{p-1}e^{-x}dx \quad (p>0) \tag{2}$$

であり、(1) に必要なガンマ関数 $\Gamma(p)$ の性質などについても紹介する。

2 動径方向の関数の積分

まず、次のような積分を考える。

$$I_K=\int_{B_K} g(|\vec{x}|)d\vec{x} \tag{3}$$

ここで、$\vec{x}=(x_1,x_2,\ldots,x_n)\in\mathrm{R}^n$、$B_K$ は原点中心で半径が K の n 次元球

$$B_K=\{\vec{x}\mid|\vec{x}|=\sqrt{x_1^2+\cdots+x_n^2}\le K\} \tag{4}$$

$g=g(r)$ は $r\ge 0$ 上の 1 変数関数である。

積分領域が球なので、通常は n 次元極座標

$$\begin{cases} x_1 &= r\cos\theta_1, \\ x_2 &= r\sin\theta_1\cos\theta_2, \\ x_3 &= r\sin\theta_1\sin\theta_2\cos\theta_3, \\ \cdots & \\ x_{n-1} &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\cos\theta_{n-1}, \\ x_n &= r\sin\theta_1\sin\theta_2\cdots\sin\theta_{n-2}\sin\theta_{n-1} \end{cases} \tag{5}$$

$$(r \geq 0,\ \ 0 \leq \theta_1 \leq \pi, \ldots,\ \ 0 \leq \theta_{n-2} \leq \pi, 0 \leq \theta_{n-1} < 2\pi)$$

に変換して積分する。このとき、この極座標 (5) のヤコビ行列式 J は

$$J = J(r, \theta_1, \ldots, \theta_{n-1}) = \left| \frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(r, \theta_1, \ldots, \theta_{n-1})} \right| = r^{n-1}\hat{J}(\theta_1, \ldots, \theta_{n-1}) \tag{6}$$

の形となる。よって、(3) は

$$I_K = \int_0^K g(r)r^{n-1}dr \int_0^\pi d\theta_1 \cdots \int_0^\pi d\theta_{n-2} \int_0^{2\pi} \hat{J}d\theta_{n-1} = A\int_0^K r^{n-1}g(r)dr \tag{7}$$

となる。

ここで、$g(r) \equiv 1$ ならば (7) は B_K の体積 $V_n(K)$ となるので、

$$V_n(K) = A\int_0^K r^{n-1}dr = \frac{AK^n}{n}$$

より A は

$$A = \frac{nV_n(K)}{K^n} = nV_n(1) \tag{8}$$

となる。[1] で見たように、半径 1 の n 次元球の体積 $V_n(1)$ は以下の式で表される。

$$V_n(1) = \begin{cases} \dfrac{2^m\pi^m}{(2m)!!} & (n = 2m) \\ \dfrac{2^{m+1}\pi^m}{(2m+1)!!} & (n = 2m+1) \end{cases} \tag{9}$$

ここで $n!!$ は n から 2 ずつ減らして 1, または 2 までかけたもの

$$n!! = n(n-2)(n-4)\cdots \tag{10}$$

である。

3 ガンマ関数

次は、前節の半径 1 の n 次元球の体積 (9) をガンマ関数 $\Gamma(p)$ で表すことを考える。

ガンマ関数 (2) は以下のような性質を持つことを講義で紹介した。

$$\Gamma(p+1) = p\,\Gamma(p)\ \ (p>0),\quad \Gamma(1)=1,\quad \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi} \tag{11}$$

これらについても説明する。

まず、(2) は $x\to\infty$ に関して広義積分であるが、$0<p<1$ の p に対しては、$p-1<0$ より $x\to+0$ の方も広義積分になっていることに注意する。

しかし、$x\to+0$ では、e^{-x} は有界 ($e^{-x}\le 1$) で x^{p-1} は可積分 (広義積分は収束) なので、$x\to+0$ での広義積分は収束する。また、$x\to\infty$ では、$h_{p-1}(x)=x^{p-1}e^{-x/2}$ は有界で $e^{-x/2}$ は可積分なので、$x\to\infty$ の広義積分も収束する。なお、$h_{p-1}(x)$ の有界性は、

$$h'_{p-1}(x) = (p-1)x^{p-2}e^{-x/2} - \frac{1}{2}x^{p-1}e^{-x/2} = \frac{1}{2}(2p-2-x)x^{p-2}e^{-x/2}$$

より $x>2p-2$ では減少することからわかる。よって (2) は $p>0$ に対して収束する。

$p>0$ に対し、部分積分により、

$$\begin{aligned}\Gamma(p+1) &= \int_0^\infty x^p e^{-x}dx = \lim_{K\to\infty}\int_0^K x^p(-e^{-x})'dx \\ &= \lim_{K\to\infty}\left\{\left[-x^pe^{-x}\right]_0^K + p\int_0^K x^{p-1}e^{-x}dx\right\} = -\lim_{K\to\infty}K^pe^{-K} + p\,\Gamma(p)\end{aligned}$$

となるが、$K^p e^{-K}$ の極限は、$K^p e^{-K/2} = h_p(K)$ が有界で、$e^{-K/2} \to 0$ となるから 0 に収束し、これで (11) の最初のものが得られる。残りのものは、直接計算して、

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty e^{-x}dx \ = \ \left[-e^{-x}\right]_0^\infty \ = \ 1, \\
\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty x^{-1/2}e^{-x}dx \quad (x = t^2) \\
&= \int_0^\infty t^{-1}e^{-t^2}2tdx \ = \ 2\int_0^\infty e^{-t^2}dt \ = \ \sqrt{\pi}
\end{aligned}$$

となる (最後の積分については、例えば [2] 参照)。

これらを用いると、自然数 m に対し、

$$\begin{aligned}
2^m\,\Gamma(m+1) &= 2^m m! \ = \ 2^m m(m-1)(m-2)\cdots 2\cdot 1\,\Gamma(1) \\
&= (2m)(2m-2)(2m-4)\cdot\cdots\cdot 4\cdot 2 \ = \ (2m)!!, \\
2^{m+1}\,\Gamma\left(m+\frac{3}{2}\right) &= 2^{m+1}\left(m+\frac{1}{2}\right)\left(m-\frac{1}{2}\right)\cdots\frac{3}{2}\cdot\frac{1}{2}\,\Gamma\left(\frac{1}{2}\right) \\
&= (2m+1)(2m-1)\cdots 3\cdot 1\sqrt{\pi} \ = \ (2m+1)!!\sqrt{\pi}
\end{aligned}$$

となるので、(9) にこれらを用いると、$n = 2m$ のときは、

$$V_{2m}(1) = \frac{2^m\pi^m}{(2m)!!} = \frac{2^m\pi^m}{2^m\,\Gamma(m+1)} = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$$

となり、$n = 2m+1$ のときは、

$$V_{2m+1}(1) = \frac{2^{m+1}\pi^m}{(2m+1)!!} = \frac{2^{m+1}\pi^m\sqrt{\pi}}{2^{m+1}\,\Gamma(m+3/2)} = \frac{\pi^{m+1/2}}{\Gamma(m+3/2)} = \frac{\pi^{n/2}}{\Gamma(n/2+1)}$$

となるので、結局 $V_n(1)$ はすべての自然数 n に対して

$$V_n(1) = \frac{\pi^{n/2}}{\Gamma(n/2+1)} \tag{12}$$

と書けることになる。よって、(8) より A は

$$A = nV_n(1) = \frac{n\pi^{n/2}}{\Gamma(n/2+1)} = \frac{n\pi^{n/2}}{(n/2)\,\Gamma(n/2)} = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

となるので、結局 (7) の I_k は

$$I_K = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^K r^{n-1} g(r) dr \tag{13}$$

と書けることになる。

$r^{n-1}g(r)$ が $r \geq 0$ で可積分ならば、(13) で $K \to \infty$ とすれば、

$$I_\infty = \int_{\mathrm{R}^n} g(\vec{x}) d\vec{x} = \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty r^{n-1} g(r) dr \tag{14}$$

も得られる。

4 カイ自乗分布

確率変数 $u_1, \ldots, u_n$ が独立で、$u_j \sim N(0,1)$ $(1 \leq j \leq n)$ のとき、$v = u_1^2 + \cdots + u_n^2$ が従う確率分布を「自由度 n のカイ自乗分布 $\chi^2(n)$」と呼ぶ。この v の密度関数 $g(v)$ を求める。

$N(0,1)$ の密度関数を

$$f_0(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \tag{15}$$

とする。$u_1, \ldots, u_n$ は独立なので、n 次元確率分布 $(u_1, \ldots, u_n)$ の密度関数は

$$f_0(u_1) \cdots f_0(u_n) = \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-|\vec{u}|^2/2}$$

となる ([3])。よって、v の分布関数 $G(v)$ は、

$$\begin{aligned} G(t) &= \mathrm{Prob}\{v \leq t\} \;=\; \mathrm{Prob}\{|\vec{u}|^2 = u_1^2 + \cdots + u_n^2 \leq t\} \\ &= \int_{\{\vec{u} \mid |\vec{u}|^2 \leq t\}} \left(\frac{1}{\sqrt{2\pi}} \right)^n e^{-|\vec{u}|^2/2} d\vec{u} \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^n \int_{\mathrm{R}^n} e^{-|\vec{u}|^2/2} \xi_{\{\vec{u} \mid |\vec{u}|^2 \leq t\}}(\vec{u}) d\vec{u} \end{aligned} \tag{16}$$

となる。ここで、$H \subset \mathrm{R}^n$ に対して、$\xi_H(\vec{x})$ を H の定義関数、すなわち

$$\xi_H(\vec{x}) = \begin{cases} 1 & (\vec{x} \in H) \\ 0 & (\vec{x} \notin H) \end{cases} \tag{17}$$

とする。今の場合、

$$\xi_{\{\vec{u}|\ |\vec{u}|^2 \leq t\}}(\vec{u}) = \xi_{\{r|\ r^2 \leq t\}}(|\vec{u}|)$$

と 1 次元の定義関数を使って書けるから、(16) の被積分関数は $|\vec{u}|$ の関数、すなわち動径方向のみに依存し、よって (14) により、

$$G(t) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \frac{2\pi^{n/2}}{\Gamma(n/2)} \int_0^\infty r^{n-1} e^{-r^2/2} \xi_{\{r|\ r^2 < t\}} dr \tag{18}$$

となり、$t \leq 0$ に対しては $G(t) = 0$, $t > 0$ に対しては、

$$G(t) = \frac{2^{1-n/2}}{\Gamma(n/2)} \int_0^{\sqrt{t}} r^{n-1} e^{-r^2/2} dr \tag{19}$$

となることがわかる。よって、v の密度関数 $g(v) = G'(v)$ は、$v < 0$ では 0 で、$v > 0$ では

$$g(v) = G'(v) = \frac{2^{1-n/2}}{\Gamma(n/2)} (\sqrt{v})^{n-1} e^{-v/2} \frac{1}{2\sqrt{v}} = \frac{2^{-n/2}}{\Gamma(n/2)} v^{n/2-1} e^{-v/2}$$

となって、これで (1) が得られたことになる。

参考文献

[1] 竹野茂治「n 次元球の体積について」(2007)
`http://takeno.iee.niit.ac.jp/~shige/math/lecture/basic3/data/ball1.pdf`

[2] 竹野茂治「正規確率変数の一次式」(2022)
`http://takeno.iee.niit.ac.jp/~shige/math/lecture/graduate/data/normal1.pdf`

[3] 竹野茂治「多次元確率分布と独立性」(2022)
`http://takeno.iee.niit.ac.jp/~shige/math/lecture/graduate/data/ndimrandvar1.pdf`