2022 年 08 月 23 日

正規母集団の平方和の標本分布

新潟工科大学 基礎教育・教養系 竹野茂治

1 はじめに

講義では、正規母集団から取った n 個のデータの平方和を母分散で割ったものが自由度 $(n-1)$ のカイ自乗分布に従うことを述べ、$n=2,3$ のときにそれを具体的に説明したが、本稿では、一般の n に対する証明を行う。

なお、本稿では現代的な公理的確率論ではなく、古典的確率論の範疇で考える。

2 定式化

自由度 n のカイ自乗分布 $\chi^2(n)$ は、$u_1,\ldots,u_n$ が $N(0,1)$ に従う確率変数で、かつ独立であるとき、

$$v=u_1^2+\cdots+u_n^2 \tag{1}$$

が従う確率分布である。

本稿で証明するのは、$x_1,\ldots,x_n$ を正規母集団から取ったデータ、すなわち x_k が $N(\mu,\sigma^2)$ に従う確率変数でかつそれらが独立であるとき、標本平均 $\bar{x}$, 平方和 S

$$\bar{x}=\frac{1}{n}\sum_{k=1}^{n}x_k,\quad S=\sum_{k=1}^{n}(x_k-\bar{x})^2 \tag{2}$$

に対して、

$$\frac{S}{\sigma^2}\sim\chi^2(n-1) \tag{3}$$

となること、である。

$n=2$ のときは

$$u_1 = \frac{1}{\sqrt{2}\,\sigma}(x_1 - x_2) \tag{4}$$

によって、$n=3$ のときは

$$u_1 = \frac{1}{\sqrt{2}\,\sigma}(x_1 - x_3), \quad u_2 = \frac{1}{\sqrt{6}\,\sigma}(-x_1 + 2x_2 - x_3) \tag{5}$$

によって

$$\frac{S}{\sigma^2} = \sum_{k=1}^{n-1} u_k^2 \tag{6}$$

と書け、かつ $u_k \sim N(0,1)$ でかつ $u_1, \ldots, u_{n-1}$ が独立であることを講義では紹介した (が、証明はしていない)。

これと同様に、一般の n に対しても、(6) となるような u_k を、$x_1, \ldots, x_n$ の 1 次式として構成し、それが独立でかつ $N(0,1)$ に従うことを示すことで (3) を示す。

3 独立な正規確率変数に関する条件

$n = 2, 3$ の場合と同様に、u_i を x_k の一次式として、

$$u_i = \frac{1}{\sigma}\sum_{j=1}^{n} a_{i,j} x_j \quad (1 \le i \le n-1) \tag{7}$$

とする。この u_i が $u_i \sim N(0,1)$ となるためには、[2] より

$$E[u_i] = \frac{1}{\sigma}\sum_{j=1}^{n} a_{i,j}\,\mu = \frac{\mu}{\sigma}\sum_{j=1}^{n} a_{i,j} = 0, \tag{8}$$

$$V[u_i] = \frac{1}{\sigma^2}\sum_{j=1}^{n} a_{i,j}^2\,\sigma^2 = \sum_{j=1}^{n} a_{i,j}^2 = 1 \tag{9}$$

が条件となる。ここで、行ベクトル $\overrightarrow{\alpha}_i, \overrightarrow{\beta}$ を

$$\overrightarrow{\alpha}_i = (a_{i,1}, \dots, a_{i,n}) \quad (1 \le i \le n-1), \quad \overrightarrow{\beta} = \frac{1}{\sqrt{n}}(1, \dots, 1) \tag{10}$$

とすると、条件 (8), (9) は

$$\overrightarrow{\alpha}_i \cdot \overrightarrow{\beta} = 0 \ \ (1 \le i \le n-1), \quad |\overrightarrow{\alpha}_i| = 1 \tag{11}$$

と書け、つまり $\overrightarrow{\alpha}_i$ はすべて $\overrightarrow{\beta}$ と垂直な単位ベクトルとなる。なお、$\overrightarrow{\beta}$ も単位ベクトルである。

また、$u_1, \dots, u_{n-1}$ が独立であることは、[3] により、$\overrightarrow{\alpha}_i$ $(1 \le i \le n-1)$ が互いに垂直であることと同値である。よって、これらをまとめると、$u_1, \dots, u_{n-1} \sim N(0,1)$ で、かつ独立であることは、

「$\overrightarrow{\alpha}_1, \dots, \overrightarrow{\alpha}_{n-1}, \overrightarrow{\beta}$ が互いに垂直な単位ベクトル」

と同値になる。そしてこれは、行列 A を

$$A = \begin{bmatrix} \overrightarrow{\alpha}_1 \\ \vdots \\ \overrightarrow{\alpha}_{n-1} \\ \overrightarrow{\beta} \end{bmatrix} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n-1,1} & \cdots & a_{n-1,n} \\ 1/\sqrt{n} & \cdots & 1/\sqrt{n} \end{bmatrix} \tag{12}$$

とすれば、A が直交行列、すなわち

$$A {}^tA = {}^tAA = E \tag{13}$$

となることと同値になる ($A {}^tA$ は A の行ベクトルの内積を成分とする行列)。

4 平方和に関する条件

次は、(6) の条件を $a_{i,j}$ に関する条件に書き直す。この式に (7) を代入すれば、

$$S = \sum_{k=1}^{n} (x_k - \bar{x})^2 = \sum_{i=1}^{n-1} \left(\sum_{j=1}^{n} a_{i,j} x_j \right)^2 \tag{14}$$

となり、両辺とも $x_1, \ldots, x_n$ の 2 次式なので、それが恒等式となるような係数の条件が求める条件となる。

まずは、(14) の左辺 S を計算する。展開すると、

$$S = \frac{1}{n^2}\sum_{k=1}^{n}(nx_k - x_1 - \cdots - x_n)^2 = \sum_{i=1}^{n} d_i x_i^2 + \sum_{1\leq i<j\leq n} p_{i,j}x_ix_j$$

の形になるが、x_i^2 の係数 d_i は、和の部分から出てくるのは、i 番目が $(n-1)^2$, それ以外の $(n-1)$ 個は 1 なので、

$$d_i = \frac{1}{n^2}\{(n-1)^2 + (n-1)\times 1\} = \frac{n^2-n}{n^2} = \frac{n-1}{n}$$

となる。x_ix_j の係数 $p_{i,j}$ は、和の部分から出てくるのは、i 番目と j 番目が $-2(n-1)$、それ以外の $(n-2)$ 個は 2 なので、

$$p_{i,j} = \frac{1}{n^2}\{-2(n-1)\times 2 + (n-2)\times 2\} = -\frac{2n}{n^2} = -\frac{2}{n}$$

となる。よって、

$$S = \frac{n-1}{n}\sum_{i=1}^{n} x_i^2 - \frac{2}{n}\sum_{1\leq i<j\leq n} x_ix_j = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 + 2\sum_{1\leq i<j\leq n} x_ix_j\right)$$

となる。これは、行列 N, 列ベクトル $\overrightarrow{x}$ を

$$N = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}, \quad \overrightarrow{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{15}$$

とすると、対称行列 $E - N/n$ に関する 2 次形式

$$S = {}^t\overrightarrow{x}\left(E - \frac{1}{n}N\right)\overrightarrow{x} \tag{16}$$

の形に書くことができる。

一方、(14) の右辺 ($= I$ とする) は、

$$\left(\sum_{j=1}^{n} a_{i,j}x_j\right)^2 = (\overrightarrow{\alpha}_i\overrightarrow{x})^2 = {}^t(\overrightarrow{\alpha}_i\overrightarrow{x})(\overrightarrow{\alpha}_i\overrightarrow{x}) = {}^t\overrightarrow{x}\left({}^t\overrightarrow{\alpha}_i\overrightarrow{\alpha}_i\right)\overrightarrow{x}$$

と書けるので、

$$Q = \sum_{i=1}^{n-1} {}^t\overrightarrow{\alpha}_i\overrightarrow{\alpha}_i = \left[\begin{array}{ccc} {}^t\overrightarrow{\alpha}_1 & \cdots & {}^t\overrightarrow{\alpha}_{n-1}\end{array}\right]\left[\begin{array}{c}\overrightarrow{\alpha}_1\\ \vdots \\ \overrightarrow{\alpha}_{n-1}\end{array}\right]$$

とすると、

$$I = {}^t\overrightarrow{x}Q\overrightarrow{x}$$

となる。よって、(14) は

$${}^t\overrightarrow{x}\left(E - \frac{1}{n}N\right)\overrightarrow{x} = {}^t\overrightarrow{x}Q\overrightarrow{x}$$

となり、これが $\overrightarrow{x}$ に関して恒等的に成り立つためには、$E - N/n = Q$ が条件となる。

一方、${}^t\overrightarrow{\beta}\overrightarrow{\beta} = N/n$ となるので、

$$Q + \frac{1}{n}N = \sum_{i=1}^{n-1} {}^t\overrightarrow{\alpha}_i\overrightarrow{\alpha}_i + {}^t\overrightarrow{\beta}\overrightarrow{\beta} = \left[\begin{array}{cccc} {}^t\overrightarrow{\alpha}_1 & \cdots & {}^t\overrightarrow{\alpha}_{n-1} & {}^t\overrightarrow{\beta}\end{array}\right]\left[\begin{array}{c}\overrightarrow{\alpha}_1\\ \vdots \\ \overrightarrow{\alpha}_{n-1}\\ \overrightarrow{\beta}\end{array}\right] = {}^tAA$$

となり、よって、$E - N/n = Q$ は $E = {}^tAA$ と同じことになり、そしてこれは (13) とも等しい。

すなわち、正規性と独立性の条件 (13) によって、(14) も自動的に得られることになり、結局 $a_{i,j}$ の満たすべき条件は A が直交行列になること、となる。

これを満たす $a_{i,j}$ はたくさんある。例えば、

$$\begin{cases} \overrightarrow{\alpha}_1 & = \dfrac{1}{\sqrt{2}}(1,-1,0,\ldots,0), \\ \overrightarrow{\alpha}_2 & = \dfrac{1}{\sqrt{6}}(1,1,-2,0,\ldots,0), \\ \overrightarrow{\alpha}_3 & = \dfrac{1}{\sqrt{12}}(1,1,1,-3,0,\ldots,0), \\ \ldots & \\ \overrightarrow{\alpha}_{n-1} & = \dfrac{1}{\sqrt{n(n-1)}}(1,1,\ldots,1,-(n-1)) \end{cases} \tag{17}$$

も一つの解となる。

$n=2,3$ で紹介した例 (4), (5) もほぼこの解と同じで、$n=3$ の方は x_2 と x_3 を入れかえて、u_2 を -1 倍すれば上の形になる。

以上により、一般の n (≥ 2) に対して (3) が示されたことになる。

参考文献

[1] 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`

[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「正規確率変数の一次式の独立性」(2022)
`http://takeno.iee.niit.ac.jp/~shige/math/lecture/graduate/data/normal2.pdf`