単相関係数の密度関数について　　2008 年 6 月 25 日

確率ベクトル $\begin{pmatrix} X_i \\ Y_i \end{pmatrix}$, $i = 1, 2, \ldots, n$, を 2 次元正規分布 $\mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right)$ からの sample とし、その標本相関係数 R を考える:

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(\sum_{i=1}^n (X_i - \bar{X})^2)^{1/2} (\sum_{i=1}^n (Y_i - \bar{Y})^2)^{1/2}}. \tag{1}$$

ただし、$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ である。

問 1. 文献 [3] p.95 にあるような 1 行目が $\left(\frac{1}{\sqrt{n}} \cdots \frac{1}{\sqrt{n}}\right)$ である直交行列 $\mathbf{\Gamma}$ により、

$$\mathbf{\Gamma}\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \sqrt{n}\bar{X} \\ \tilde{X}_1 \\ \vdots \\ \tilde{X}_{n-1} \end{pmatrix}, \qquad \mathbf{\Gamma}\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} \sqrt{n}\bar{Y} \\ \tilde{Y}_1 \\ \vdots \\ \tilde{Y}_{n-1} \end{pmatrix}$$

と定義すると、

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^{n-1} \tilde{X}_i^2, \qquad \sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^{n-1} \tilde{Y}_i^2, \qquad \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \sum_{i=1}^{n-1} \tilde{X}_i \tilde{Y}_i$$

となることを確かめよ。

問 2. $\begin{pmatrix} \tilde{X}_i \\ \tilde{Y}_i \end{pmatrix}$, $i = 1, 2, \ldots, n$, は独立で 2 次元正規分布 $\mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right)$ に従うことを確かめよ。

以下、$\tilde{X}_i, \tilde{Y}_i$ を単に X_i, Y_i と書くこととし、$N = n - 1$ とする。また、$\sigma^2 = \sigma_2^2(1 - \rho^2)$ とする。

問 3. $X_i = x_i$ で条件付けたとき Y_i は $\mathrm{N}\left(\frac{\sigma_2\rho}{\sigma_1}x_i, \sigma^2\right)$ に従うことを示せ。

問 4. $B := \dfrac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N X_i^2}$, $\dfrac{U}{\sigma^2} := \dfrac{1}{\sigma^2}\sum_{i=1}^N (Y_i - BX_i)^2$ は、$X_i = x_i$, $i = 1, \ldots, N$, の条件の下、独立で、それぞれ $N\left(\frac{\sigma_2\rho}{\sigma_1}, \frac{\sigma^2}{c^2}\right)$, χ^2_{N-1} に従うことを示せ。[1]ただし、$c^2 = \sum_{i=1}^N x_i^2$ である。

[ヒント] $e_i = Y_i - E[Y_i | X_i = x_i] = Y_i - \frac{\sigma_2\rho}{\sigma_1^2}x_i$ は $X_i = x_i$ の条件の下、$\mathrm{N}(0, \sigma^2)$ に従うこと、

$$\sum_{i=1}^N e_i^2 = \sum_{i=1}^N (Y_i - BX_i)^2 + \left(B - \frac{\sigma_2\rho}{\sigma_1}\right)^2 c^2$$

となること、及び、1 行目が $\left(\frac{x_1}{c} \cdots \frac{x_N}{c}\right)$ である直交行列を考え、[2] p.186 もしくは [3] p.95 の議論を用いよ。

問 5. 確率変数 T の $X = x$ での条件付け密度関数 $f_{T|X}(t|x)$ が x に依存しないとき、T と X は独立で T の密度関数 $f_T(t)$ は $f_{T|X}(t|x)$ に一致することを示せ。

以下、しばらく $\rho = 0$ と仮定する。

問 **4** において $T := \dfrac{\frac{c}{\sigma}B}{\sqrt{\frac{U/\sigma^2}{N-1}}}$ は、$X_i = x_i$, $i = 1, \ldots, N$, の条件の下、自由度 $N - 1$ の t 分布に従う。これは $X_i = x_i$, $i = 1, \ldots, N$, に依存していないので、問 **5** より条件付けをしなくても自由度 $N - 1$ の t 分布に従う。(以下、また (1) 式での記号を用いる。)

[1] χ_k^2 に従うとは、自由度 k のカイ自乗分布に従うの意味である。

問 6. $\dfrac{cB}{\sqrt{U}} = \dfrac{R}{\sqrt{1-R^2}}$ を示し、$\sqrt{n-2}\dfrac{R}{\sqrt{1-R^2}}$ が自由度 $n-2$ の t 分布に従うことを示せ。さらに、R の確率密度関数 $f_R(r)$ が $f_R(r) = \dfrac{\Gamma(\frac{N-1}{2})}{\Gamma(\frac{N-2}{2})\sqrt{\pi}}(1-r^2)^{(N-4)/2}$ $(-1 \le r \le 1)$ となることを示せ。[2]

再び $\rho = 0$ とは限らないとし、問 **2** の下に書いた記号を用いる。このとき、

$$W_{11} = \sum_{i=1}^N X_i^2, \qquad W_{22} = \sum_{i=1}^N Y_i^2, \qquad W_{12} = \sum_{i=1}^N X_i Y_i \tag{2}$$

とし、(W_{11}, W_{12}, W_{22}) の同時密度関数 $f_W(w_{11}, w_{12}, w_{22})$ を求めよう。(この分布を自由度 N の Wishart 分布という。[3])

問 7. $A = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{pmatrix}$ を $AA' = \mathbf{\Sigma} := \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}$ により定め[4]、$\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix} = A^{-1}\begin{pmatrix} X_i \\ Y_i \end{pmatrix}$, $i = 1, \ldots, N$, とおく。このとき、行列 A を求め、$\begin{pmatrix} Z_{1i} \\ Z_{2i} \end{pmatrix}$, $i = 1, \ldots, N$, は独立で 2 次元正規分布 $\mathrm{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\Big)$ に従うことを示せ。

次に、$S_{kl} = \sum_{i=1}^N Z_{ki}Z_{li}$, $k, l = 1, 2$, と定める。

問 **4** により、$Z_{1i} = z_{1i}$, $i = 1, \ldots, N$, の条件の下、$\dfrac{S_{12}}{S_{11}^{1/2}}$, $S_{22\cdot 1} := S_{22} - \dfrac{S_{12}^2}{S_{11}}$ は独立でそれぞれ $\mathrm{N}(0,1)$, χ^2_{N-1} に従う。ここで、これらは $Z_{1i} = z_{1i}$, $i = 1, \ldots, N$, によらないから、問 **5** と同様にこれらは Z_{1i}, $i = 1, \ldots, N$, と独立、即ち、S_{11} と独立となり、S_{11}, $S_{12}/S_{11}^{1/2}$, $S_{22\cdot 1}$ は独立でそれぞれ χ^2_N, $\mathrm{N}(0,1)$, χ^2_{N-1} に従う。

問 8. $T = \begin{pmatrix} T_{11} & 0 \\ T_{21} & T_{22} \end{pmatrix}$ を $TT' = \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix}$ と定めるとき、(T_{11}, T_{21}, T_{22}) の同時密度関数 $f_T(t_{11}, t_{21}, t_{22})$ が

$$f_T(t_{11}, t_{21}, t_{22}) = c_T^{-1} t_{11}^{N-1} t_{22}^{N-2} \exp\Big\{-\frac{1}{2}(t_{11}^2 + t_{21}^2 + t_{22}^2)\Big\}, \qquad t_{11}, t_{22} > 0, t_{21} \in \mathbf{R}$$

となることを示せ。ただし、$c_T = 2^{N-2}\sqrt{\pi}\Gamma(\frac{N}{2})\Gamma(\frac{N-1}{2})$ である。

問 9. $\tilde{T} = \begin{pmatrix} \tilde{T}_{11} & 0 \\ \tilde{T}_{21} & \tilde{T}_{22} \end{pmatrix}$ を $\tilde{T} = AT$ と定める。このとき、$(\tilde{T}_{11}, \tilde{T}_{21}, \tilde{T}_{22})$ の同時密度関数 $f_{\tilde{T}}(\tilde{t}_{11}, \tilde{t}_{21}, \tilde{t}_{22})$ が

$$f_{\tilde{T}}(\tilde{t}_{11}, \tilde{t}_{21}, \tilde{t}_{22}) = c_T^{-1}|\mathbf{\Sigma}|^{-N/2}\tilde{t}_{11}^{N-1}\tilde{t}_{22}^{N-2}\exp\Big\{-\frac{1}{2|\mathbf{\Sigma}|}(\sigma_2^2\tilde{t}_{11}^2 - 2\sigma_1\sigma_2\rho\tilde{t}_{11}\tilde{t}_{21} + \sigma_1^2(\tilde{t}_{21}^2 + \tilde{t}_{22}^2))\Big\}$$

$(\tilde{t}_{11}, \tilde{t}_{22} > 0, \tilde{t}_{21} \in \mathbf{R})$ となることを示せ。ただし、$|\mathbf{\Sigma}|$ は $\mathbf{\Sigma}$ の行列式である。

問 10. $\begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix} = \sum\limits_{i=1}^N \begin{pmatrix} X_i \\ Y_i \end{pmatrix}\begin{pmatrix} X_i & Y_i \end{pmatrix} = \tilde{T}\tilde{T}'$ を示し、

$$f_W(w_{11}, w_{12}, w_{22}) = c_W^{-1}|\mathbf{\Sigma}|^{-\frac{N}{2}}|w_{11}w_{22} - w_{12}^2|^{\frac{N-3}{2}}\exp\Big\{-\frac{1}{2|\mathbf{\Sigma}|}(\sigma_2^2 w_{11} - 2\sigma_1\sigma_2\rho w_{12} + \sigma_1^2 w_{22})\Big\}$$

$(w_{11}, w_{22} \ge 0, w_{11}w_{22} - w_{12}^2 \ge 0)$ となることを示せ。ただし、$c_W = 2^N\sqrt{\pi}\Gamma(\frac{N}{2})\Gamma(\frac{N-1}{2})$ である。[5]

参考文献

[1] Anderson, T.W.: An introduction to multivariate statistical analysis 2nd ed., John Wiley & Sons, New York, 1984.

[2] 浅野長一郎, 江島伸興, 李賢平: 基本統計学, 森北出版, 1993.

[3] 稲垣宣生: 数理統計学 改訂版, 裳華房, 2003.

[4] 国沢清典編: 確率統計演習 2 統計, 培風館, 1966.

[5] 竹村彰通: 多変量推測統計の基礎, 共立出版, 1991.

[2] この導出法は [1] Ch.4 による。あわせて [4] 第 7 章の関連する問題を解いてください。

[3] [1] によると、この密度関数は Fisher により導出され、Wishart はこれを p 次元正規分布に対する場合に拡張したとある。

[4] A' は A の転置行列を表す。$a_{11}, a_{22} > 0$ ととるものとする。この分解を Σ の三角分解という。

[5] この導出法は [5] による。これを用いて [4] **7-1** 要項 2 b. も示されるが、漸近理論を用いるので略す。詳しくは [5] 第 5 章 およびその引用文献を参照のこと。