⑨ 重相関係数の標本分布

1952.10.16.

阪大理学部数学教室　小川潤次郎

母平均ベクターが $m = \begin{bmatrix} m_1 \\ \vdots \\ m_k \end{bmatrix}$ で積率行列が Λ なる nonsingular な k次元正規母集団を考える。 その標本ベクターを縦にならべて

$$\mathbb{X} = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}$$

とし，$\mathbb{X}$ の列ベクターを夫々 $\mathfrak{x}_i$，$i = 1, 2, \cdots, k$ とする。$\mathfrak{x}_i$ の平均ベクターを $\bar{\mathfrak{x}}_i$ で表わし

$$\mathbb{X}^* = [\,\mathfrak{x}_1 - \bar{\mathfrak{x}}_1,\ \mathfrak{x}_2 - \bar{\mathfrak{x}}_2,\ \cdots,\ \mathfrak{x}_k - \bar{\mathfrak{x}}_k\,]$$

とする。

$$\mathbb{X}^{*\prime}\mathbb{X}^* = A = (a_{ij})$$

そうすると，x_1 と $x_2, x_3, \cdots, x_k$ の標本重相関係数は

$$r_{1(23\cdots k)} = \sqrt{1 - \frac{A}{a_{11}A_{11}}}$$

である。 これに対応する母集団重相関係数は

$$\rho_{1(23\cdots k)} = \sqrt{1 - \frac{\Lambda}{\lambda_{11}\Lambda_{11}}}$$

以下本稿では特に必要なき限り

$$r = r_{1(23\cdots k)}, \quad \rho = \rho_{1(23\cdots k)}$$

として記することにする。

$\rho = 0$ の special case にわ r の分布は比較的簡単に求められる。 その確率要素は

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k-1}{2}\right)}(r^2)^{\frac{k-3}{2}}(1-r^2)^{\frac{n-k-2}{2}}dr^2 \tag{1}$$

である。 例えば H. Cramér ; Math. Methods of Stat. 29.13. p. 414 を参照せよ。

$\rho \neq 0$ の一般の場合は R.A. Fisher[1], S.S. Wilks[2], 鍋谷清治[3], P.A.P. Moran[4] の諸氏によって夫々異った方法で derivation が与えられている。 結果わ

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(r^2)^{\frac{k-3}{2}}(1-r^2)^{\frac{n-k-2}{2}}\sum_{m=0}^{\infty}\frac{1}{m!}\frac{\Gamma\left(\frac{k-1}{2}\right)\Gamma^2\left(\frac{n-1}{2}+m\right)}{\Gamma^2\left(\frac{n-1}{2}\right)\Gamma\left(\frac{k-1}{2}+m\right)}(\rho^2 r^2)^m dr^2 \tag{2}$$

又わ超幾何級数を用いて

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\Gamma\left(\frac{k-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(r^2)^{\frac{k-3}{2}}(1-r^2)^{\frac{n-k-2}{2}}F\left(\frac{n-1}{2}, \frac{n-1}{2}, \frac{k-1}{2}; \rho^2 r^2\right)dr^2 \tag{3}$$

である。

著者はかつて，多変量解析に現れる諸統計量の標本分布を正規回帰論の方法によつて統一的に導くことを企てたことがあった。[5]

しかし，この方法は既に 1947年に G. Elfving[6] によって，より完全に成功していることを知つて，一応中止して了つたのであった。 しかし Elfving も著者も共に $\rho \neq 0$ の場合の重相関係数の標本分布は求めていなかつた。

この論文では，重相関係数の一般の場合の標本分布も亦正規回帰論的方法で導けることを示すのが目的である。

この方法が前記の著者達の方法に比較してより簡明であるかどうかは知らないが，方法の統一という観点からすれば，全然興味なしとはしないであろう。

以下で必要なことがらを補助定理として述べておく。

Lemma 1. x_i, $i=1,2,\cdots,n$ は夫々 $N(a_i,\sigma^2)$ に従って，互に独立に分布するとき統計量

$$\chi'^2 = \sum_{i=1}^{n} x_i^2/\sigma^2$$

の確率要素は

$$\frac{e^{-\lambda}}{2}\left(\frac{1}{2}\chi'^2\right)^{\frac{n-2}{2}} exp\left(-\frac{1}{2}\chi'^2\right)\sum_{m=0}^{\infty}\frac{\left(\frac{1}{2}\lambda\chi'^2\right)^m}{m!\Gamma\left(\frac{n}{2}+m\right)}d(\chi'^2) \tag{4}$$

で與えられる。 こゝに

$$\lambda = \frac{1}{2\sigma^2}\sum_{i=1}^{n} a_i^2 \tag{5}$$

χ'^2 は所謂“非心カイ自乗分布”である。

Lemma 2. χ'^2 は (4) で與えられる分布に従い，χ^2 は χ'^2 と独立な自由度 p のカイ自乗分布に従う確率変数とすれば

$$G = \frac{\chi'^2}{\chi^2} \tag{6}$$

の確率要素は

$$\sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\left(\frac{G}{1+G}\right)^{\frac{n+2m-2}{2}}\left(\frac{1}{1+G}\right)^{\frac{p+2}{2}}\frac{\Gamma\left(\frac{n+p}{2}+m\right)}{\Gamma\left(\frac{n}{2}+m\right)\Gamma\left(\frac{p}{2}\right)}dG \tag{7}$$

である：

Lemma 3.[7] 統計量

$$\lambda^* = \frac{A_{11}}{2A}\sum_{pq=2}^{k}\frac{A_{1p}A_{1q}}{A_{11}^2}a_{pq} \tag{8}$$

の確率要素は

$$\frac{1}{\left(\frac{\rho^2}{1-\rho^2}\right)^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{\lambda^*}{\frac{\rho^2}{1-\rho^2}}}(\lambda^*)^{\frac{n-1}{2}-1}d\lambda^* \tag{9}$$

である。

$x_2, \cdots\cdots, x_k$ を固定した x_1 の Conditional distribution で考えると

$$\eta_1 = x_1 - m_1 - \beta_{12}(x_2 - m_2) - \cdots\cdots - \beta_{1k}(x_k - m_k) \tag{10}$$

は $N(0, \sigma_{1.23\cdots k}^2)$ からの任意標本である。 パラメーターは最小自乗法で推定し，

$$\beta = \begin{bmatrix} \beta_{12} \\ \vdots \\ \beta_{1k} \end{bmatrix}$$

の推定値を

$$b = \begin{bmatrix} b_{12} \\ \vdots \\ b_{1k} \end{bmatrix}$$

とすると，b の決定方程式は

$$X_1^{*\prime} X_1^* b = X_1^{*\prime}(x_1 - \bar{x}_1)$$

$$b = A_{11}^{-1} X_1^{*\prime}(x_1 - \bar{x}_1) \tag{11}$$

そうすると勿論

$$q_1 = (x_1 - \bar{x} - X_1^* b)'(x_1 - \bar{x}_1 - X_1^* b) \tag{12}$$

と b とは互に独立である。

$$q_2 = (A_{11}^{\frac{1}{2}} b)'(A_{11}^{\frac{1}{2}} b) = (x_1 - \bar{x}_1)' X_1^* A_{11}^{-1} X_1^* (x_1 - \bar{x}_1) \tag{13}$$

これらは共に $\sigma_{1.23\cdots k}^2$ で割れば夫々自由度 $n-k$, $k-1$ のカイ自乗分布，非心カイ自乗分布であつて

$$q_1 = a_{11} - (x_1 - \bar{x})' X_1^* A_{11}^{-1} X_1^* (x_1 - \bar{x}_1)$$

であるから

$$G = \frac{q_2^{k-1}}{q_1} = \frac{(x_1-\bar{x}_1)' X_1^* A_{11}^{-1} X_1^* (x_1-\bar{x}_1)}{a_{11}-(x_1-\bar{x}_1)' X_1^* A_{11}^{-1} X_1^* (x_1-\bar{x}_1)}$$

$$= \frac{r^2}{1-r^2}$$

$$r^2 = \frac{G}{1+G}$$

だから（$x_2, \cdots, x_k$）を固定したときの r^2 の Conditional な確率要素は，（7）から

$$\sum_{m=0}^{\infty} e^{-\lambda^*} \frac{\lambda^{*m}}{m!} (r^2)^{\frac{k-3}{2}} (1-r^2)^{\frac{n-k-2}{2}} (r^2)^m \frac{\Gamma(\frac{n-1}{2}+m)}{\Gamma(\frac{k-1}{2}+m)\Gamma(\frac{n-k}{2})} d(r^2) \qquad (14)$$

ところで λ^* の分布は（9）で與えられるから求める r^2 の分布は

$$\frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n-k}{2})\Gamma(\frac{k-1}{2})} (1-\rho^2)^{\frac{n-1}{2}} (r^2)^{\frac{k-3}{2}} (1-r^2)^{\frac{n-k-2}{2}} \times$$

$$\sum_{m=0}^{\infty} \frac{1}{m!} \frac{\Gamma^2(\frac{n-1}{2}+m)\Gamma(\frac{k-1}{2})}{\Gamma^2(\frac{n-1}{2})\Gamma(\frac{k-1}{2}+m)} (\rho^2 r^2)^m d(r^2)$$

参　考　文　献

(1) R.A. Fisher : The general compling distribution of the multiple correlation coefficients. Proc. Roy. Soc. London. Vol. 121. (1928). pp. 654-673

(2) S. S. Wilks : On the sampling distribution of the multiple correlation coefficient. Ann. of Math. Stat. Vol. 3 (1932) pp. 196-203.

(3) 鍋谷清治： 重相関係数の標本分布について. 統数研講究録. Vol. 4. No. 9 (1948) pp 381-385

(4) P.A.P. Moran : The distribution of the multiple correlation coefficient. Proc. Cambridge Phil. Soc. Vol. 46 (1950). pp. 521-522.

(5) 小川潤次郎 ： 正規回帰論及その応用。 統数研講究録. Vol. 3. (1947) .

(6) G. Elfving : A simple method of deducing certain distributions concerned with multivariate sampling. Skandinavisk Aktuarietidskrift (1947).

(7) この証明には λ^* の c.f を計算して, P. Levy の反逆公式を用いればよい。