重相関係数の標本分布について

鍋谷　清治

$x_0, x_1, \cdots\cdots x_p$ は $p+1$ 次元の正規分布に従つて分布するものとし，簡単の為に各変数の平均値は 0 とする。

而して x_0 と $x_1, x_2, \cdots\cdots x_p$ との間の重相関係数を ρ，此の母集団より取つた大いさ n の標本に於ける x_0 と $x_1, x_2 \cdots\cdots x_p$ との間の重相関係数を R とするとき R^2 の分布は

(1)

$$\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}F\left(\frac{n-1}{2},\frac{n-1}{2};\frac{p}{2};\rho^2R^2\right)d(R^2)$$

但し

$$F(a,b;c;x)=1+\frac{ab}{c}x+\frac{a(a+1)b(b+1)}{c(c+1)}x^2+\cdots\cdots$$

によつて與へられる。此の分布は既に次の論文で導かれてゐる。

R.A. Fisher: The General Sampling Distribution of the multiple Correlation Coefficients — Proc. Roy. Soc. London. vol 121 (1928) p.p. 654−673.

S.S. Wilks: On the Sampling Distribution of the multiple Correlation Coefficient — Annals.

Math. Stat. vol. 3 (1932) p.p. 196-203.

併し其の証明は何れも相当厄介なので茲に亦一つの證明法を紹介する。此の分布は上記論文に於ける Fisher (A) - distribution であつて最近正準相関係数の標本分布の理論で重要な役割をなしてゐる。

先づ x_0 の或る常数倍 y_0 と $x_1, x_2 ----- x_p$ の一次独立な一次形式 $y_1, y_2 ---- y_p$ を適当にとれば，y_0 と $y_1 ---- y_p$ の間の分散行列は

$$\left(\begin{array}{cc|ccc} 1 & \rho & & & \\ \rho & 1 & & 0 & \\ \hline & & 1 & & 0 \\ & 0 & & \ddots & \\ & & 0 & & 1 \end{array}\right)$$

となる事が，正準相関論より知られてゐる。此の y_0 と $y_1, y_2 ----- y_p$ との間の標本重相関係数の分布を求めればよい訳である。

此の y に関する観測値を $y_{i\alpha}$ $(i=0,1,---p\,;\ \alpha=1,2---n)$ とし，これ等の値に対して適当な直交変換を施して平均値を消去する。即ち $(C_{\alpha\beta})$ を各組成分子が常数であるやうな n 次の直交行列で

$$C_{\alpha n} = \frac{1}{\sqrt{n}} \qquad (\alpha = 1, 2, ---- n)$$

とするとき

$$y'_{i\beta} = \sum_{\alpha=1}^{n} C_{\alpha\beta}\, y_{i\alpha} \qquad \left(\begin{array}{l} i = 0, 1, ---- p \\ \beta = 1, 2, ---- n-1 \end{array}\right)$$

と置く。

次に $y'_{i\beta}$ $(i = 0, 1, ---- p\,;\ \beta = 1, 2 ---- n-1)$ に対して，

$y'_{0\beta}$, $y'_{1\beta}$ ($\beta = 1, 2, \cdots n-1$) のみに depend する直交変換を施してそれを次のやうな形にする。

$$Z'_{i\gamma} = \sum_{\beta=1}^{n-1} d_{\beta\gamma}\, y'_{i\beta} \qquad \left(\begin{matrix} i = 0, 1, \cdots p \\ \gamma = 1, 2, \cdots n-1 \end{matrix} \right)$$

$$Z_{01} = a r \,,\quad Z_{02} = a\sqrt{1-\gamma^2},\ Z_{03} = \cdots = Z_{0\,n-1} = 0$$

$$Z_{11} = b,\quad Z_{12} = Z_{13} = \cdots Z_{1,\,n-1} = 0$$

茲に $d_{\beta\gamma}$ は $y'_{0\beta}$, $y'_{1\beta}$ ($\beta = 1, 2, \cdots n-1$) のみに depend する変数であり, r は y_0 と y_1 との間の標本相関係数である。

此の $Z_{i\gamma}$ ($i = 2, \cdots p;\ \gamma = 1, 2, \cdots n-1$) は a, b, r とは独立に，しかも互に独立に $N(0, 1)$ の分布をなす事は明らかである。扨

$$\begin{matrix} ar, & a\sqrt{1-\gamma^2}, & 0 & \cdots & 0 \\ b\ , & 0 & , 0 & \cdots & 0 \\ Z_{21}, & Z_{22} & , Z_{23} & \cdots & Z_{2,n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Z_{p1}, & Z_{p2}, & Z_{p3} & \cdots & Z_{p,n-1} \end{matrix}$$

なる行列に於て最後の k 行 k 列の作る小行列と其の転置行列を掛けた k 次の行列の行列式を $\Delta_{k,k}$ で表はせば, R^2 の定義と直交変換の性質から

$$1 - R^2 = \frac{\Delta_{p+1,\,n-1}}{a^2 \Delta_{p,\,n-1}}$$

然るに

$$\Delta_{p+1,\, n-1} = a^2 b^2 (1-r^2)\, \Delta_{p-1,\, n-3}$$

$$\Delta_{p,\, n-1} = b^2 \Delta_{p-1,\, n-2}$$

なる事は簡單な計算から容易にわかるから

$$1 - R^2 = (1-r^2)\ \frac{\Delta_{p-1,\, n-3}}{\Delta_{p-1,\, n-2}}$$

此の最後の行列式の比を W で表はせば，W の分布は次の如く與へられておる。

(2)

$$\frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)} W^{\frac{n-p-1}{2}-1}(1-W)^{\frac{p-1}{2}-1} dW \qquad (0 \leqq W \leqq 1)$$

(S. S. Wilks : Mathematical Statistics P. 237 参照)

而して r の分布は

(3)

$$\frac{(1-\rho^2)^{\frac{n-1}{2}}(1-r^2)^{\frac{n-4}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} \sum_{i=0}^{\infty} \frac{(2\rho r)^i}{i!} \Gamma^2\left(\frac{n-1+i}{2}\right) dr \qquad (-1 \leqq r \leqq 1)$$

で與へられる。r と W は独立に分布するからそれ等の同時分布は (2) (3) の積で與へられる。これを

$$W = \frac{1-R^2}{1-r^2}\ , \qquad dW\,dr = -\frac{1}{1-r^2}\, dr\, d(R^2)$$

なる関係に依って R^2 と r の同時分布に直せば

$$\frac{(1-\rho^2)^{\frac{n-1}{2}}(1-R^2)^{\frac{n-p-3}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}(R^2-r^2)^{\frac{p-3}{2}}\sum_{i=0}^{\infty}\frac{(2\rho r)^i}{i!}\Gamma^2\left(\frac{n-1+i}{2}\right)dr\,d(R^2)$$

これを r について $-R \leqq r \leqq R$ の範囲で積分すれば R^2 の分布が得られる。

$$r = Rt^{\frac{1}{2}}, \qquad dr = \frac{1}{2}Rt^{-\frac{1}{2}}$$

なる変換を用ひて其の積分を実行すれば

$$\frac{(1-\rho^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}\int_0^1(1-t)^{\frac{p-3}{2}}t^{-\frac{1}{2}}\sum_{i=0}^{\infty}\frac{(2\rho R)^{2i}t^i}{(2i)!}\Gamma\left(\frac{n-1}{2}+i\right)dt\,d(R^2)$$

$$=\frac{(1-\rho^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}}{\sqrt{\pi}\,\Gamma\left(\frac{p-1}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)}\sum_{i=0}^{\infty}\frac{2^{2i}}{(2i)!}\frac{\Gamma\left(i+\frac{1}{2}\right)\Gamma\left(\frac{p-1}{2}\right)}{\Gamma\left(\frac{p}{2}+i\right)}\Gamma^2\left(\frac{n-1}{2}+i\right)(\rho^2R^2)^i d(R^2)$$

$$=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}\sum_{i=0}^{\infty}\frac{1}{i!}\frac{\Gamma\left(\frac{p}{2}\right)}{\Gamma\left(\frac{p}{2}+i\right)}\frac{\Gamma^2\left(\frac{n-1}{2}+i\right)}{\Gamma^2\left(\frac{n-1}{2}\right)}(\rho^2R^2)^i d(R^2)$$

$$=\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{n-p-1}{2}\right)}(1-\rho^2)^{\frac{n-1}{2}}(R^2)^{\frac{p-2}{2}}(1-R^2)^{\frac{n-p-3}{2}}\Gamma\left(\frac{n-1}{2},\frac{n-1}{2};\frac{p}{2};\rho^2R^2\right)d(R^2)$$

これで上記の分布 (1) が得られた訳である。