(16) "r-DISTRIBUTIONS"に関する注意

樋口　伊佐夫

佐藤良一郎先生の "r-distributions & r-tests" に自由度についての careless mistakes があつたのを指摘申上げた際に，説明に用いた記法は複雑なもの機械的に見通しをよくするものと思う。　別段変つた面白いものでもないが小川潤次郎先生のおすゝめもあつたので本講究録に寄稿することにした。

§1. 標本相関係数の記述

x, y なる二つの scalar 量の組の size n の標本を (x_i, y_i) $(i=1,2,\cdots\cdots,n)$ とし $\mathfrak{x} \equiv (x_1, \cdots\cdots, x_n)$ $\mathfrak{y} \equiv (y_1, \cdots\cdots, y_n)$ とする。

x と y との標本相関係数を r とすると

$$r^2 = [L_0 \mathfrak{x}'(\mathfrak{x} L_0 \mathfrak{x}')^{-1} \mathfrak{x} L_0] : [L_0 \mathfrak{y}'(\mathfrak{y} L_0 \mathfrak{y}')^{-1} \mathfrak{y} L_0] \cdots\cdots \quad (1.1)$$

となる。　ここで L_0 は，$C_0 \equiv \underbrace{(1, 1, \cdots\cdots, 1)}_{n\text{個}}$

I を単位 n, n matrix とした時

$$L_0 \equiv I - C_0'(C_0 C_0')^{-1} C_0 \quad \cdots\cdots\cdots\cdots \quad (1.2)$$

で与えられる idenpotent symmetric matrix である。

(' は transposed をあらわす。)

: は二つの matrix の double sum の記号である。

次に y の代りに $\vec{y}$ なる k-vector 量を考え x と $\vec{y}$ との size n の標本を

$$\begin{pmatrix} x_1 & x_2 & \cdots\cdots & x_n \\ y_{11} & y_{12} & \cdots\cdots & y_{1n} \\ y_{21} & y_{22} & \cdots\cdots & y_{2n} \\ \cdots & \cdots & \cdots\cdots & \cdots \\ y_{k1} & y_{k2} & \cdots\cdots & y_{kn} \end{pmatrix}$$

とし，この行列の第一行を除いた k,n 行列をYとすると x と $\vec{y}$ との標本重相関係数 R は

$$R^2 = [L_0 \mathscr{Y}'(\mathscr{Y}_0 L_0 \mathscr{Y}')^{-1} \mathscr{Y} L_0] : [L_0 Y'(Y L_0 Y')^{-1} Y L_0] \cdots\cdots (1.3)$$

により與えられる。

更に x の代りに $\vec{x}$ なる l-vector 量を考える時，$(\vec{x}, \vec{y})$ の size n の標本を $S = \{s_{ij}\}$

$$s_{ij} = \begin{cases} x_{ij} & (i = 1, 2, \cdots, l;\ j = 1, 2, \cdots, n) \\ y_{i-l,j} & (i = l+1, l+2, \cdots, l+k;\ j = 1, 2, \cdots, n) \end{cases}$$

とする。S の最初の l 行のつくる l, n matrix をX，残りの k 行のつくる k, n matrix をYとすると $\vec{x}$ と $\vec{y}$ との増山氏の標本相関係数は

$$R^2 = \frac{1}{\min(k,l)} [L_0 X'(X L_0 X')^{-1} X L_0] : [L_0 Y'(Y L_0 Y')^{-1} Y L_0] \cdots\cdots (1.4)$$

で與えられる。

§2. 一次変換に対する性質

§1 の記法の一つの長所は一次変換に対する性質がみやすいということである。matrix の double sum に関して一般に

$$(ABC):D = B:(A'DC')$$

が成立つから，例えば n 次元直交変換 P により $XP = X^*$
$P'L_0P = L_0^*$ $YP = Y^*$ とすると

$$(1.4)\text{式} = \frac{1}{min(k,l)}\left[L_0^*X^{*'}(X^*L_0^*X^{*'})^{-1}X^*L_0\right]:\left[L_0^*Y^{*'}(Y^*L_0Y^{*'})^{-1}Y^*L_0^*\right]$$

となることは直ちにわかる。　これは L_0 の所がどんな matrix（但し rank は $max(k,l)$ より小さくない）であつても成立つ性質であることもわかる。

同様のことが (1.3), (1.1) についても成立つことも明かである。　又 non-singular な変換 A, B により $AX = X^*$
$BY = Y^*$ とする時も

$$(1.4\text{式}) = \frac{1}{min(k,l)}\left\{L_0X^{*'}(X^*L_0X^{*'})^{-1}X^*L_0\right]:\left[L_0Y^{*'}(X^*L_0Y^{*'})^{-1}Y^*L_0\right]$$

となることは一目瞭然である。　(1.1)(1.3) についても同様である。　これも L_0 の特性によるわけではない。

§3. r - distribution に関する諸定理.

基本定理

$x_1, x_2, \cdots\cdots, x_n$ を互いに独立に夫々 $N(0.\sigma^2)$ に従う変量とし．α を $\|\alpha\| \neq 0$ なる constant n − 横 vector,
$\varphi = (x_1, \cdots\cdots, x_n)$ とし　$r^2 = \varphi'(\varphi\varphi')^{-1}\varphi : \alpha'(\alpha\alpha')^{-1}\alpha$
とすると，r は 自由度 $n-1$ の r 分布（佐藤先生の定義による）に従う。

説明：　$\alpha P = (a_1^*, 0, 0, \cdots\cdots, 0)$ となるような constant orthogonal 変換 P により $\varphi P = \varphi^*$ とすると直ちにわかるように

$r^2 = \dfrac{x_1^{*2}}{\sum_{i=1}^{n} x_i^{*2}}$　　となる　ところで φ^* の components の

同時分布は $\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\mathfrak{X}\mathfrak{X}^*\right\}$ であるから

$q \equiv x_1^{*2}/\sigma^2 \quad p \equiv \left(\sum_{i=2}^{n} x_i^{*2}\right)/\sigma^2$ とすると p, q は夫々自由度 $n-1,\ 1$ の χ^2 分布に従つて互いに独立に分布するから $r^2 = q/(p+q)$ の分布は p, q の同時分布から出発すれば幾何学的考察をすることなしに容易に初等積分で求まる。 結果は

$$f(r)\,dr = const \times (1-r^2)^{\frac{n-1}{2}-1}\,dr \quad\text{------}\quad (3.1)$$

である。(n 次元空間での幾何学的考察はあまり直観的でもないので往々にして誤りを犯すことがある。) 又 $r^2/(1-r^2) = q/p$ であるから $\dfrac{\sqrt{n-1}\;r}{\sqrt{1-r^2}}$ は自由度 $n-1$ の t 分布に従うことも明かである。

次に $y_1, y_2, \text{-----}, y_n$ を $\mathfrak{X}$ とは独立に互いに独立に，$N(0, \sigma_2^2)$ に従う変量とする。 $\mathfrak{Y} \equiv (y_1, \text{------}, y_n)$ とし $r^2 = (\mathfrak{X}'(\mathfrak{X}\mathfrak{X}')^{-1}\mathfrak{X}) : (\mathfrak{Y}'(\mathfrak{Y}\mathfrak{Y}')^{-1}\mathfrak{Y})$ とすれば r の分布もやはり (3.1) である。 この事もよく知られた事であるが $\mathfrak{Y}$ だけに depend する直交変換をほどこしても $\mathfrak{X}$ の分布にかわりないことから明かである。

定　理　1.　基本定理の假定の下で L を idenpotent symmetric n, n matrix. $\operatorname{rank} L = n - s$ とする。

$$r^2 = \left[L\mathfrak{X}'(\mathfrak{X}L\mathfrak{X}')^{-1}\mathfrak{X}L\right] : \left[\mathfrak{A}'(\mathfrak{A}\mathfrak{A}')^{-1}\mathfrak{A}\right]$$

とすると r の分布は自由度 $n-s-1$ の r 分布である。

証　明　　L を orthogonal 変換で対角線形にするこの変換は勿論 $\mathfrak{X}, \mathfrak{A}$ に depend しない。 L^* の対角線上にならぶものは $n-s$ 個の 1 と s 個の 0 である。

この変換で y は y^*，Ω は Ω^* にうつるとすると，y^* の分布は y の分布と同じであり，γ^2 は L^* の対角線上の 1 に対応する y^* の components から基本定理に於ける γ^2 をつくつたものと同等である。 従つて基本定理に於て n が s だけ減ずるだけである。 以上

普通出てくる L は次のような形のものである。

$L = I - C'(CC')^{-1}C$ 　ここで I は単位 $n-n$ matrix をあらわし C は constants の $s.n$ matrix で rank $C = s$ である。

即ち $E(x_i) = \xi_i$，$\xi \equiv (\xi_1, \cdots, \xi_n)$ とし $\theta \equiv (\theta_1, \cdots, \theta_s)$ を未知の parameter の s vector とし既知の C を以て，$\xi = \theta C$ なる関係にあるものを取扱うことが多い。

例えば独立な二つの regression は，Z_{ij} を既知として

$$E(x_i) = \alpha_{11} Z_{1i} + \alpha_{12} Z_{2i} + \cdots + \alpha_{1k} Z_{ki} + \beta_1 \quad (i = 1, \cdots, m)$$

$$E(x_i) = \alpha_{21} Z_{1i} + \alpha_{22} Z_{2i} + \cdots + \alpha_{2l} Z_{li} + \beta_2 \quad (i = m+1, \cdots, n)$$

でこの時，$\theta = (\alpha_{11}, \alpha_{12}, \cdots, \alpha_{1k}, \beta_1, \alpha_{21}, \alpha_{22}, \cdots, \alpha_{2l}, \beta_2)$

$$C = \begin{pmatrix} Z_{11} & Z_{12} & \cdots & Z_{1n} & 0 & 0 & \cdots & 0 \\ Z_{21} & Z_{22} & \cdots & Z_{2n} & 0 & 0 & \cdots & 0 \\ \cdots & & & & & & & \cdots \\ Z_{k1} & Z_{k2} & \cdots & Z_{kn} & 0 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & Z_{1\,m+1} & Z_{1\,m+2} & \cdots & Z_{1n} \\ \cdots & & & & & & & \cdots \\ 0 & 0 & \cdots & 0 & Z_{l\,m+1} & Z_{l\,m+2} & \cdots & Z_{ln} \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

となる。 C の選び方に関して拙論（註４）に於ても二三の例を示しておいた。

さて　　$\xi(I-C'(CC')^{-1}C)=\xi-\xi C'(CC')^{-1}C$

$$=\theta C-\theta CC'(CC')^{-1}C=0$$

であるから，$(\mathfrak{X}-\xi)(I-C'(CC')^{-1}C)=\mathfrak{X}(I-C'(CC')^{-1}C)$

従つてLがこのようなものであれば定理1に於けるx_iの平均がすべて0であるという假定は不要である。　分散だけ等しければ平均はすべて異つていてもよい。　従つて，

定　理　2.　x_iを互いに独立に$N(\xi_i;\sigma^2)$に従つて分布するvariateとし，$\mathfrak{a}$を$\|\mathfrak{a}\|\neq 0$なるconstant vector，Iを單位$n\,n$ matrix Cをconstant $s\,n$ matrix rank $C=s$とする。　$L=I-C(CC')^{-1}C$

$$r^2=(L\mathfrak{X}'(\mathfrak{X}L\mathfrak{X}')^{-1}\mathfrak{X}L):\mathfrak{a}'(\mathfrak{a}\mathfrak{a}')^{-1}\mathfrak{a}$$

とするとrの分布は自由度$n-s-1$の分布である。

以上の諸定理により佐藤先生のAnnals vol II NO 2に於ける御論文の中の定理がすべて含まれることになる。　自由度がCのrankによつてきまるので種々なexampleに於て一々厄介な幾何学的考察を行うことがいらない。

ついでに$\mathfrak{y}=(y_1,\text{------},y_n)$とし$y_i$が互いに独立に且$\mathfrak{X}$とは独立に夫々$N(\eta_i,\sigma_2^2)$に従つて分布し$\eta=\theta^*C$　$\xi=\theta C$がみたされるとき，やはり$L=I-C'(CC')^{-1}C$として

$$r^2=(L\mathfrak{X}'(\mathfrak{X}L\mathfrak{X}')^{-1}\mathfrak{X}L):(L\mathfrak{y}'(\mathfrak{y}L\mathfrak{y}')^{-1}\mathfrak{y}L)\text{ ---- }(3.2)$$

とすればrは自由度$n-s-1$のr分布に従うことも明かである。

従つて，(1.1)に於ては$s=1$だから，よく知られた結果が出る。

尚　$\eta=\theta^*C$　$\xi=\theta C$のときは普通の相関係数(1.1)を用い

るより（3.2）を用いた方がよいことは今迄の議論からも裏づけられるが，$\eta=\theta^{*}C^{*}$　$\xi=\theta C$ という様な場合はどうであろうか　$L_1=I-C'(CC')^{-1}C$　$L_2=I-C^{*'}(C^{*}C^{*'})^{-1}C^{*}$ として，$(L_1\xi'(\xi L_1\xi')^{-1}\xi L_1):(L_2\eta'(\eta L_2\eta')^{-1}\eta L_2)$ とすればよいようであるが正規母集団の場合でも r 分布（或いは ρ のある場合の分布）を用いて推論する場合に今までの場合のように簡単にゆかない。　L_1 と L_2 とが同時に主軸変換出來る場合，即ち L_1 と L_2 とが可換のときは，

$$(L_1\xi'(\xi L_1L_2\xi')^{-1}\xi L_1):(L_2\eta'(\eta L_1L_2\eta')^{-1}\eta L_2)\cdots\cdots(3.3)$$

で以て定義すればどうであろうか，これは

$$(L_1L_2\xi'(\xi L_1L_2\xi')^{-1}\xi L_1L_2):(L_1L_2\eta'(\eta L_1L_2\eta')^{-1}\eta L_1L_2)$$

に等しいから今までと同様の推論が可能である。　（3.3）は一つの私案である。

§4.　重相関係数に対応するもの．

單相関係数の場合にはむしろ複雑と思われる記法を用いたのは重相関係数の場合にも平行して議論するためでもあつた。

先づ§3の基本定理に対応するものは，

定理 3　$\xi=(x_1,\cdots\cdots x_n)$

を各 Component が互いに独立に $N(0,\sigma^2)$ に従つて分布する variate とし，A を rank k の constant $k\ n$ matrix（$k<n$）とすると

$$R^2=(\xi'(\xi\xi')^{-1}\xi):(A'(AA')^{-1}A)\quad の分布は$$

$$const\times(R^2)^{\frac{k}{2}-1}(1-R^2)^{\frac{n-k}{2}-1}d(R^2)\ \cdots\cdots\ (4.1)$$

である。

証　明　k次の直交行列Qとn次の直交行列Pとを適当にとることにより

$$QAP=\left(\begin{array}{cccc|c} a_1 & & & 0 & \\ & a_2 & & & 0 \\ 0 & & \ddots & & \\ & & & a_k & \end{array}\right)$$

なる形に出来る。　ところで　$QAP=A^*$, $\mathfrak{X}P=\mathfrak{X}^*$ とすると§2により

$$R^2=\left(\mathfrak{X}^{*\prime}(\mathfrak{X}^*\mathfrak{X}^{*\prime})^{-1}\mathfrak{X}^*\right):\left(A^{*\prime}(A^*A^{*\prime})^{-1}A^*\right)$$

又 $A^{*\prime}(A^*A^{*\prime})^{-1}A^*=\begin{pmatrix} \underbrace{1 \;\; 1}_{k\text{個}} & 0 & 0 \\ 0 & & 0 \end{pmatrix}$

だから

$$R^2=\left(\sum_{i=1}^{k}x_i^{*2}\right)\Big/\left(\sum_{i=1}^{n}x_i^{*2}\right)$$

又Pは$\mathfrak{X}$にdependしない直交変換だから$\mathfrak{X}^*$の分布も$\mathfrak{X}$の分布も変らない。　従つて

$$\frac{1}{\sigma^2}\sum_{i=1}^{k}x_i^{*2}\equiv q \qquad \frac{1}{\sigma^2}\sum_{i=k+1}^{n}x_i^{*2}\equiv p$$

とすると　p と q とは互いに独立に夫々自由度 $n-k$ 及びk の χ^2分布に従つて分布する。　故に

$R^2=\dfrac{q}{p+q}$　の分布は p, qの同時分布から出発すれば容易に求まる。　以上

又 $\dfrac{k}{n-k}\left(\dfrac{1}{R^2}-1\right)=\dfrac{p/(n-k)}{q/k}$　　だから

$F_{k,n-k}\equiv\left(\dfrac{n-k}{k}\right)\left(\dfrac{R^2}{1-R^2}\right)$　とすると $F_{k,n-k}$ は自由度 k, $n-k$ のF分布に従うことが直ちにわかる。

定 理 4. Xを平均0のk変数正規母集団からのsize nの標本matrix（kn-matrix）とし，αを$\|\alpha\|\neq 0$なるconstant vectorとするとき

$$R^2 = (X'(XX')^{-1}X):(\alpha'(\alpha\alpha')^{-1}\alpha)$$

の分布もやはり（4.1）である。

証 明． 直接證明することも出来るが，$\vec{x}$とyとが独立の時の平均0の正規分布からの

$$R^2 = (X'(XX')^{-1}X):(y'(yy')^{-1}y)$$

の分布を仲介にすれば容易にわかる。 このR^2の分布もやはり（4.1）である。 何故ならXとyとの同時分布は

$$f(X, y) = \frac{(\det \Phi)^{\frac{n}{2}}}{\sqrt{2\pi}^{kn}} \exp\left\{-\frac{1}{2}\Phi : XX'\right\}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} yy'\right\}$$

Xだけにdependする直交変換をほどこし$y \to y^*$となったとしてもy^*の分布はやはり

$$f(y^*) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2} y^* y^{*\prime}\right\}$$

R^2は直交変換に対して不変だから定理3の時の如くXだけにdependする適当な直交変換により

$$R^2 = \sum_{i=1}^{k} y_i^{*2} \Big/ \left(\sum_{i=1}^{n} y_i^{*2}\right)$$

となり，この分布はy^*の分布にだけ依存することから明かにR^2の分布も（4.1）となる。

さて $R_1^2 = (X'(XX')^{-1}X):(\alpha'(\alpha\alpha')^{-1}\alpha)$ と
$R_2^2 = (X'(XX')^{-1}X):(y'(yy')^{-1}y)$ との二つ考えてみよう。

$\mathfrak{a} P_1 = (a_1^*, 0, \cdots\cdots, 0)$ ならしめる直交変換 P_1 で $X P_1 = X_1^*$　$\mathfrak{y} P_2 = (y_1^*, 0, 0, \cdots\cdots, 0)$ ならしめる直交変換で $X P_2 = X_2^*$ となるとする。

$$R_1^2 = \{ (X_1^{*\prime} (X_1^* X_1^{*\prime})^{-1} X_1^*) \text{ の } 1,1 \text{ 要素} \}$$

$$R_2^2 = \{ (X_2^{*\prime} (X_2^* X_2^{*\prime})^{-1} X_2^*) \text{ の } 1,1 \text{ 要素} \}$$

となる，ところで X_1^*, X_2^* の分布は

$$f(X_1^*) = \frac{(\det \Phi_1)^{\frac{n}{2}}}{\sqrt{2\pi}^{\,kn}} \exp\left\{-\frac{1}{2} \Phi_1 : X_1^* X_1^{*\prime}\right\}$$

$$f(X_2^*) = \frac{(\det \Phi_2)^{\frac{n}{2}}}{\sqrt{2\pi}^{\,kn}} \exp\left\{-\frac{1}{2} \Phi_2 : X_2^* X_2^{*\prime}\right\}$$

となることは明かで且つ R_2^2 の分布は Φ_2 に無関係に常に (4.1) であるから R_1^2 の分布も (4.1) であることがわかる。　以上

定理 5. 定理3，定理4の假定の下で L を $rank\ n-s$ の idenpotent symmetric $n\ n$ matrix とする時

$$R^2 = (L X' (X L X')^{-1} X) : (\mathfrak{a}' (\mathfrak{a} \mathfrak{a}')^{-1} \mathfrak{a})$$

$$R^2 = (L \mathfrak{Y}' (\mathfrak{Y} L \mathfrak{Y}')^{-1} \mathfrak{Y} L) : (A' (A A')^{-1} A)$$

の分布は何れも

$$\text{constant} \times (R^2)^{\frac{k}{2}-1} (1-R^2)^{\frac{n-s-k}{2}-1} d(R^2)$$

である。証明は述べる迄もない。

更に $\vec{\mathfrak{x}}$, $\mathfrak{x}$ の平均が 0 でない時

$$\mathbf{L} = I - C' (C C')^{-1} C$$

により §3 と同様の議論が出来ることも言うまでもない。

これ等の応用については佐藤先生が *Annals* の次の号にお書きになるかも知れないし徒らに紙数を費すことになるので割愛するが，これ等は定理３の証明の後に書いた注意からも結局 F-test と同等であることがわかる。

尚二つの vector 量の相関係数に相当するものについては簡単にゆかない。

参　考　書

1) R. Satō： r-distributions and r-tests
Annals of the Institute of Statistical Math.
Vol II NO. 2.

2) 鍋谷清治： 重相関係数の標本分布について
講究録第４巻，第９号

3) 伏見康治： 確率論及統計論　河出書房
§80（ヴェクトル量相関に関する増山氏の説）

4) 樋口伊佐夫： 正規 Parameters の比の推定について
講究録第七巻，第六号