(1555)

変量分析法に於て取扱はれる統計量の独立性と自由度の判定法に就て

昭和二十一年五月十二日　所員　坂元平八

I. 独立性と自由度の判定條件

筆者は前に「統計量の独立性に就て」(註1)と題する論文で一次形式並びに二次形式統計量の独立性と自由度に関する問題を取扱つた。此の論文で取扱はれた二次形式統計量 $Q=(A\mathfrak{X},\mathfrak{X})$ は非常に一般的で、独立性を取扱ふ場合に関してはAには対称行列といふ以外に何等制限はなかつた。唯Qが χ^2 分布に従ふための特別な條件としては独立性とは無関係な $AVA=A$ なる制限を加へたのである。

然し吾々が應用上当面する問題にはかくまで広い條件は必要とせず、且 $AVA=A$ でなければQは χ^2 分布をなさず、従つてQの自由度を考へることも実際問題としては余り意味がないので、吾々は以下変量分析法に於てよく取扱はれる統計量の独立性と自由度の簡単な判定法の研究をしようと思ふ。この特別な場合の判定法としてはCochran(註2)の有名な定理があるがこの定理は形としては非常にキレイであるにも拘らず行列の階数を決めるのに厳密でない点であるため実際應用上には不便な様に思へる。例へばWilksの著書(註3)中の随所に見られるCochranの定理の應用例の階数決定の処を参照すれば容易にこの点が首けることゝ思ふ。

依て筆者は今行列の使用を避けた簡単な判定條件を与へる定理を述べようと思ふ。

〔定理I〕 $x_1, x_2, \cdots, x_n$ を母集團分布法則が

$$f(x)\,dx=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2\sigma^2}}$$

で与へられるn個の独立に観測された値とし $Q_i^{(1)}, Q_i^{(2)}, \cdots, Q_i^{(m)}$ $(i=1,2,\cdots,n)$ を $x_1, x_2, \cdots, x_n$ より作れるmn個の一次形式とする。然るときかゝる $Q_i^{(j)}$ $(i=1,2,\cdots,n;\ j=1,$ $j=1,2,\cdots,m)$ より作れるm個の統計量

$$S_j=\sum_i^n\{Q_i^{(j)}\}^2,\quad(j=1,2,\cdots,m)$$

が相互に独立なるための必要充分なる條件は

$$\sum_{i=1}^{n} \theta_i^{(K)} \theta_i^{(\ell)} \equiv 0 \quad (K \neq \ell ; K=1,2,\cdots,m ; \ell=1,2,\cdots m).$$

なることである。特にこの場合

$$S_1 + S_2 + \cdots + S_m = \sum x_i^2$$

或は

$$S_1 + S_2 + \cdots\cdots + S_m = \sum (x_i - \bar{x})^2 \quad (\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i).$$

なる関係が与へられれば

$\frac{S_1}{\sigma^2}, \frac{S_2}{\sigma^2}, \cdots, \frac{S_m}{\sigma^2}$ はすべて χ^2 分布をなし且その自由度は

$$E\left(\frac{S_K}{\sigma^2}\right) = \gamma_K, \quad (K=1,2,\cdots,n)$$

で与へられる(註4)

(証明) 筆者の前論文(註5)の定理I,定理II,及び定理IIIを用ゐる。

$\theta_i^{(j)} = \sum_{K=1}^{n} a_{iK}^{(j)} x_K$ と置けば

$(\theta_1^{(j)}, \theta_2^{(j)}, \cdots, \theta_n^{(j)}) = A_j' \mathfrak{x}$ であらはされる。

但しここに $\mathfrak{x} = (x_1, x_2, \cdots, x_n)$ で

$$A_j = \begin{pmatrix} a_{11}^{(j)} & a_{12}^{(j)} & \cdots & a_{1n}^{(j)} \\ a_{21}^{(j)} & a_{22}^{(j)} & \cdots & a_{2n}^{(j)} \\ \vdots & \vdots & & \vdots \\ a_{n1}^{(j)} & a_{n2}^{(j)} & \cdots & a_{nn}^{(j)} \end{pmatrix}$$

とし A_j' は A_j の transposed matrix とする。

然るとき

$$S_j = \sum_{i=1}^{n} \{\theta_i^{(j)}\}^2 = (A_j'\mathfrak{x}, A_j'\mathfrak{x}) = (A_j A_j' \mathfrak{x}, \mathfrak{x}) = (B_j \mathfrak{x}, \mathfrak{x})$$

である。但しここに

$B_j = A_j A_j'$ とする。

B_j が対称行列なることは明らかである。亦

$\sum_{i=1}^{n} \theta_i^{(K)} \theta_i^{(l)} \equiv 0$ は $A_l A_K' = 0$ $(K \neq l;\ K=1,2,\cdots,m;\ l=1,2,\cdots,m)$

と等値である。併して $B_l B_K = A_l A_l' A_K A_K'$ であるから

$B_l B_K = 0$ $(K \neq l;\ K=1,\cdots,m;\ l=1,\cdots,m)$ と

$A_l A_K' = 0$ $(K \neq l;\ K=1,\cdots,m;\ l=1,\cdots,m)$ とは

等値である。(何となれば $A_l A_K' = 0 \longrightarrow B_l B_K = 0$ は明らか。逆に $B_l B_K = 0$ ならば $A_l A_l' A_K A_K' = 0 \longrightarrow A_K' A_l A_l' A_K A_K' A_l = 0 \longrightarrow A_K' A_l A_l' A_K A_K' A_l A_l' A_K = 0 \longrightarrow A_K' A_l A_l' A_K = 0 \longrightarrow A_K' A_l = 0$ が言へる。依て $A_l A_K' = 0$) 故に

$\sum_{i=1}^{n} \theta_i^{(K)} \theta_i^{(l)} \equiv 0$ と $B_l B_K = 0$ $(K \neq l;\ K=1,2,\cdots,m,\ l=1,2,\cdots,m)$

は等値である。従って前論文の定理 I に依り $V = \sigma^2 E$ なることを考慮すれば $S_1, S_2, \ldots, S_m$ が相互に独立である為の必要充分條件は

$\sum_{i=1}^{n} \theta_i^{(K)} \theta_i^{(l)} \equiv 0$ $(K \neq l;\ K=1,\ldots,m;\ l=1,\cdots,m)$

なる事が言へる。次に特に

$S_1 + S_2 + \cdots\cdots + S_n = \sum_{i=1}^{n} x_i^2$ ならば

$B_1 + B_2 + \cdots\cdots + B_m = E$ であるから

この両辺に $B_K\ (K=1,2,\cdots,m)$ を乗じて

$B_1 B_K + B_2 B_K + \cdots + B_K^2 + \cdots + B_m B_K = B_K$

$B_K B_l = 0$ $(K \neq l,\ K=1,2,\cdots,m;\ l=1,2,\cdots,m)$ であるから

$B_K^2 = B_K$ である。従って前論文定理 II に依り

$\frac{S_K}{\sigma^2}$ $(K=1,\cdots,m)$ は χ^2 分布をなすことが言へる。

亦 $S_1 + S_2 + \cdots + S_m = \sum_{i=1}^{n} (x_i - \bar{x})^2$ とすれば

$\sum (x_i - \bar{x})^2 = S_1 + S_2 + \cdots + S_m$ と $\sum \bar{x}^2 = S_{m+1}$ とは

相互に独立である。何となれば

$\sum_{i=1}^{n} (x_i - \bar{x}) \bar{x} \equiv 0$ となり、本定理の独立條件を満足するからである。

従って $S_1+S_2+\cdots+S_m=\sum_{i=1}^{n}x_i^2-\sum\bar{x}^2$ より

$S_1+S_2+\cdots+S_m+S_{m+1}=\sum_{i=1}^{n}x_i^2$ である。

今 $S_m=(B_{m+1}\mathfrak{x},\mathfrak{x})$ であらはせば

$B_1+B_2+\cdots+B_m+B_{m+1}=E$ で

$(B_1+B_2+\cdots+B_m)B_{m+1}=0$ が上で云へたから

此の両辺に B_K を掛けて

$$B_K(B_1+B_2+\cdots+B_m)B_{m+1}=0$$

$$(B_KB_1+B_KB_2+\cdots+B_KB_m)B_{m+1}=0$$

$$B_KB_l=0\quad(K\neq l;\ K=1,\cdots,m;\ l=1,\cdots,m)$$

だから $B_K^2B_{m+1}=0$ 従って $B_KB_K'B_{m+1}=0$

故に $B_{m+1}'B_KB_K'B_{m+1}=0$ これから $B_K'B_{m+1}=B_KB_{m+1}=0$

が言へる。依て $S_1,S_2,\cdots,S_{m+1}$ は相互に独立である

依て $B_K^2=B_K,\ (K=1,2,\cdots,m+1)$ は前と同様にして得られる。

従って $\frac{S_K}{\sigma^2},\ (K=1,2,\cdots,m+1)$ は亦、χ^2分布をなすことが言へる。

次に B_K の階数が r_K ならば適当な直交行列 T を選んで

$$D=T'B_KT=\begin{pmatrix}1 & & & & & 0\\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ 0 & & & & & 0\end{pmatrix}$$

ならしめ得る。依て

$E\left(\frac{S_K}{\sigma^2}\right)=E\left\{\frac{(B_K\mathfrak{x},\mathfrak{x})}{\sigma^2}\right\}$ で $\mathfrak{x}=T\mathfrak{y}$ $\mathfrak{y}=(y_1,y_2,\cdots,y_n)$

と置けば

$$E\left(\frac{S_K}{\sigma^2}\right)=E\left\{\frac{(B_KT\mathfrak{y},T\mathfrak{y})}{\sigma^2}\right\}=E\left\{\frac{(T'B_KT\mathfrak{y},\mathfrak{y})}{\sigma^2}\right\}=E\left\{\frac{(D\mathfrak{y},\mathfrak{y})}{\sigma^2}\right\}$$

$$=E\left(\frac{y_1^2+\cdots\cdots+y_{r_K}^2}{\sigma^2}\right)=$$

而して $\mathfrak{x}=T\mathfrak{y}$ に依り $E(y_i^2)=E(x^2)=\sigma^2\ (i=1,2,\cdots,r_K)$ なる

ことは容易に分るから

$$E\left(\frac{S_K}{\sigma^2}\right)=E\left(\frac{y_1^2+\cdots+y_{r_K}^2}{\sigma^2}\right)=\sum_{i=1}^{r_K}\frac{E(y_i^2)}{\sigma^2}=\frac{r_K\sigma^2}{\sigma^2}=r_K$$

なることが言へ 従って S_K の自由度は $E\left(\frac{S_K}{\sigma^2}\right)=\gamma_K$ で与へられることが言へた。

〔定理Ⅱ〕〔定理Ⅰ〕の $x_1, x_2, \cdots, x_n$ より作れる $(m+1)$ 個の統計量

$$S_1=(A_1\mathfrak{x},\mathfrak{x}),\ S_2=(A_2\mathfrak{x},\mathfrak{x}),\cdots,S_m=(A_m\mathfrak{x},\mathfrak{x}),$$

$$S_{m+1}=(B\mathfrak{x},\mathfrak{x}),$$ ($A_1, A_2, \cdots, A_m$ 及び B はすべて対称行列であるとする) が

$$S_1+S_2+\cdots+S_{m+1}=\sum_{i=1}^{n}x_i^2$$

で且 $\frac{S_1}{\sigma^2}, \frac{S_2}{\sigma^2}, \cdots, \frac{S_m}{\sigma^2}$ が相互に独立で夫々自由度 $\gamma_1, \gamma_2, \cdots \cdots \gamma_m$ の χ^2 分布をなすことが分つてゐるとする。然るとき $S_1, S_2, \cdots, S_m, S_{m+1}$ は相互に独立で且 $\frac{S_{m+1}}{\sigma^2}$ は自由度 $n-(\gamma_1+\gamma_2+\cdots+\gamma_m)=\gamma_{m+1}$ の χ^2 分布をなす。

〔定理Ⅲ〕(註6) 〔定理Ⅰ〕の $x_1, x_2, \cdots, x_n$ より作れる二つの統計量 $S_1=(A\mathfrak{x},\mathfrak{x})$ $S_2=(B\mathfrak{x},\mathfrak{x})$ があり $S_1+S_2=\sum_{i=1}^{n}x_i^2$ で且 $\frac{S_1}{\sigma^2}$ が自由度が γ の χ^2 分布をなしてゐることが分つてゐるとする。然るとき $\frac{S_2}{\sigma^2}$ は $\frac{S_1}{\sigma^2}$ と独立で且自由度 $n-\gamma$ の χ^2 分布をなす。

〔定理Ⅱの証明〕

$$S_1+S_2+\cdots+S_m+S_{m+1}=\sum x_i^2$$ であるから

$$A_1+A_2+\cdots+A_m+B=E \text{ ----- ①}$$ である

$\frac{S_1}{\sigma^2}, \frac{S_2}{\sigma^2}, \text{----}, \frac{S_m}{\sigma^2}$ は相互に独立で且 χ^2 分布をなすから $V=\sigma^2E$ なる事を考慮すれば筆者の前論文定理Ⅰ及び定理Ⅱに依り

$$\left.\begin{array}{ll} A_iA_j=0 & (i\neq j) \\ A_i^2=A_i \text{ ---------- ②} & \end{array}\right\} \text{------ ②}$$

であるから①式に依り

$$A_i(A_1+A_2+\cdots+A_m)+A_iB=A_i$$

従って②より $A_i^2+A_iB=A_i$ $\therefore A_iB=0$ $(i=1,2,\cdots,m-1)$

依て $S_1, S_2, \cdots, S_{m+1}$ は相互に独立である。
亦①式より

$$(A_1 + A_2 + \cdots + A_m)B + B^2 = B$$

従って③より　　$B^2 = B$

で $\frac{S_{m+1}}{\sigma^2}$ が χ^2 分布をなすことが容易に判る。且此が自由度 Σ_{m+1} の χ^2 分布をなすことは筆者の前論文定理Ⅲに依り極めて明かである。〔定理Ⅲの証明〕は〔定理Ⅰ〕の証明に依り明かである。

Ⅲ. 変量分析法への應用例.

〔定理Ⅲ〕の應用例は Cochran の論文にあるから本論文では省略する。〔定理Ⅱ〕の應用例は〔定理Ⅰ〕より範囲が狭い。尚本論文の定理は変量分析法以外に於ても應用は極めて有効なことを注意して置く。
此処では定理Ⅰの應用例を述べて置く。

変量分析法に於ては $N = ab$ 個の独立に観測された正規分布変量の値は或る適当な計畫に従って次の通り a 行 b 行に配列されるのが普通である。

$$\begin{array}{cccc}
x_{11}, & x_{12}, & \cdots\cdots & , x_{1b} \\
x_{21}, & x_{22}, & \cdots\cdots & , x_{2b} \\
\vdots & & & \vdots \\
x_{a1}, & x_{a2}, & \cdots\cdots & , x_{ab}
\end{array}$$

今 $\bar{x}_{j\circ}$, $\bar{x}_{\circ k}$, 及び $\bar{x}$ を夫々 j 行の平均値, k 列の平均値, 及び全観測値の平均値をあらはすものとすれば

$$\sum_1^a \sum_1^b (x_{jk} - \bar{x})^2 = \sum_1^a \sum_1^b (x_{jk} - \bar{x}_{j\circ} - \bar{x}_{\circ k} + \bar{x} + \bar{x}_{j\circ} - \bar{x} + \bar{x}_{\circ k} - \bar{x})^2$$

$$= b\sum_1^a (\bar{x}_{j\circ} - \bar{x})^2 + a\sum_1^b (\bar{x}_{\circ k} - \bar{x})^2 + \sum_1^a \sum_1^b (x_{jk} - \bar{x}_{j\circ} - \bar{x}_{\circ k} + \bar{x})^2$$

$$= S_1 + S_2 + S_3$$

が x の $N = ab$ 個の値に無関係に成立する。
これは $\theta_{jk}^{(1)} = \bar{x}_{j\circ} - \bar{x}$, $\theta_{jk}^{(2)} = \bar{x}_{\circ k} - \bar{x}$, 及び $\theta_{jk}^{(3)} = x_{jk} - \bar{x}_{j\circ} - \bar{x}_{\circ k} + \bar{x}$
と置くとき

$$x_{jk}-\bar{x}=\theta_{jk}^{(1)}+\theta_{jk}^{(2)}+\theta_{jk}^{(3)} \qquad (j=1,2,\cdots,a\ ;\ k=1,2,\cdots,b)$$

で

$$\sum_1^a\sum_1^b(x_{jk}-\bar{x})^2=\sum_1^a\sum_1^b(\theta_{jk}^{(1)}+\theta_{jk}^{(2)}+\theta_{jk}^{(3)})^2$$
$$=\sum_1^a\sum_1^b(\theta_{jk}^{(1)})^2+\sum_1^a\sum_1^b(\theta_{jk}^{(2)})^2+\sum_1^a\sum_1^b(\theta_{jk}^{(3)})^2$$

なる関係式が x の $N=ab$ 個の値に無関係に成立するのは

$$\sum_1^a\sum_1^b\theta_{jk}^{(1)}\theta_{jk}^{(2)}\equiv 0,\quad \sum_1^a\sum_1^b\theta_{jk}^{(2)}\theta_{jk}^{(3)}\equiv 0 \quad 及び$$

$$\sum_1^a\sum_1^b\theta_{jk}^{(3)}\theta_{jk}^{(1)}\equiv 0$$ が成立するからである。

従つて定理Iより

$$S_1=\sum_1^a\sum_1^b(\theta_{jk}^{(1)})^2,$$
$$S_2=\sum_1^a\sum_1^b(\theta_{jk}^{(2)})^2,$$
$$S_3=\sum_1^a\sum_1^b(\theta_{jk}^{(3)})^2$$

が相互に独立なることが云へ、且つ $S_1+S_2+S_3=\sum_1^a\sum_1^b(x_{jk}-\bar{x})^2$ であるから $\frac{S_1}{\sigma^2}$, $\frac{S_2}{\sigma^2}$ 及び $\frac{S_3}{\sigma^2}$ は皆 χ^2 分布をなす。且つ

$$E\left(\frac{S_1}{\sigma^2}\right)=a-1,$$
$$E\left(\frac{S_2}{\sigma^2}\right)=b-1,$$
$$E\left(\frac{S_3}{\sigma^2}\right)=(a-1)(b-1)$$

なることが容易に計算出来る。依て $\frac{S_1}{\sigma^2}$, $\frac{S_2}{\sigma^2}$ 及び $\frac{S_3}{\sigma^2}$ は相互に独立で且つ夫々自由度 $(a-1)$, $(b-1)$ 及び $(a-1)(b-1)$ の χ^2 分布をなすことが言へる。

或は $E\left(\frac{S_3}{\sigma^2}\right)$ を計算しなくとも、$\frac{S_1}{\sigma^2}$, $\frac{S_2}{\sigma^2}$ が夫々自由度 $(a-1)$, $(b-1)$ の χ^2 分布をなすことより定理IIに依り

$$\sum_1^a\sum_1^b(x_{jk}-\bar{x})^2$$

が自由度 $N-1$ の χ^2 分布をなすことより、$\frac{S_3}{\sigma^2}$ が自由度 $N-$

$$N-1-(a-1)-(b-1)=ab-a-b+1=(a-1)(b-1)$$

の χ^2 分布をなすことを知る。

〔註1〕 坂元平八「統計量の独立性に就て」 昭和十八年七月・数物年会講演，統計数理研究所講究録 第一巻第九号 昭和十九年十一月十五日発行．

〔註2〕 G. C. Cochran, The distributeon of Covariance, Proc. Camb. of Phil. Soc, Vol. 30 (1934), pp. 178–191.

〔註3〕 S. S. Wilks Mathematical Statistics, Princeton, N.J., 1943

〔註4〕 A. T. Craig, A Certain mean-valve Problem in Statistics. Bulletin of the American Mathematical Society XIII, 1936, pp. 670–674.

〔註5〕 〔註1〕と同じ．

〔註6〕 〔註2〕と同じ．

〔訂正〕 P. 156, 12段の $(\theta_1^{(j)}, \theta_2^{(j)}, \cdots, \theta_n^{(j)}) = A_j' x$ は正確には $(\theta_1^{(j)}, \theta_2^{(j)}, \cdots, \theta_n^{(j)}) = \theta$ として $\theta' = A_j' x'$ と書くべきである。すると

$$S_j = \sum_{i=1}^{n} \{\theta_i^{(j)}\}^2 = (A_j' x', A_j' x') = (A_j A_j' x', x')$$

$= (B_j x', x')$ となり，以下の証明もすべて書き直されるが，その書き変へは容易であらう。結果は勿論少しも変らない。