ON TRANSFORMATIONS USEFUL IN THE DISTRIBUTION PROBLEMS OF LEAST SQUARES

By C. RADHAKRISHNA RAO
Statistical Laboratory, Calcutta

SECTION 1

One of the problems in univariate distributions is to find the distribution of the minimum value of

$$R^2 = \sum_{1}^{n} (y_i - a_{i1}\tau_1 - \ldots - a_{ik}\tau_k)^2$$

with respect to parameters $\tau_1, \tau_2, \ldots$ subject to s independent conditions

$$\begin{aligned} f_{11}\tau_1 + \ldots + f_{1k}\tau_k &= g_1 \\ \cdot \quad \ldots \quad \cdot \quad & \quad \cdot \\ f_{s1}\tau_1 + \ldots + f_{sk}\tau_k &= g_s \end{aligned} \qquad \ldots \quad (1.0)$$

on the assumption, that y_i is an independent observation from a normal population with variance σ^2 not depending on i and expectation equal to

$$a_{i1}\tau_1 + \ldots + a_{ik}\tau_k$$

the compounding co-efficients a_{ij} being known.

The complete solution to this problem is given by the author in two earlier papers (Rao, 1946, 1952). An elegant proof using a suitable orthogonal transformation is given here. This transformation is fundamental in the derivation of most of the univariate and multivariate distributions.

The following definitions and results will be used. (i) If P is a matrix of m columns and rank r then there exists a matrix $\bar{P}$ of m columns and $(m-r)$ rows with rank $(m-r)$ such that

$$\bar{P}P' = 0 \quad \text{and} \quad \bar{P}\bar{P}' = I \qquad \ldots \quad (1.1)$$

The matrix $\bar{P}$ is not unique but it is necessary that the matrix $\bar{P}'\bar{P}$ be unique for if $\bar{P}_1$ is another matrix satisfying (1.1) then $\bar{P}_1 = C\bar{P}$ where C is an orthogonal matrix. Therefore $\bar{P}_1'\bar{P}_1 = \bar{P}'\bar{P}$. Also if Q is such that $QP' = 0$ then there exists a matrix Λ satisfying the equation

$$Q = \Lambda\bar{P} \qquad \ldots \quad (1.2)$$

(ii) If P is a matrix (q, m) and rank q then there exists a non-singular matrix C, (q, q) such that

$$CPP'C' = I \qquad \ldots \quad (1.3)$$

This follows from the fact that PP' is a positive definite matrix.

(iii) The equations

$$\underset{\sim}{x} P' = \underset{\sim}{d}$$

always admit a solution if the rank of P is equal to the number of its rows.

(iv) Ranks of $A A'$, ACA' (where C is positive definite) and A are the same.

Now in the above problem we consider the matrices

$$A' = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \cdot & \dots & \cdot \\ a_{n1} & \dots & a_{nk} \end{pmatrix} \qquad F = \begin{pmatrix} f_{11} & \dots & f_{1k} \\ \cdot & \dots & \cdot \\ f_{s1} & \dots & f_{sk} \end{pmatrix}$$

$$D = AA',\ B = \bar{D}F',\ \bar{B},\ B\bar{B}' = \bar{B}F\,\bar{D}' = 0$$

Then $\bar{B}F = \Lambda D = \Lambda AA'$. By (1.2) Λ exists. Let the number of rows in matrix $\bar{B}$ be t. The matrix $\bar{B}$ is of the type (t, s) with rank t and $\bar{B}F$ has also the same rank since $\bar{B}FF'\bar{B}'$ has the same rank (by result iv) as $\bar{B}$. Then ΛD is of type (t, k) with rank t or $\Lambda AA'$ is of the type (t, k) with rank t, which implies that ΛA is a matrix with t rows and has the same rank. There exists by (1.3) a matrix M such that

$$M(\Lambda A)(A'\Lambda')M' = I$$

Let $M\Lambda A = G$ and $H = \left(\frac{\bar{A}}{G}\right)$. Then the compound matrix

$$T = \left(\frac{H}{\bar{H}}\right) \qquad \dots \quad (1.4)$$

is orthogonal because

$$\bar{A}\bar{A}' = I,\ GG' = I,\ \bar{A}G' = (\bar{A}A')\Lambda'M' = 0 \ \text{ or } \ HH' = I$$

and by construction $H\bar{H}'=0$ and $\bar{H}\bar{H}'=I$. Let the ranks of A and G be r and t in which case the ranks of $\bar{A}$ and $\bar{H}$ are $(n-r)$ and $(r-t)$.

Consider the transformation

$$\underset{\sim}{x} = \underset{\sim}{y}\, T$$

The vector $\underset{\sim}{\tau} A$ transforms to

$$\underset{\sim}{\xi} = \underset{\sim}{\tau} A(\bar{A}' \,|\, G' \,|\, \bar{H}')$$

$$= (0 \,|\, \underset{\sim}{\tau} AG' \,|\, \underset{\sim}{\tau} A\bar{H}')$$

Now

$$\underset{\sim}{\tau} AG' = \underset{\sim}{\tau} AA'\Lambda'M' = \underset{\sim}{\tau} F'\bar{B}'M' = \underset{\sim}{g}\,\bar{B}'M' = (\eta_1, \dots, \eta_t)$$

where $\underset{\sim}{g} = (g_1, \dots, g_s)$ is the vector of constants in the restrictions (1.0) and the vector $\underset{\sim}{\tau} AG'$ has specified values $(\eta_1, \dots, \eta_t)$ as elements when the restrictions hold. Let $\zeta = (\zeta_1, \dots, \zeta_{r-t}) = \underset{\sim}{\tau} A\bar{H}'$. Since the transformation is orthogonal

$$(\underset{\sim}{y} - \underset{\sim}{\tau} A)^2 = \Sigma(y_i - a_{1i}\tau_1 - \dots - a_{ki}\tau_k)^2 \qquad \dots \quad (1.5)$$

where we put

$$\Delta = (r_1-r_2)(r_2-r_3)(r_3-r_1). \qquad \dots \quad (10)$$

If we put further

$$r_1 = d, \quad r_2 = 2d \quad \text{and} \quad r_3 = 3d, \qquad \dots \quad (11)$$

then the following simple formula are obtained:

$$\hat{\phi} = A_3 - 3(A_2 - A_1), \qquad \dots \quad (12)$$

$$\hat{\lambda} = (8A_2 - 5A_1 - 3A_3)/(2d), \qquad \dots \quad (13)$$

and

$$\hat{\nu} = (A_1 + A_3 - 2A_2)/(2\pi d^2). \qquad \dots \quad (14)$$

Note that $\hat{\phi}$ is independent of d. Although this result is simple, but this design may not be "optimal" to estimate ϕ. To define the optimal design, we must take into consideration various circumstances which we might meet in actual survey. However, we shall confine ourselves to discuss only the variance of estimate.

6. The sampling error is estimated as a random fluctuation of counts per visited point (see Appendix B). We shall denote the standard deviation in population by σ with variate in bracket. Then we can put

$$r_1 = 1, \ r_2 = 1+X \ (X>0), \ \text{and} \ r_3 = 1+X+Y \ (Y>0) \qquad \dots \quad (15)$$

and

$$\sigma(A_2) = (1+\xi)\sigma(A_1), \quad \sigma(A_3) = (1+\eta)\sigma(A_1), \qquad \dots \quad (16)$$

without loss of generality, because ϕ is a function of the ratio $r_1:r_2:r_3$. Then, if the estimates $\hat{p}_i$'s are mutually uncorrelated—generally speaking, it need not be so—the ratio $\sigma^2(\hat{\phi})/\sigma^2(A_1)$ is given by

$$f(X, Y) = \left(\frac{1+X}{X}\right)^2 \left(\frac{1+X+Y}{X+Y}\right)^2 + \frac{(1+\xi)^2(1+X+Y)^2}{X^2Y^2} + \frac{(1+\eta)^2(1+X)^2}{Y^2(X+Y)^2} \qquad \dots \quad (17)$$

As X and Y are positive the equations

$$XY - (1+\xi)(1+X+Y) = (X-1)(Y-1) - 2 - \xi(1+X+Y), \qquad \dots \quad (18)$$

and

$$Y(X+Y) - (1+\eta)(1+X) = (1+X+Y)(Y-1) - \eta(1+X) \qquad \dots \quad (19)$$

suggest that if both ξ and η are non-negative and finite*, X and Y should be greater than 1 to minimize $f(X, Y)$. $f(X, Y)$ is always greater than 1 and is equal to 1, when both X and Y tend to infinity. Thus the greater both X and Y are, the better. In

* We assume here that ξ and η are finite. But in general ξ and η are an increasing function of X and of $(X+Y)$ respectively. Hence $f(X, Y)$ may or may not have an extremum. We owe this remark to Dr. D. J. Finney.

where $H = \bar{A}$. The rank of AH' is the same as that of A. $\underset{\sim}{\tau} A$ transforms to

$$\underset{\sim}{\tau} A(\bar{A}' | H') = (0 | \underset{\sim}{\tau} AH')$$

and $(\underset{\sim}{y} - \underset{\sim}{\tau} A)^2$ transforms to

$$\begin{aligned} & x_1^2 + \ldots + x_{n-r}^2 \\ & + (x_{n-r+1} - \zeta_1)^2 + \ldots + (x_n - \zeta_r)^2 \end{aligned} \qquad \ldots \quad (2.1)$$

where $\underset{\sim}{\tau} AH' = (\zeta_1, \ldots, \zeta_r)$. The minimum value is

$$x_1^2 + \ldots + x_{n-r}^2$$

when the second expression is identically zero, that is when

$$(x_{n-r+1}, \ldots \; x_n) = \underset{\sim}{\tau} AH'$$

which admits a solution since the matrix AH' has the rank r equal to the number of its rows. The minimum is distributed as $\sigma^2 \chi^2$ with $(n-r)$ degrees of freedom.

The expression for the minimum sum of squares is

$$x_1^2 + \ldots + x_{n-r}^2 = \underset{\sim}{y} \bar{A}' \bar{A} \underset{\sim}{y}' \qquad \ldots \quad (2.2)$$

and this is unique for all choices of $\bar{A}$ by result (i) in section 1. Also the difference of the expressions (1.7) and (2.2) is

$$(x_{n-r+1} - \eta_1)^2 + \ldots + (x_{n-r+t} - \eta_t)^2 \qquad \ldots \quad (2.3)$$

which is distributed as $\sigma^2 \chi^2$ with t degrees of freedom independently of (2.2).

SECTION 3

A second problem is that of determining the distribution of the ratio

$$U = \frac{\min \sum_1^q (y_i - a_{1i}\tau_1 - \ldots - a_{ki}\tau_k)^2}{\min \sum_1^n (y_i - a_{1i}\tau_1 - \ldots - a_{ki}\tau_k)^2}$$

where $n > q$ so that the numerator is only the minimum value of a partial sum of squares, the assumption on y_i being the same as in the first problem.

This test is useful in some problems of specification as indicated later. Also the distributions of a number of multivariate statistics can be deduced by using the orthogonal transformation constructed for solving this problem. Consider, the matrices

$$A'_1 = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \cdot & \cdots & \cdot \\ a_{q1} & \cdots & a_{qk} \end{pmatrix} \text{ rank } r_1; \qquad \text{rank } \bar{A}_1 = (q-r_1)$$

$$A'_2 = \begin{pmatrix} a_{q+1,1} & \cdots & a_{q+1,k} \\ \cdot & \cdots & \cdot \\ a_{n1} & \cdots & a_{nk} \end{pmatrix} \text{ rank } r_2$$

$$A = (A_1 | A_2) \qquad \text{rank } r, \qquad \text{rank } AA' = r$$

and

$$H = \left(\frac{\bar{A}_1 | 0}{A_1 | A_2}\right) \quad \text{rank } p, \qquad \text{rank } H\,H' = p$$

$$H\,H' = \left(\frac{I}{0}\middle|\frac{0}{A\,A'}\right)$$

This has rank equal to $(q-r_1+\text{rank } A\,A')=(q-r_1+r)$ Therefore $p=q+r-r_1$ and the rank of $\bar{H}$ is $n-p = n-q-r+r_1$. Also

$$H\,\bar{H}' = O = \left(\frac{\bar{A}_1 | O}{A}\right)\bar{H}' = \left(\frac{(\bar{A}_1|O)\bar{H}'}{A\bar{H}'}\right)$$

which implies that $A\bar{H}' = O$. The transformation

$$\underset{\sim}{x} = \underset{\sim}{y}\left(\frac{A'_1}{O}\middle|\bar{H}'\middle|\bar{R}'\right) \quad \text{where} \quad R = \left(\frac{\bar{A}_1|O}{H}\right)$$

is orthogonal by construction and this changes $\underset{\sim}{\tau} A$ to

$$\underset{\sim}{\tau} A\left(\frac{\bar{A}_1'}{O}\middle|\bar{H}'|\bar{R}'\right) = (O|O|\underset{\sim}{\tau} A\bar{R}')$$

The sum of squares

$$\sum_1^n (y_i - a_{i1}\tau_1 - \ldots)^2.$$

changes over to

$$\begin{aligned} & x_1^2 + \ldots + x^2_{q-r_1} \\ & + x^2_{q-r_1+1} + \ldots + x^2_{n-r} \\ & + (x_{n-r+1} - \zeta_1)^2 + \ldots + (x_n - \zeta_r)^2 \end{aligned} \qquad \ldots \quad (3.1)$$

where $(\zeta_1, \ldots \zeta_r) = \underset{\sim}{\tau} A\,\bar{R}'$. The minimum value is as in section 2

$$\begin{aligned} & x_1^2 + \ldots + x^2_{q-r_1} \\ & + x^2_{q-r_1+1} + \ldots + x^2_{n-r} \end{aligned} \qquad \ldots \quad (3.2)$$

Also by (2.2) the expression

$$x_1^2+\ldots+x^2_{q-r_1} = (y_1, \ldots, y_q)A_1'A_1(y_1, \ldots, y_q)$$

is the minimum value of the partial sum of squares

$$\sum_1^q (y_i-a_{i1}\tau_1-\ldots)^2$$

The statistic U is then the ratio

$$\frac{x_1^2+\ldots+x^2_{q-r_1}}{(x^2_1+\ldots+x^2_{q-r_1})+(x^2_{q-r_1+1}+\ldots+x^2_{n-r})}$$

which is distributed as

$$\chi^2_1/(\chi^2_1+\chi^2_2)$$

where χ_1^2 and χ_2^2 are independently distributed with degrees of freedom $(q-r_1)$ and $(n-q-r+r_1)$. The distribution of U is of the beta form

$$B\left(\frac{q-r_1}{2}, \frac{n-q-r+r_1}{2}\right)dU$$

$$= \text{const. } U^{\frac{q-r_1}{2}-1}(1-U)^{\frac{n-q-r+r_1}{2}-1}dU \quad \ldots \quad (3.3)$$

It may be observed that this result can be deduced from the general proposition in section 1 by considering $2k$ parameters in the expectations

$$E(y_i) = a_{i1}\tau_1+\ldots+a_{ik}\tau_k, \quad i = 1, \ldots, q$$

$$= a_{i1}\tau'_1+\ldots+a_{ik}\tau'_k, \quad i = q+1, \ldots, n$$

If

$$R_1^2 = \min \sum_1^q (y_i-a_{i1}\tau_1-\ldots)^2$$

$$R_2^2 = \min \sum_{q+1}^n (y_i-a_{i1}\tau'_1-\ldots)^2$$

and

$$R^2 = \min \sum_1^n (y_i-a_{i1}\tau_1-\ldots)^2$$

then R_1^2, R_2^2 are independently distributed as $\sigma^2\chi^2$ with $(q-r_1)$ and $(n-q-r_2)$ degrees of freedom and R^2, the conditional minimum when $\tau_i = \tau'_i$, is distributed as $\sigma^2\chi^2$ with $(n-r)$ degrees of freedom, and by result (2.3) of section 2

$$R^2-(R_1^2+R_2^2)$$

is distributed as $\sigma^2\chi^2$ with

$$n-r-(q-r_1)-(n-q-r_2) = (r_1+r_2-r)$$

degrees of freedom independently of R_1^2 and R_2^2. Therefore R_1^2 and $R^2-R_1^2$ are distributed with degrees of freedom $(q-r_1)$ and $(r_1+r_2-r)+(n-q-r_2)=(n-q-r+r_1)$. Therefore

$$U = R_1^2/R^2$$

has the distribution derived in (3.3). This test is useful in situations where accepting the specification of the expected values of $y_1,\dots,y_q$ we may have to test whether the expected values of $y_{q+1}, \dots, y_n$ have the given specification in terms of the unknown parameters $\tau_1, \dots, \tau_k$. It is thus a test of the specied values of the compounding co-efficients of the unknowns $\tau_1,\dots,\tau_k$ occurring in the expected values of

$$y_{q+1}, \dots, y_n$$

As an illustration of the use of the above distribution we consider the distribution problem of the ratio of determinants

$$|S'_{ij}|_p \div |S_{ij}|_p$$

where

$$S'_{ij} = \sum_1^q x_{ir}x_{jr}, \quad S_{ij} = \sum_1^n x_{ir}x_{jr}$$

$$i, j = 1, 2, \dots, p$$

and

$$x_{1r}, \dots, x_{pr}$$
$$r = 1, \dots, n$$

are n independent sets of observations from a multivariate normal distribution with zero mean values. Consider the conditional distribution of x_p given $x_1, \dots, x_{p-1}$ in which

$$E(x_p) = \beta_1 x_1 + \dots + \beta_{p-1}x_{p-1}$$

The minimum value of

$$\Sigma\,(x_{pr}-\beta_1 x_{1r}-\dots)^2$$

when the summation is from 1 to $q(>p)$ is

$$|S'_{ij}|_p \div |S'_{ij}|_{p-1}$$

and when the summation is from 1 to n is

$$|S_{ij}|_p \div |S_{ij}|_{p-1}$$

The distribution of the ratio

$$R_p = M_p \div M_{p-1} = \frac{|S'_{ij}|_p}{|S_{ij}|_p} \div \frac{|S'_{ij}|_{p-1}}{|S_{ij}|_{p-1}}$$

is

$$B\left(\frac{q-p+1}{2},\ \frac{n-q}{2}\right)$$

using the result (3.3). Changing p to $(p-1)$ we find the distribution of R_{p-1} and so on. All these are independently distributed so that the distribution of the product

$$R_p R_{p-1} \dots R_1 = |S'_{ij}|_p \div |S_{ij}|_p$$

is same as that of the product of p beta variables having the distibutions

$$B\left(\frac{q-p+1}{2}, \frac{n-q}{2}\right) B\left(\frac{q-p+2}{2}, \frac{n-q}{2}\right) \dots B\left(\frac{q}{2}, \frac{n-q}{2}\right)$$

If $q = (n-1)$ the product itself reduces to the beta form

$$B\left(\frac{n-p}{2}, \frac{1}{2}\right)$$

as shown in (Rao, 1952, page 46).

REFERENCES

RAO, C. R. (1946): On the linear combination of observations and the general theory of least squares. *Sankhyā*, 7, 237.

RAO, C. R. (1952): *Advanced Statistical Methods in Biometric Research*, John Wiley & Sons, New York.

Paper received : December, 1952.