## DISCUSSION

T. W. ANDERSON[1]

*Stanford University*

As stated by the author in Section 3, his emphasis in the study of

$$Dy = \Gamma(\theta) = \sum_{a=1}^{s} \theta_a A_a \tag{1}$$

is different from mine [Anderson (1969, 1970, 1973)]. A brief exposition of this other point of view might put the present paper into a larger perspective. I shall use the notation of the present paper as well as the assumption of $Ey = 0$ and corresponding modifications in statements.

Anderson considered $N$ observations from $N(0, \Gamma(\theta))$; in the present paper $N = 1$. We suppose $A_1, \dots, A_s$ to be symmetric matrices and that for some vector $\theta = (\theta_1, \dots, \theta_s)'$ the matrix $\sum_{a=1}^{s} \theta_a A_a$ is positive definite. Then the derivative equations for the maximum likelihood estimate of $\theta$ are

$$\operatorname{tr}\left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} A_b = y'\left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} A_b \left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} y, \qquad b = 1, \dots, s. \tag{2}$$

If there are several vectors $\hat{\theta}$ satisfying (2), the vector minimizing $|\sum_{a=1}^{s} \hat{\theta}_a A_a|$ is taken. Rewriting (2) as

$$\begin{aligned} &\sum_{c=1}^{s} \operatorname{tr}\left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} A_b \left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} A_c \hat{\theta}_c \\ &\qquad = y'\left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} A_b \left(\sum_{a=1}^{s} \hat{\theta}_a A_a\right)^{-1} y, \qquad b = 1, \dots, s, \end{aligned} \tag{3}$$

suggests an iterative procedure for solving (2). Let the vector $\hat{\theta}^{(0)}$ be an initial estimate, and define $\Gamma(\hat{\theta}^{(i)}) = \sum_{a=1}^{s} \hat{\theta}_a^{(i)} A_a$, $i = 0, 1, \dots$. The $i$th stage of the iteration is solving the linear equations

$$\begin{aligned} &\sum_{c=1}^{s} \operatorname{tr} \Gamma^{-1}(\hat{\theta}^{(i-1)}) A_b \Gamma^{-1}(\hat{\theta}^{(i-1)}) A_c \hat{\theta}_c^{(i)} \\ &\qquad = y' \Gamma^{-1}(\hat{\theta}^{(i-1)}) A_b \Gamma^{-1}(\hat{\theta}^{(i-1)}) y, \qquad b = 1, \dots, s. \end{aligned} \tag{4}$$

This is the method of scoring.

If the matrices $\{A_a\}$ commute [the author's (c)], there exists an orthogonal matrix $B$ such that $A_a = B \Lambda_a B'$ where $\Lambda_a$ is diagonal with diagonal elements $\lambda_{ta}$, $t = 1, \dots, n$. Then

$$B'\Gamma(\theta) B = \sum_{a=1}^{s} \theta_a \Lambda_a \tag{5}$$

[1]Work sponsored by the Army Research Office Contract No. DAAG 29-85-0239 at Stanford University and written while the author was a National Research Council Resident Research Associate at the Naval Postgraduate School at Monterey, California.

is diagonal. If $z = B'y$, then (3), for example, can be written

$$\sum_{c=1}^{s}\sum_{t=1}^{n}\frac{\lambda_{tb}\lambda_{tc}}{\left(\sum_{a=1}^{s}\hat{\theta}_a^{(i-1)}\lambda_{ta}\right)^2}\hat{\theta}_c^{(i)} = \sum_{t=1}^{n}\frac{\lambda_{tb}z_t^2}{\left(\sum_{a=1}^{s}\hat{\theta}_a^{(i-1)}\lambda_{ta}\right)^2}. \tag{6}$$

To take account of the multiplicities of roots we can order them so that we can write

$$\Lambda_a = \begin{bmatrix} \nu_{1a}I & 0 & \cdots & 0 \\ 0 & \nu_{2a}I & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \nu_{ma}I \end{bmatrix}, \tag{7}$$

where the orders of the $I$'s on the main diagonal are $n_1, \dots, n_m$, respectively, and $\sum_{j=1}^{m} n_j = n$. That the matrix

$$\begin{bmatrix} \nu_{11} & \nu_{12} & \cdots & \nu_{1s} \\ \nu_{21} & \nu_{22} & \cdots & \nu_{2s} \\ \vdots & \vdots & & \vdots \\ \nu_{m1} & \nu_{m2} & \cdots & \nu_{ms} \end{bmatrix} \tag{8}$$

is of rank $m$ is equivalent to $\{A_a\}$ being linearly independent. Define $V_j$ by

$$n_j V_j = \sum_{t=n_1+\cdots+n_{j-1}+1}^{n_1+\cdots+n_j} z_t^2, \qquad j = 1, \dots, m. \tag{9}$$

Then (6) can be written

$$\sum_{c=1}^{s}\sum_{j=1}^{m}\frac{n_j\nu_{jb}\nu_{jc}}{\left(\sum_{a=1}^{s}\hat{\theta}_a^{(i-1)}\nu_{ja}\right)^2}\hat{\theta}_c^{(i)} = \sum_{j=1}^{m}\frac{n_j\nu_{jb}V_j}{\left(\sum_{a=1}^{s}\hat{\theta}_a^{(i-1)}\nu_{ja}\right)^2}, \qquad b = 1, \dots, s. \tag{10}$$

If $m = s$ and (8) is nonsingular, then

$$\sum_{c=1}^{s}\nu_{jc}\hat{\theta}_c = V_j, \qquad j = 1, \dots, m, \tag{11}$$

can be solved uniquely and $\hat{\theta}^{(i-1)} = \hat{\theta}$; it is the unique maximum likelihood estimate. What the author means by "in general" $\Gamma(\theta)$ has $s$ distinct eigenvalues [the author's (d)] is that (8) is square and nonsingular. In this case for any $\hat{\theta}^{(0)}$ the solution of (10) for $i = 1$ is $\hat{\theta}$ [Szatrowski (1980)].

Since $S_1, \dots, S_m$ are linear combinations of $A_1, \dots, A_m$, $B'S_1B = T_1, \dots, B'S_mB = T_m$, say, are diagonal. Because $T_1, \dots, T_m$ are idempotent as well as diagonal, their diagonal elements are 0's and 1's. In fact, $T_\alpha$ has $I$ in the $\alpha$th diagonal block, corresponding to the partitioning of (7). The author's equation $A_a S_\alpha = p_{\alpha a} S_\alpha$ is equivalent to $\Lambda_a T_\alpha = p_{\alpha a} T_\alpha$. Hence, $p_{\alpha a} = \nu_{\alpha a}$. In this canonical form the maximum likelihood estimates (in the author's notation) are $\hat{\xi}_\alpha = V_\alpha (d_\alpha = n_\alpha)$. There is a considerable literature on optimal estimation of variance components, that is, of $\theta$. However, when $m = s$ and (8) is nonsingular, the maximum likelihood estimators $\hat{\xi}_\alpha$ are minimum variance unbiased estimators; $d_\alpha \hat{\xi}_\alpha / \xi_\alpha$ has a $\chi^2$-distribution with $d_\alpha$ degrees of freedom and $\hat{\xi}_1, \dots, \hat{\xi}_s$ are independent.

Whenever the distribution of the observable vector $y$ is needed in this paper, it is assumed to be normal. However, an analysis of variance makes sense within the context of elliptically contoured distributions. In this case if $y$ has a density, it can be written

$$|\Gamma|^{-1/2} g(y'\Gamma^{-1}y), \tag{12}$$

where $g(x'x)$ is a density in $R^n$. If

$$\int_0^\infty v^{n/2} g(v)\, dv < \infty, \tag{13}$$

the first two moments of $y$ exist and $Ey = 0$, $Eyy' = \Gamma$. The likelihood function has a maximum at $\xi = (n/v_g)\hat{\xi}$, where $v_g$ is the value of $v$ maximizing $v^{n/2}g(v)$ and $\hat{\xi}$ is the maximum likelihood estimator under normality [Anderson, Fang and Hsu (1986, Theorem 1)]. The uncorrelatedness of $S_\alpha y_t$ and $S_\beta y_u$, $\alpha \neq \beta$, holds, but in general, independence of quadratic forms does not hold. For example, if $y'S_\alpha y$ and $y'(I - S_\alpha)y$ are independent, the distribution of $y$ must be normal [Anderson and Fang (1987, Theorem 1)]. Nevertheless, $F$-tests are valid [Anderson, Fang and Hsu (1986, Theorem 2)].

## REFERENCES

ANDERSON, T. W. (1969). Statistical inference for covariance matrices with linear structure. In *Multivariate Analysis* (P. R. Krishnaiah, ed.) **2** 55–66. Academic, New York.

ANDERSON, T. W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In *Essays in Probability and Statistics* (R. C. Bose et al., eds.) 1–24. Univ. of North Carolina Press, Chapel Hill, N.C.

ANDERSON, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Ann. Statist.* **1** 135–141.

ANDERSON, T. W. and FANG, K.-T. (1987). Distributions of quadratic forms and Cochran's theorem for elliptically contoured distributions. *Sankhyā*. To appear.

ANDERSON, T. W., FANG, K.-T. and HSU, H. (1986). Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. *Canad. J. Statist.* **14** 55–59.

SZATROWSKI, T. H. (1980). Necessary and sufficient conditions for explicit solutions in the multivariate normal estimation problem for patterned means and covariances. *Ann. Statist.* **8** 802–810.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

R. A. BAILEY

*Rothamsted Experimental Station*

It is a pleasure to read this unified account of the analysis of variance, and the relationship between its many facets, for variance models based on association schemes. The theory of association schemes is an elegant piece of mathematics, as the recent book by Bannai and Itô (1984) shows, with many areas of