Ann. Inst. Statist. Math.
30 (1978), Part A, 81-88

EXPLICIT SOLUTIONS, ONE ITERATION CONVERGENCE AND AVERAGING IN THE MULTIVARIATE NORMAL ESTIMATION PROBLEM FOR PATTERNED MEANS AND COVARIANCES

TED H. SZATROWSKI

(Received July 20, 1977; revised Feb. 21, 1978)

1. Introduction

Anderson [1], [2] studies the problem of estimation of the mean vector and covariance matrix of a multivariate normal distribution when the mean vector and covariance matrix have linear structure. In the [2] paper, he presents the likelihood equations (L.E.) and suggests an iterative algorithm for finding the solutions of the L.E. based on the method of scoring. In the present study, sufficient conditions are presented for (1) the existence of explicit maximum likelihood estimates (M.L.E.), (2) the convergence of the iterative procedure proposed by Anderson [2] in one iteration from any positive definite starting point, and (3) the solution to the L.E. being found by solving a set of averaging equations. These results are applied to some well-known problems involving structured mean vectors and covariance matrices including the block and non-block cases of complete, compound and circular symmetry. In these cases, the M.L.E. for the elements of the patterned mean vector may be found by averaging corresponding elements in the sample mean vector; M.L.E. for the elements of the patterned covariance matrix (P.C.M.) may be found by averaging corresponding elements in the sample covariance matrix.

In Section 2, the problem under consideration is described in detail along with the iterative algorithm proposed by Anderson for its solution and the method of averaging. The main results appear in Theorem 1 of Section 3. Applications of Theorem 1 are given in Section 4.

2. Likelihood equations and averaging

Let X be a p-component column vector with multivariate normal

Some key words: Patterned covariance matrices, patterned means, maximum likelihood estimation, explicit solutions, convergence, averaging, initial estimates, complete symmetry, compound symmetry, circulant.
AMS 1970 subject classification: 62H05, 62H15.

distribution such that the mean vector $\mathcal{E}X=\mu$ and covariance matrix $\mathcal{C}(X)=\mathcal{E}(X-\mu)(X-\mu)'=\Sigma$ may be expressed in the forms $\mu=\sum_{j=1}^{r}\beta_j z_j$ and $\Sigma=\sum_{g=0}^{m}\sigma_g G_g$. Here the β's and σ's are unknown scalars, the z's and G's are known linearly independent p-component vectors and known linearly independent symmetric $p\times p$ matrices, respectively. We assume that there exists at least one set of σ's such that Σ is positive definite. Estimates of μ and Σ in these forms are desired based on N independent p-dimensional observations, $x_1,\cdots,x_N$.

The likelihood equations (L.E.) for this problem as given by Anderson [2] are

$$\sum_{i=1}^{r}(z_j'\hat{\Sigma}^{-1}z_i)\hat{\beta}_i=z_j'\hat{\Sigma}^{-1}\bar{x}\ , \qquad j=1,\cdots,r\ , \tag{2.1}$$

$$\sum_{h=0}^{m}(\operatorname{tr}\hat{\Sigma}^{-1}G_g\hat{\Sigma}^{-1}G_h)\hat{\sigma}_h=\operatorname{tr}\hat{\Sigma}^{-1}G_g\hat{\Sigma}^{-1}C\ , \qquad g=0,\cdots,m\ , \tag{2.2}$$

where $\bar{x}=(1/N)\sum_{\alpha=1}^{N}x_\alpha$, is the sample mean and

$$C=(1/N)\sum_{\alpha=1}^{N}(x_\alpha-\bar{x})(x_\alpha-\bar{x})'+(\bar{x}-\hat{\mu})(\bar{x}-\hat{\mu})'\ .$$

Anderson points out that the L.E. written in this form suggest an iterative scheme (noted by Rao [8] to correspond to the method of scoring) wherein from an initial estimate of Σ one can solve the linear equations in the β_i's and σ_h's to yield the next estimate of Σ. A well-known result concerning the existence of a unique solution to these linear equations is given by

LEMMA 1. *If $\hat{\Sigma}$, and thus $\hat{\Sigma}^{-1}$, are chosen positive definite, then the linear iterative equations* (2.1) *and* (2.2) *have a unique solution.*

The "method of averaging" involves solving the L.E. (2.1) and (2.2) using $\hat{\Sigma}=I$, the identity matrix. The motivation for studying the method of averaging arises from its relationship to naively averaging elements of the sample mean and covariance matrix to obtain estimates of the β's and σ's. Specifically, the method of averaging results in estimates of the β's and σ's which may also be obtained by averaging groups of elements in the sample mean and covariance matrix when the non-zero elements of the z vectors are all equal and the non-zero elements of the G matrices are all equal. This last condition is equivalent to being able to reparameterize the z vectors and G matrices so that they consist only of zeroes and ones.

Examples where the solutions to the averaging equations and solutions obtained by averaging groups of elements in the sample mean and covariance differ are easy to find. Ignoring the problem of estimation of the mean vector, we need only consider a covariance matrix with at least one G matrix having ones and twos. For example, let Σ be of the form,

$$\Sigma=\begin{bmatrix} b & 0 \\ 0 & 2b \end{bmatrix}=b\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

The averaging equations result in the estimate of b given by $(c_{11}+2c_{22})/5$, where the c's are elements of the 2×2 sample covariance matrix C. Ordinary averaging would lead to the estimate $(c_{11}+c_{22})/3$. Neither of these solutions is the M.L.E. given by $(2c_{11}+c_{22})/4$. In this particular case, it is interesting to note that the method of scoring gives us the M.L.E. in one iteration from any positive definite starting point.

Note also that the method of averaging does not always yield a positive definite covariance matrix.

$$\Sigma=\begin{bmatrix} a & b & c \\ b & a & c \\ c & c & a \end{bmatrix}; \qquad C=\begin{bmatrix} 9 & 8 & 0 \\ 8 & 9 & 0 \\ 0 & 0 & 3 \end{bmatrix}; \qquad \Sigma_{\mathrm{AVE}}=\begin{bmatrix} 7 & 8 & 0 \\ 8 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix} \not> 0 .$$

3. Main results

In this section, sufficient conditions for explicit M.L.E., one-iteration convergence and averaging yielding the M.L.E. are given in Theorem 1. This theorem involves transforming the patterned mean vector μ and covariance matrix Σ by an orthogonal matrix P, resulting in a transformed mean and covariance matrix of the form $\nu=P\mu=(\nu_1',\cdots,\nu_q')'$ and $\Psi=P\Sigma P'=\mathrm{diag}\,(\Psi_1,\cdots,\Psi_q)$. The components of ν and Ψ have corresponding dimensions, i.e. if ν_j is $p_j\times 1$, Ψ_j is $p_j\times p_j$, $j=1,\cdots,q$. The corresponding components of ν and Ψ, (ν_j,Ψ_j), $j=1,\cdots,q$, also have special structure. We use the Kronecker product, defined by $A\otimes B=(a_{ij}B)$, in specifying this special structure. Ψ_j has a block diagonal form, $\Psi_j=D_j\otimes U_j$ where D_j is a diagonal matrix with known positive elements of dimension $e_j\times e_j$ and U_j is an $f_j\times f_j$ symmetric positive definite matrix of unkown parameters. The mean vector ν_j is decomposed into e_j blocks of size f_j corresponding to the e_j U_j matrices that form the block diagonal Ψ_j matrix. The special structure of ν_j and its correspondence with the structure of Ψ_j is given by

$$\nu_j=((w_j\otimes\varphi_j)',\ \vdots\gamma_j',\ \vdots 0')', \tag{3.1}$$

$$\Psi_j=(D_j\otimes U_j)=\mathrm{diag}\,(d_{j1}U_j,\cdots,d_{jg_j}U_j,\ \vdots d_{jg_j+1}U_j,\cdots,d_{jg_j+h_j}U_j,\ \vdots d_{jg_j+h_j+1}U_j,\cdots,d_{je_j}U_j)\ .$$

The first element in the structure of ν_j, $(w_j \otimes \varphi_j)$, consists of g_j repetitions of $\varphi_j(f_j \times 1)$ which correspond to the first g_j blocks of $U_j(f_j \times f_j)$ on the diagonal of Ψ_j. Here w_j is a $g_j \times 1$ vector of knowns and φ_j is an $f_j \times 1$ vector of unknowns. The next element in the structure of ν_j, γ_j, is an $h_j f_j \times 1$ vector of unknowns corresponding to the next h_j blocks of U_j on the diagonal of Ψ_j. The final 0 vector corresponds to the final $e_j-(g_j+h_j)$ blocks of U_j on the diagonal of Ψ_j. A minimum of one of the three structures of which ν_j is composed must be present.

THEOREM 1. *Suppose there exists an orthogonal matrix P with known elements with the properties:*

1. $\nu = P\mu = (\nu_1', \cdots, \nu_q')'$, ν_j *given in* (3.1), $j=1, \cdots, q$, *where*

a. ν_j, w_j, φ_j *and* γ_j *are column vectors of dimensions* p_j, g_j, f_j *and* $h_j f_j$ *respectively*; w_j *has known elements,* φ_j *and* γ_j *have unknown elements and* $g_j \geqq 0$, $h_j \geqq 0$, $g_j + h_j \leqq e_j$ *where* e_j *is the dimension of the diagonal matrices* D_j *defined in* 2 *below,*

b. *there exists a one-to-one correspondence between the scalars* $\beta_1, \cdots, \beta_r$ *and the elements of the vectors* $\varphi_1, \gamma_1, \cdots, \varphi_q, \gamma_q$ *with* $r = \sum_{j=1}^{q} (a_j + h_j) f_j$ *where* $a_j = 1$ *if* $g_j > 0$, 0 *otherwise.*

2. $\Psi = P\Sigma P' = \operatorname{diag}(\Psi_1, \cdots, \Psi_q)$, $\Psi_j = D_j \otimes U_j$, $j=1, \cdots, q$, *where*

a. Ψ_j *is* $p_j \times p_j$; D_j *is an* $e_j \times e_j$ *diagonal matrix with known positive diagonal elements*; *and* U_j *is an* $f_j \times f_j$ *symmetric positive definite matrix with* $f_j(f_j+1)/2$ *unknown elements,* $e_j f_j = p_j$;

b. *there exists a one-to-one correspondence between* $\sigma_0, \cdots, \sigma_m$ *and the upper triangular elements of* $U_1, \cdots, U_q$ *with* $m+1 = \sum_{j=1}^{q} f_j(f_j+1)/2$.

Under these hypotheses, we conclude that

A. μ *and* Σ *have explicit M.L.E.,*

B. *the method of scoring converges in one iteration to the explicit M.L.E. from any set* $\{\sigma_0, \cdots, \sigma_m\}$ *yielding a positive definite covariance matrix* Σ_0 *as a starting point for the iteration,*

C. *the explicit M.L.E. may be obtained by the method of averaging in the case where the* D_j *matrices may be expressed by* $D_j = t_j I_{e_j}$, *where* t_j *is a known positive scalar.*

PROOF. We prove first conclusions A and B by showing that (i) from any positive definite Σ starting point, there is a unique scoring iterate and (ii) that this unique scoring iterate is the explicit solution of the M.L.E. Part (i) follows from Lemma 1.

To prove (ii) we derive the explicit M.L.E. solution and show this solution is the first iterate solution of the L.E. We start with the L.E. in the means

$$z_i'\Sigma^{-1}\mu = z_i'\Sigma^{-1}\bar{x}\ , \qquad i=1, \cdots, r\ .$$

Letting $c_i=Pz_i$ and $\bar{y}=P\bar{x}$, we rewrite these equations as

$$(3.2)\qquad c_i'\Psi^{-1}\nu=c_i'\Psi^{-1}\bar{y}\ ,\qquad i=1,\cdots,r\ .$$

Since $\nu=P\mu=\sum_{i=1}^{r}\beta_iPz_i=\sum_{i=1}^{r}\beta_ic_i=(\nu_1',\cdots,\nu_q')'$, where $\nu_j=((w_j\otimes\varphi_j)',\gamma_j',0')'$, $j=1,\cdots,q$, we know c_i is of the form

$$c_i=(c_{i1}',\cdots,c_{iq}')'\ ,\qquad c_{ij}=((w_j\otimes a_{ij})',b_{ij}',0')'\ ,$$
$$j=1,\cdots,q,\ i=1,\cdots,r\ .$$

Also, since $\Psi=P\Sigma P'=\text{diag}\,(\Psi_1,\cdots,\Psi_q)=\text{diag}\,(D_1\otimes U_1,\cdots,D_q\otimes U_q)$

$$=\sum_{g=0}^{m}\sigma_gPG_gP'=\sum_{g=0}^{m}\sigma_g\Lambda_g\ ,$$

where $\Lambda_g=PG_gP'$, we know Λ_g is of the form

$$\Lambda_g=\text{diag}\,(D_1\otimes\Lambda_{g1},\cdots,D_q\otimes\Lambda_{gq})\ ,\qquad g=0,\cdots,m\ .$$

Noting these forms of ν and Ψ, we may rewrite equation (3.2) yielding

$$\sum_{j=1}^{q}((w_j\otimes a_{ij})',b_{ij}',0')\,\text{diag}\,(d_{j1}^{-1}U_j^{-1},\cdots,d_{je_j}^{-1}U_j^{-1})((w_j\otimes\varphi_j)',\gamma_j',0')'$$
$$=\sum_{j=1}^{q}((w_j\otimes a_{ij})',b_{ij}',0')\,\text{diag}\,(d_{j1}^{-1}U_j^{-1},\cdots,d_{je_j}^{-1}U_j^{-1})\bar{y}_j\ ,$$
$$i=1,\cdots,r\ .$$

Further multiplication results in the simplification:

$$\sum_{j=1}^{q}\left[a_{ij}'U_j^{-1}\left(\sum_{k=1}^{g_j}w_{jk}^2d_{jk}^{-1}\right)\varphi_j+\sum_{k=1}^{h_j}b_{ijk}'d_{j(g_j+k)}^{-1}U_j^{-1}\gamma_{jk}\right]$$
$$=\sum_{j=1}^{q}\left[a_{ij}'U_j^{-1}\left(\sum_{k=1}^{g_j}w_{jk}d_{jk}^{-1}\bar{y}_{jk}\right)+\sum_{k=1}^{h_j}b_{ijk}'d_{j(g_j+k)}^{-1}U_j^{-1}\bar{y}_{j(g_j+k)}\right]\ ,$$
$$i=1,\cdots,r\ .$$

We have partitioned b_{ij} into b_{ijk}, $k=1,\cdots,h_j$ where b_{ijk} is $f_j\times 1$.

In this form, we can see that

$$(3.3)\qquad \varphi_j=\frac{\sum_{k=1}^{g_j}w_{jk}\bar{y}_{jk}/d_{jk}}{\sum_{k=1}^{g_j}w_{jk}^2/d_{jk}}\ ;\qquad \gamma_{jk}=\bar{y}_{j(g_j+k)}\ ;\ k=1,\cdots,h_j,\ j=1,\cdots,q\ ,$$

satisfy these equations and that these solutions are independent of the initial value of the positive definite covariance matrix Σ.

We have shown that in one iteration of the L.E. for the mean μ from a positive definite starting point we arrive at the M.L.E., $\hat{\mu}$. We form the sample covariance matrix C using $\hat{\mu}$, $C=(1/N)\sum_{i=1}^{N}(x_i-\bar{x})(x_i-\bar{x})'+(\bar{x}-\hat{\mu})(\bar{x}-\hat{\mu})'$. The proof of one iteration convergence of the second set of equations in Σ (2.2) involves the substitution of $E=PCP'$, $\Psi=P\Sigma P'$ and $\Lambda_g=PG_gP'$ into these equations and simplifications similar to those used above. This yields the M.L.E. of the U's given by

$$U_j^*=\frac{1}{e_j}\sum_{k=1}^{e_j} d_{jk}^{-1}(E_{jj})_{kk}\ , \qquad j=1,\cdots,q\ , \tag{3.4}$$

where $(E_{jj})_{kk}$ are the diagonal blocks of E_{jj} corresponding to the U_j's in $\Psi_j=D_j\otimes U_j$.

To prove part C, we note that under the new restrictions, there exist $\sigma_0,\cdots,\sigma_m$ so that $\Sigma=I$ is a possible positive definite patterned covariance matrix. Substitution of $\Sigma=I$ as an initial estimate in the scoring equations results in the averaging equations

$$\sum_{i=1}^{r}\beta_i(z_j'z_i)=z_j'\bar{x}\ , \qquad j=1,\cdots,r\ ,$$

$$\sum_{h=0}^{m}\sigma_h \operatorname{tr} G_hG_g=\operatorname{tr} CG_h\ , \qquad g=0,\cdots,m\ .$$

Thus we know the averaging equations yield the explicit M.L.E.

Note that Miller [5] proves part B for the covariance case with $D_j=I_{e_j}$, $j=1,\cdots,q$. Also the form of the mean vector in Theorem 1 can be extended to include duplications of the first structure.

To see that the additional restrictions in part C are necessary, we rewrite the averaging equations in terms of the rotated coordinates and solve for ν and Ψ without the additional constraint on the D matrices. These solutions are then compared to the explicit M.L.E. derived earlier. Rewriting the averaging equations, we find

$$c_j'\nu=c_j'\bar{y}\ , \qquad j=1,\cdots,r\ ; \tag{3.5}$$

$$\operatorname{tr}\Psi\Lambda_g=\operatorname{tr}E\Lambda_g\ , \qquad g=0,\cdots,m\ . \tag{3.6}$$

We rewrite these equations using the special forms of the c's, Λ's, ν and Ψ. Since (3.5) is the same as (3.2) with Ψ replaced by I, we observe the solution of (3.5) is given by

$$\varphi_j=\frac{\sum_{k=1}^{g_j} w_{jk}\bar{y}_{jk}}{\sum_{k=1}^{g_j} w_{jk}^2}\ ; \qquad \gamma_{jk}=\bar{y}_{j(g_j+k)},\ k=1,\cdots,h_j,\ j=1,\cdots,q\ . \tag{3.7}$$

Rewriting the second equation (3.6) yields

$$\sum_{j=1}^{q} \operatorname{tr}\left(\sum_{k=1}^{e_j} d_{jk}^2\right) U_j \Lambda_{gj} = \sum_{j=1}^{q} \operatorname{tr}\left(\sum_{k=1}^{e_j} d_{jk}(E_{jj})_{kk}\right) \Lambda_{gj}, \qquad g=0,\cdots,m.$$

From this equation, we observe the solutions

$$U_j = \frac{\sum_{k=1}^{e_j} d_{jk}(E_{jj})_{kk}}{\sum_{k=1}^{e_j} d_{jk}^2}, \qquad j=1,\cdots,q. \tag{3.8}$$

Comparing the solutions of the averaging equations to those of the M.L.E. given by (3.3) and (3.4), we observe that these solutions are the same if and only if the d_{jk} are independent of k, i.e. $d_{jk}=t_j$, $k=1,\cdots,e_j$, $j=1,\cdots,q$.

4. Applications

In this section, Theorem 1 is applied to several well-known examples of patterned mean and covariance matrices: the cases of complete, compound and circular symmetry. In these cases, averaging the elements of the sample mean and covariance matrix which correspond to a single value in the population mean and covariance matrix yields the M.L.E. This result has not been given in previous studies of block compound symmetry (Arnold [3], [4]) and circular symmetry (Olkin and Press [7], Olkin [6]). It should be noted that these well-known patterns cannot be used to study properties of convergence of the iterative algorithms of Anderson [2] because convergence takes place in one iteration from any positive definite starting point.

The case of complete symmetry (see Wilks [11]) involves μ unstructured or $\mu=(1,\cdots,1)'\otimes\beta$ and a covariance matrix Σ of the form $\Sigma=I_e\otimes(A-B)+J_{e,e}\otimes B$ where $J_{e,e}$ is an $e\times e$ matrix of ones, β is an $f\times 1$ column vector of unknowns and A and B are symmetric $f\times f$ matrices of unknowns. When A and B are scalars, $(f=1)$, we have the non-block case of complete symmetry, for $f>1$, the block case. Using an orthogonal matrix P of the form $P=\Gamma\otimes I_f$, where Γ is any $e\times e$ orthogonal matrix with identical elements in the first row, we find

$$\nu=P\mu=(\sqrt{e}\,\beta', 0', \cdots, 0')'=(\gamma', 0')',$$

$$\Psi=P\Sigma P'=\operatorname{diag}(A+(e-1)B, I_{e-1}\otimes(A-B))=\operatorname{diag}(U_1, I_{e-1}\otimes U_2).$$

Thus all hypotheses of Theorem 1 are satisfied and the M.L.E. may be obtained by averaging.

Compound symmetry (see Votaw [10]) involves several groups of random vectors and structures arising from the assumption of exchangeability of random vectors within groups. Although more complicated in form, the construction of an orthogonal matrix P needed in Theorem 1 only involves Kronecker products and orthogonal matrices with elements of the first row equal as in the case of complete symmetry. For a more complete explanation of compound symmetry, both block and non-block forms, see Arnold [3], [4] and Szatrowski [9].

The case of circular symmetry is studied by Olkin and Press [7] in the non-block form and Olkin [6] in the block form. They show that the circular symmetric covariance matrix can be rotated into a block diagonal form. It is easily seen that this form satisfies all hypotheses of Theorem 1 so that averaging yields the M.L.E.

Acknowledgements

I would like to express my appreciation to my thesis advisor, Professor Ingram Olkin, for suggesting this area of research and for many helpful discussions. Thanks are also due to Professor John J. Miller for useful comments and conversations. This research was partially supported by the National Science Foundation both under a Competitive Fellowship and Contract Number MPS-7509450000.

RUTGERS UNIVERSITY

REFERENCES

[1] Anderson, T. W. (1969). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices, *Essays in Probability and Statistics*, University of North Carolina Press, Chapel Hill, 1-24.

[2] Anderson, T. W. (1973). Asymptotic efficient estimation of covariance matrices with linear structure, *Ann. Statist.*, **1**, 135-141.

[3] Arnold, S. F. (1973). Application of the theory of products of problems to certain patterned covariance matrices, *Ann. Statist.*, **1**, 682-699.

[4] Arnold, S. F. (1976). Applications of products to the generalized compound symmetry problem, *Ann. Statist.*, **3**, 227-233.

[5] Miller, J. J. (1975). Computing of MLE of variance components by functional iteration and scoring, with comments on large and small sample properties, *Rutgers University Technical Report*.

[6] Olkin, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks, *Proceedings of the Symposium on Multivariate Analysis*, Dalhousie, Nova Scotia, 183-195.

[7] Olkin, I. and Press, S. J. (1969). Testing and estimation for a circular stationary model, *Ann. Math. Statist.*, **40**, 1358-1373.

[8] Rao, J. N. K. (1973). Personal communication to T. W. Anderson.

[9] Szatrowski, T. H. (1976). Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and non-linear structure, *Technical Report* No. 107, Stanford University.

[10] Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution, *Ann. Math. Statist.*, **19**, 447-473.

[11] Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances and equality of covariances in a normal multivariate distribution, *Ann. Math. Statist.*, **17**, 257-281.