The Annals of Statistics
1977, Vol. 5, No. 4, 746–762

ASYMPTOTIC PROPERTIES OF MAXIMUM LIKELIHOOD ESTIMATES IN THE MIXED MODEL OF THE ANALYSIS OF VARIANCE

BY JOHN J. MILLER[1]

Rutgers University

We show that in the mixed model of the analysis of variance, there is a sequence of roots of the likelihood equations which is consistent, asymptotically normal, and efficient in the sense of attaining the Cramér–Rao lower bound for the covariance matrix. These results follow directly by an application of a general result of Weiss (1971, 1973) concerning maximum likelihood estimates. This problem differs from standard problems in that we do not have independent, identically distributed observations and that estimates of different parameters may require normalizing sequences of different orders of magnitude. We give some examples and comment briefly on likelihood ratio tests for these models.

1. Introduction. The estimation of the parameters in the mixed model of the analysis of variance is a problem of considerable interest to statisticians and many different methods of estimation have been proposed. The maximum likelihood method received little attention until recently because the complexity of the likelihood equations precluded their use in practical problems. The development of high speed computers has made feasible the solution of the likelihood equations. Therefore it is of interest to discuss the properties of the maximum likelihood estimates in the mixed model. Hartley and Rao (1967) proposed a computational algorithm for the solution of the likelihood equations and proved that under certain restrictions the estimates were consistent and asymptotically normal as the size of the experimental design increased. Anderson (1969, 1971) considered maximum likelihood estimates in a more general class of models (multivariate models where the covariance matrix has linear structure) and proposed a different method of solution; he proved that the estimates were consistent and asymptotically normal as the entire design was repeated. In this paper we consider asymptotic properties of the maximum likelihood estimates for a large class of design sequences whose size increases to infinity; this class of design sequences contains all sequences treated by Hartley and Rao and most sequences which could occur in practice. We take the basic model of Hartley and Rao, rewrite it in the form used by Anderson and prove consistency and asymptotic normality of the estimates in this model.

In order to obtain asymptotic results in the mixed model, the number of levels

Received April 1974; revised December 1976.

[1] Research partially supported by the Office of Naval Research Contract N00014-67-A-0112-0030.

AMS 1970 *subject classifications*. 62E20, 62J10.

Key words and phrases. Analysis of variance, mixed model, maximum likelihood estimates, consistency, asymptotic normality.

of each random factor must increase to infinity. One way this can be accomplished is by considering repetitions of a given experiment; in this case Anderson's results apply. More often in the analysis of variance a conceptual sequence of experiments with the number of levels of each of the random factors increasing to infinity is considered. Hartley and Rao treat such sequences. However, one of their assumptions is that the number of observations at any level of any factor must remain less than some fixed constant for all designs in the sequence. This assumption eliminates many crossed designs where the number of observations at a given level of one factor is proportional to the number of levels of another factor. We loosen the assumptions to allow such sequences.

Using this larger class of design sequences introduces several problems into consideration of asymptotic results. Results on maximum likelihood estimation with independent, identically distributed observations do not apply because, as in any sequence of mixed model designs, the observations in a particular design are not independent. A more important difficulty is the possibility that estimates of different parameters may require normalizing sequences which differ in order of magnitude. For example, if θ_1 and θ_2 are two parameters and if $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$ are their estimates, there may be no single function of n, $\nu(n)$, increasing to infinity such that $\nu^{\frac{1}{2}}(n)[(\hat{\theta}_{1n} - \theta_1), (\hat{\theta}_{2n} - \theta_2)]'$ converges in distribution to a bivariate normal distribution. It may be necessary to use two such sequences, $\nu_1(n)$ and $\nu_2(n)$, where $[\nu_1^{\frac{1}{2}}(n)(\hat{\theta}_{1n} - \theta_1), \nu_1^{\frac{1}{2}}(n)(\theta_{2n} - \theta_2)]'$ converges to a bivariate normal distribution but where $\nu_1(n)/\nu_2(n)$ converges to zero or infinity. Asymptotic results must be modified to allow for this possibility. A general theorem of Weiss (1971, 1973) on maximum likelihood estimates allows us to overcome both of these difficulties. We will be able to show that for a reasonable set of conditions for the design sequences, the assumptions required for the theorem of Weiss are satisfied. Then the consistency, asymptotic normality and efficiency will follow for the mixed model analysis of variance as a corollary to Weiss' theorem.

In Section 2 we discuss the basic analysis of variance model and assumptions about it and give the likelihood equations and Weiss' theorem. In Section 3 we give and explain the restrictions on the design sequences and we state and give an outline of the proof of Theorem 3.1, which yields the consistency and asymptotic normality of the maximum likelihood estimates. In Section 4 we give two simple examples of the application of asymptotic theory. In Section 5 we make some comments on the asymptotic efficiency of the maximum likelihood estimates and on likelihood ratio tests. Appendix A contains details from the proof of Theorem 3.1. Appendix B contains a sufficient condition on the design sequence to guarantee the positive definiteness of the matrix $\mathbf{C}_1$.

2. Basic results. The basic model we shall use in the mixed model analysis of variance is that given by Hartley and Rao (1967); it can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}_1\mathbf{b}_1 + \mathbf{U}_2\mathbf{b}_2 + \cdots + \mathbf{U}_{p_1}\mathbf{b}_{p_1} + \mathbf{e}\,, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times p_0$ matrix of known constants (the design matrix for the fixed effects); $\boldsymbol{\alpha}$ is a $p_0 \times 1$ vector of unknown constants; $\mathbf{U}_i$ is an $n \times m_i$ matrix of known constants (a design matrix for a random effect), $i = 1, 2, \cdots, p_1$; $\mathbf{b}_i$ is an $m_i \times 1$ random vector, $i = 1, 2, \cdots, p_1$; $\mathbf{e}$ is an $n \times 1$ random vector. Let $\mathbf{G}_i = \mathbf{U}_i\mathbf{U}_i'$, $i = 1, 2, \cdots, p_1$, and $\mathbf{G}_0 = \mathbf{I}_n$. The following assumptions are made about the model.

ASSUMPTION 2.1. The random vectors $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_{p_1}, \mathbf{e}$ are mutually independent, with $\mathbf{e} \sim \mathscr{N}_n(\mathbf{0}, \sigma_0\mathbf{I}_n)$ and $\mathbf{b}_i \sim \mathscr{N}_{m_i}(\mathbf{0}, \sigma_i\mathbf{I}_{m_i})$, $i = 1, 2, \cdots, p_1$.[2]

ASSUMPTION 2.2. The matrix $\mathbf{X}$ has full rank p_0.

ASSUMPTION 2.3. $n \geqq p_0 + p_1 + 1$.

ASSUMPTION 2.4. The partitioned matrix $[\mathbf{X}: \mathbf{U}_i]$ has rank greater than p_0, $i = 1, 2, \cdots, p_1$.

ASSUMPTION 2.5. The matrices $\mathbf{G}_0, \mathbf{G}_1, \cdots, \mathbf{G}_{p_1}$ are linearly independent; that is, $\sum_{i=0}^{p_1} \tau_i \mathbf{G}_i = \mathbf{0}$ implies $\tau_i = 0$, $i = 0, 1, \cdots, p_1$.

ASSUMPTION 2.6. The matrix $\mathbf{U}_i$ consists only of zeros and ones and there is exactly one 1 in each row and at least one 1 in each column, $i = 1, 2, \cdots, p_1$.

Note that Assumption 2.2 can always be satisfied by a suitable reparameterization of the problem. Assumption 2.4 requires that the fixed effects not be confounded with any of the random effects. Assumption 2.5 requires that the random effects not be confounded with each other. Assumption 2.6 states that the $\mathbf{U}_i$ are standard design matrices and it has three consequences: $\mathbf{U}_i'\mathbf{U}_i = \mathbf{D}_i$, an $m_i \times m_i$ nonsingular diagonal matrix; $\mathbf{U}_i$ has full rank m_i; and $m_i \leqq n$. Assumptions 2.1—2.5 are sufficient to guarantee estimability of the parameters. Assumption 2.6 is added for convenience.

It follows that $\mathbf{y} \sim \mathscr{N}_n(\mathbf{X}\boldsymbol{\alpha}, \boldsymbol{\Sigma}(\boldsymbol{\sigma}))$ where $\boldsymbol{\Sigma}(\boldsymbol{\sigma}) = \sum_{i=0}^{p_1} \sigma_i \mathbf{G}_i$. The objective is to observe $\mathbf{y}$ and estimate $\boldsymbol{\alpha}, \sigma_0, \sigma_1, \cdots, \sigma_{p_1}$ by the method of maximum likelihood.

The parameter space is defined as follows: Let $p \equiv p_0 + p_1 + 1$ and let $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \cdots, \sigma_{p_1})'$. Then $\Theta \subset R^p$ is the parameter space, where

$$\Theta = \{\boldsymbol{\theta} \in R^p \mid \boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\sigma}'); \boldsymbol{\alpha} \in R^{p_0}; \sigma_0 > 0; \sigma_i \geqq 0, i = 1, 2, \cdots, p_1\} . \tag{2}$$

The vector $\boldsymbol{\theta}$ may be represented by its components θ_i, by its partitioned forms $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ and their components α_j and σ_i, or by mixed expressions (for example, $\partial\lambda(\mathbf{y}, \boldsymbol{\theta})/\partial\theta_i$, $|\hat{\sigma}_i - \sigma_i|$, or $\partial\lambda(\mathbf{y}, \boldsymbol{\theta})/\partial\boldsymbol{\alpha}$). The log-likelihood function $\lambda(\mathbf{y}, \boldsymbol{\theta})$ is given by

$$\lambda(\mathbf{y}, \boldsymbol{\theta}) = -\tfrac{1}{2}n \log 2\pi - \tfrac{1}{2} \log |\boldsymbol{\Sigma}| - \tfrac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) . \tag{3}$$

Anderson (1970) proved that in a less restricted model (no restrictions on the

[2] This differs from the usual convention of using σ_i^2. σ_i is used as a *variance* to simplify notation. This also follows the notation of Anderson (1969, 1971, 1973).

σ_i) $\hat{\boldsymbol{\theta}}$ may be calculated by solving the likelihood equations, which are

$$\begin{aligned} (4) \qquad & [\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}]\boldsymbol{\alpha} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}\,, \\ & \operatorname{tr} \boldsymbol{\Sigma}^{-1}\mathbf{G}_i = (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\,, \qquad i = 0, 1, \cdots, p_1\,, \end{aligned}$$

and $\boldsymbol{\Sigma}$ is taken as a function of $\boldsymbol{\sigma}$. If there is more than one solution to (4), $\hat{\boldsymbol{\theta}}$ is taken to be that solution which maximizes $\lambda(\mathbf{y}, \boldsymbol{\theta})$. In the present model it is necessary to insure that $\hat{\boldsymbol{\theta}}$ belongs to the parameter space. If any solutions of the above equations have negative estimates of a variance component, it is necessary to obtain solutions along the boundaries of the parameter space and compare their values to obtain $\hat{\boldsymbol{\theta}}$. It should be noted that this poses no essential problem because the restricted or reduced model (with one or more σ_i set to zero) is another model of the same form. Therefore the techniques of solution of the likelihood equations may be used in the reduced model. $\hat{\boldsymbol{\theta}}$ is that estimate which is a solution of the full or any reduced likelihood equations which maximizes $\lambda(\mathbf{y}, \boldsymbol{\theta})$. Numerical techniques for solution of the likelihood equations have been discussed by Hartley and Rao (1967), Hartley and Vaughn (1972), Anderson (1973) and Miller (1973); these techniques will not be discussed here. The asymptotic results proved here are not affected by the "truncation" of the variance estimates since the asymptotic results concern roots of the likelihood equation when the true parameter point is an interior point of the parameter space. The consistency of these roots means that with high probability no truncation will be required.

Weiss (1971, 1973) gave a very general theorem on asymptotic properties of maximum likelihood estimates which we can apply to our sequence of mixed model analysis of variance designs. We will paraphrase Weiss' statement of the theorem to fit our needs and notation. $\mathbf{Y}_n$ will be the observations from a design in our sequence of designs and $\boldsymbol{\theta}$ will be as defined by (2).

THEOREM 2.1 (Weiss (1971, 1973)). *Consider a sequence of random variables* $\mathbf{Y}_n$ *with density* $L_n(\mathbf{Y}_n, \boldsymbol{\theta})$ *where* $\boldsymbol{\theta} \in \Theta$, $\boldsymbol{\theta}$ *is* $p \times 1$. *Let* $\lambda(\mathbf{Y}_n, \boldsymbol{\theta}) = \log L_n(\mathbf{Y}_n, \boldsymbol{\theta})$. *Suppose* $\boldsymbol{\theta}_0$, *the true parameter point, is an interior point of* Θ *and let there be* $2p$ *sequences of nonrandom positive quantities* $n_i(n)$ *and* $q_i(n)$, $i = 1, 2, \cdots, p$ *such that* $\lim_{n\to\infty} n_i(n) = \infty$, $\lim_{n\to\infty} q_i(n) = \infty$, $\lim_{n\to\infty} [q_i(n)/n_i(n)] = 0$, $i = 1, 2, \cdots, p$. [$q_i(n)$ *may depend on* $\boldsymbol{\theta}_0$.] *Further assume that there exist nonrandom quantities* $J_{ij}(\boldsymbol{\theta}_0)$ *such that* $-[1/n_i(n)n_j(n)][\partial^2\lambda(\mathbf{Y}_n, \boldsymbol{\theta})/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0}]$ *converges stochastically to* $J_{ij}(\boldsymbol{\theta}_0)$ *as* $n \to \infty$, $i, j = 1, 2, \cdots, p$. $\mathbf{J}(\boldsymbol{\theta}_0)$ *is assumed to be a continuous function of* $\boldsymbol{\theta}_0$ *and to be positive definite. Now let* $N_n(\boldsymbol{\theta}_0)$ *denote the set of all vectors* $\boldsymbol{\theta}$ *such that* $|\theta_i - \theta_{0i}| \leqq q_i(n)/n_i(n)$, $i = 1, 2, \cdots, p$. *Denote* $-[1/n_i(n)n_j(n)][\partial^2\lambda(\mathbf{Y}_n, \boldsymbol{\theta})/\partial\theta_i\,\partial\theta_j] - J_{ij}(\boldsymbol{\theta}_0)$ *by* $\varepsilon_{ij}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, n)$. *For any given positive value* γ *let* $R_n(\boldsymbol{\theta}_0, \gamma)$ *denote the region in* $\mathbf{Y}_n$ *space where* $\sum_{i=1}^{p}\sum_{j=1}^{p} q_i(n)q_j(n)\sup_{\boldsymbol{\theta}\in N_n(\boldsymbol{\theta}_0)} |\varepsilon_{ij}(\boldsymbol{\theta}, \boldsymbol{\theta}_0, n)| < \gamma$. *Assume there exist sequences* $\{\gamma(n, \boldsymbol{\theta}_0)\}$, $\{\delta(n, \boldsymbol{\theta}_0)\}$ *of nonrandom positive quantities with* $\lim_{n\to\infty} \gamma(n, \boldsymbol{\theta}_0) = 0$, $\lim_{n\to\infty} \delta(n, \boldsymbol{\theta}_0) = 0$ *such that for each* n $P_{\boldsymbol{\theta}}\{R_n[\boldsymbol{\theta}_0, \gamma(n, \boldsymbol{\theta}_0)]\} > 1 - \delta(n, \boldsymbol{\theta}_0)$ *for all* $\boldsymbol{\theta} \in N_n(\boldsymbol{\theta}_0)$. *It then follows that there exists a sequence of estimates of* $\hat{\boldsymbol{\theta}}(n)$ (*which are roots of the equations* $\partial\lambda(\mathbf{Y}_n, \theta)/\partial\theta_i = 0$,

$i = 1, 2, \cdots, p$) *such that the vector whose* ith *component is* $n_i(n)[\hat{\theta}_i(n) - \theta_{0i}]$ *converges in distribution to a normal random vector with mean vector* $\mathbf{0}$ *and covariance matrix* $\mathbf{J}^{-1}(\boldsymbol{\theta}_0)$. *That is, the sequence* $\hat{\boldsymbol{\theta}}(n)$ *is consistent, asymptotically normal, and efficient.*

Now all we must do is prove that the conditions of this theorem are met, which we do in the next section. (Note that although n will approach infinity, it will not generally behave as an index $n = 1, 2, 3, \cdots$ but may move in increasing jumps; this, of course, does not affect the asymptotic results.)

3. Main result.

In this section the assumptions used to carry out the asymptotic theory will be stated and briefly explained and the main result of this paper will be stated and an outline of its proof given. Consider a sequence of experiments each following the model (1). An experiment in this sequence may be an extension of previous experiments or an entirely different design. However, all such sequences must have the following properties.

ASSUMPTION 3.1. n and each m_i, $i = 1, 2, \cdots, p_1$, tend to infinity; each m_i can be considered a function of n. (Note that all matrices and vectors in the experiment now should properly be denoted to depend on n; that is we should write $\mathbf{y}_n$, $\mathbf{X}_n$, $\mathbf{U}_{i(n)}$, etc. For convenience, these dependencies on n will not be explicitly carried in the notation.)

ASSUMPTION 3.2. Let $m_0 \equiv n$; then for each $i, j = 0, 1, \cdots, p_1$, either $\lim_{n\to\infty} m_i/m_j \equiv \rho_{ij}$ or $\lim_{n\to\infty} m_j/m_i \equiv \rho_{ji}$ exists. (If $\rho_{ij} = 0$, then let $\rho_{ji} = \infty$ for notational convenience.)

Now without loss of generality, let the $\mathbf{U}_i$ be labeled so that for $i < j$, $\rho_{ij} > 0$; i.e., the m_i are in decreasing order of magnitude. Generate a partition of the integers $\{0, 1, \cdots, p_1\}$, $S_0, S_1, \cdots, S_c$, so that for indices i in the same set S_s, the associated m_i's have the same order of magnitude. Such a partition is generated as follows:

(5)

i) $i_0 \equiv 0\,; \quad S_0 \equiv \{0\}\,; \quad i_1 \equiv 1\,.$

ii) For $s = 1, 2, \cdots,$ it is true that $i_s \in S_s$. Then for $i = i_s + 1$, $i_s + 2, \cdots,$ include i in S_s until $\rho_{i_s, i} = \infty$; call the first value of i where this occurs i_{s+1}; then $i_{s+1} \in S_{s+1}$.

iii) Continue as in step ii until p_1 has been placed in a set. Call this set S_c.

There are then $c + 1$ sets in the partitions, $S_0, S_1, \cdots, S_c$, and $S_s = \{i_s, \cdots, i_{s+1} - 1\}$ (where $i_{c+1} \equiv p_1 + 1$ to insure S_c is correct).

For each $i = 1, 2, \cdots, p_1$, $i \in S_s$ for some $s = 1, 2, \cdots, c$. Define sequences ν_i (depending on n) as follows:

$$\begin{aligned} \nu_i &\equiv \operatorname{rank}[\mathbf{U}_{i_s} : \mathbf{U}_{i_s+1} : \cdots : \mathbf{U}_{p_1}] \\ &\qquad - \operatorname{rank}[\mathbf{U}_{i_s} : \cdots : \mathbf{U}_{i-1} : \mathbf{U}_{i+1} : \cdots : \mathbf{U}_{p_1}]\,, \qquad i = 1, 2, \cdots, p_1\,, \\ \nu_0 &\equiv n - \operatorname{rank}[\mathbf{U}_1 : \cdots : \mathbf{U}_{p_1}]\,. \end{aligned} \tag{6}$$

(The ν_i so defined are closely related to the degrees of freedom of sums of squares in the analysis of variance.)

Assumption 3.3. Let $r_i \equiv \lim_{n\to\infty} \nu_i/m_i$, $i = 0, 1, \cdots, p_1$: then each of the r_i exists and is positive.

Now let $\boldsymbol{\theta}_0' = (\boldsymbol{\alpha}_0', \boldsymbol{\sigma}_0')$ be the true parameter point, where $\boldsymbol{\sigma}_0 = (\sigma_{00}, \sigma_{01}, \cdots, \sigma_{0p_1})'$. Let $\boldsymbol{\Sigma}_0 \equiv \sum_{j=0}^{p_1} \sigma_{0j}\mathbf{G}_j$ be the true covariance matrix.

Assumption 3.4. There exists a sequence ν_{p_1+1} (depending on n) increasing to infinity such that the $p_0 \times p_0$ matrix $\mathbf{C}_0$ defined by

$$\mathbf{C}_0 = \lim_{n\to\infty} [\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X}]/\nu_{p_1+1} \tag{7}$$

exists and is positive definite.

Define the $(p_1 + 1) \times (p_1 + 1)$ matrix $\mathbf{C}_1$ by

$$[\mathbf{C}_1]_{ij} = \tfrac{1}{2} \lim_{n\to\infty} [\operatorname{tr} \boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i \boldsymbol{\Sigma}_0^{-1}\mathbf{G}_j]/\nu_i^{\frac{1}{2}}\nu_j^{\frac{1}{2}}\,, \qquad i, j = 0, 1, \cdots, p_1\,. \tag{8}$$

Assumption 3.5. Each of the limits used in defining $[\mathbf{C}_1]_{ij}$ in (8) exists, $i, j = 0, 1, \cdots, p_1$. The matrix $\mathbf{C}_1$ is positive definite.

The object of these assumptions is to rule out certain sequences of experiments for which the limiting distributions either degenerate or "blow up." For example, asymptotic theory requires an expanding sequence of experiments, which is what Assumption 3.1 requires. Assumption 3.2 requires that the expansion should be orderly—sizes of various parts of the design should relate to each other in an orderly way.

The remaining assumptions require that the sequence not be a degenerate one. The ν_i defined by (6) is the dimension of the part of the linear space spanned by the columns of $\mathbf{U}_i$ which is orthogonal to the space spanned by the columns of the other $\mathbf{U}_j$ where $i_s \leqq j \leqq p_1$, $j \neq i$, and $i \in S_s$. (i_s and S_s are defined by (5).) Thus ν_i is the dimension of the part of $\mathbf{U}_i$ not dependent on the other $\mathbf{U}_j$. Assumption 3.3 says that this part remains an integral part of $\mathbf{U}_i$; it does not get overwhelmed by the other columns of $\mathbf{U}_i$. It could be said that this assumption requires that the ith effect not be "asymptotically confounded" with the effects associated with the other $\mathbf{U}_j$ mentioned above. This assumption implies that ν_i and m_i are of the same order of magnitude and hence that $\nu_i \to \infty$, $i = 0, 1, \cdots, p_1$ by Assumption 3.1.

The matrices $\mathbf{C}_0$ and $\mathbf{C}_1$ defined by (7) and (8) determine the asymptotic covariance matrix of the estimates of the fixed and random effects respectively. Assumptions 3.4 and 3.5 insure the existence and positive definiteness of these matrices. If either $\mathbf{C}_0$ or $\mathbf{C}_1$ does not exist or is not positive definite then its associated estimates do not converge to a nondegenerate distribution. It should be noted that Assumption 3.5 states that the limits given by (8) exist; it is easily shown from Assumptions 3.1—3.3 that the lim inf and lim sup exist so that the assumption only requires the additional fact that the lim inf equal the lim sup. Appendix B contains conditions on the design sequence sufficient to guarantee positive definiteness of $\mathbf{C}_1$. It appears that any design or set of designs that might be used in practice can be imbedded in a sequence satisfying Assumptions 3.1—3.5. This is of some importance because asymptotic optimality properties are usually cited as one justification for the use of maximum likelihood estimates.

Theorem 3.1 is the main result of this paper. It states that under the conditions given above the maximum likelihood estimates are consistent and asymptotically normal.

THEOREM 3.1. *Consider a sequence of experiments each described by the model* (1) *and each satisfying Assumptions* 2.1—2.6. *Suppose the sequence satisfies Assumptions* 3.1—3.5. *Let the parameter space* Θ *be given by* (2) *and the* log-*likelihood function* $\lambda(\mathbf{y}, \boldsymbol{\theta})$ *be given by* (3). *Suppose that the true parameter point* $\boldsymbol{\theta}_0$ *is an interior point of* Θ; (i.e., $\sigma_{0i} > 0$, $i = 0, 1, \cdots, p_1$). *Define the* $p \times p$ *matrix* $\mathbf{J}$ *by* $\mathbf{J} = \left[\begin{smallmatrix} \mathbf{C}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_1 \end{smallmatrix}\right]$, *where* $\mathbf{C}_0$ *and* $\mathbf{C}_1$ *are defined by* (7) *and* (8) *respectively. It follows that there exist sequences* n_i, $i = 0, 1, \cdots, p_1 + 1$ (*depending on* n) *increasing to infinity and a sequence of estimates of* $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n(\mathbf{y}) \equiv [\hat{\boldsymbol{\alpha}}_n'(\mathbf{y}), \hat{\boldsymbol{\sigma}}_n'(\mathbf{y})]'$ *with the following properties.*

i) *Given* $\varepsilon > 0$ *there exists* $b = b(\varepsilon)$ *such that* $0 < b < \infty$ *and* $n_0 = n_0(\varepsilon)$ *such that for all* $n > n_0$

$$P\left\{\frac{\partial\lambda(\mathbf{y}, \boldsymbol{\theta})}{\partial\theta_i}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n(\mathbf{y})} = 0,\ i = 1, 2, \cdots, p;\right.$$
$$|[\hat{\boldsymbol{\alpha}}_n(\mathbf{y})]_j - \alpha_{0j}| < \frac{b}{n_{p_1+1}},\ j = 1, 2, \cdots, p_0;$$
$$\left.|[\hat{\boldsymbol{\sigma}}_n(\mathbf{y})]_i - \sigma_{0i}| < \frac{b}{n_i},\ i = 0, 1, \cdots, p_1\right\} \geqq 1 - \varepsilon\,.$$

ii) *The* $p \times 1$ *vector whose first* p_0 *components are* $n_{p_1+1}\{\hat{\boldsymbol{\alpha}}_n(\mathbf{y}) - \boldsymbol{\alpha}_0\}$ *and whose* $(p_0 + i + 1)$th *component is* $n_i\{[\hat{\boldsymbol{\sigma}}_n(\mathbf{y})]_i - \sigma_{0i}\}$, $i = 0, 1, \cdots, p_1$, *converges in distribution to a* $\mathscr{N}_p(\mathbf{0}, \mathbf{J}^{-1})$ *random variable.*

PROOF. To prove this theorem we need only prove that the conditions of Theorem 2.1 hold from which the conclusion of this theorem will follow immediately. We begin by defining the sequences n_i by using the ν_i as defined by (6) and (7):

$$n_i(n) = [\nu_i(n)]^{\frac{1}{2}}\,, \qquad i = 0, 1, \cdots, p_1 + 1\,. \tag{9}$$

(For the remainder of the proof all notation of dependence on n will be suppressed unless that dependence is to be emphasized.) Then for $\lambda(\mathbf{y}, \boldsymbol{\theta})$ defined by (3) we observe that the derivatives of λ with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\sigma}$ are given below. (All indices i and j run from 0 to p_1.)

$$\partial\lambda/\partial\boldsymbol{\alpha} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\,, \tag{10}$$

$$\partial\lambda/\partial\sigma_i = [-\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{G}_i + (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})]/2\,, \tag{11}$$

$$\partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}' = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\,, \tag{12}$$

$$\partial^2\lambda/\partial\sigma_i\,\partial\boldsymbol{\alpha} = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})\,, \tag{13}$$

$$\partial^2\lambda/\partial\sigma_i\,\partial\sigma_j = [\operatorname{tr}\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{G}_j - 2(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\boldsymbol{\Sigma}^{-1}\mathbf{G}_i\boldsymbol{\Sigma}^{-1}\mathbf{G}_j\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})]/2\,, \tag{14}$$

which are respectively a $p_0 \times 1$ vector, a scalar, a $p_0 \times p_0$ matrix, a $p_0 \times 1$ vector and a scalar.

Observe that (12)—(14) with (7) and (8) show that the matrix $\mathbf{J}$ in Theorem 3.1 is indeed the $\mathbf{J}(\boldsymbol{\theta}_0)$ required for Theorem 2.1 in the sense that (i) $-\mathscr{E}_{\boldsymbol{\theta}_0}(\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0})/(n_i n_j) \to [\mathbf{J}]_{ij}$, (ii) $\mathbf{J}(\boldsymbol{\theta}_0)$ is positive definite, and (iii) $\mathbf{J}(\boldsymbol{\theta}_0)$ is continuous in $\boldsymbol{\theta}_0$. (Requirements (i) and (ii) are true by Assumptions 3.4 and 3.5 along with (7) and (8); (iii) can be shown to be true by arguments similar to those used in the subsequent proof.) Now define for each n $\kappa = \max_{i,j} |-[\mathscr{E}_{\boldsymbol{\theta}_0}(\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0})]/[n_i n_j] - [\mathbf{J}]_{ij}|$ and set

$$q_i = q = \min(n_0^{\frac{1}{4}}, n_1^{\frac{1}{4}}, \cdots, n_p^{\frac{1}{4}}, n_{p_1+1}^{\frac{1}{4}}, \kappa^{-\frac{1}{4}}), \quad i = 0, 1, \cdots, p_1 + 1. \tag{15}$$

(*Note.* It is convenient to have q_i equal for all i and violates no requirement of Theorem 2.1.) It follows that $q_i \to \infty$ because $n_i \to \infty$ by Assumptions 3.1 and 3.3 and $\kappa \to 0$ by Assumptions 3.4 and 3.5; obviously $q_i/n_i \to 0$ for all i.

To prove the conditions of Theorem 2.1 it then suffices to prove that

$$|-(\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0})/(n_i n_j) - [\mathbf{J}]_{ij}| \to_{P_{\boldsymbol{\theta}_0}} 0 \qquad \text{and that}$$

$$(q_i q_j)\sup_{\boldsymbol{\theta}_1 \in N_n(\boldsymbol{\theta}_0)} |-(\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_1})/(n_i n_j) - [\mathbf{J}]_{ij}| \to_{P_{\boldsymbol{\theta}_2}} 0 \qquad \text{for all} \quad \boldsymbol{\theta}_2 \in N_n(\boldsymbol{\theta}_0) \quad \text{and}$$

all i, j where by $T_n \to_{P_{\theta^*}} 0$ for all $\theta^* \in N_n(\boldsymbol{\theta}_0)$ we mean that for any fixed $\varepsilon > 0$, $\delta > 0$ there exists $n_0(\varepsilon, \delta)$ such that for all $n > n_0$ and all $\boldsymbol{\theta}^* \in N_n(\boldsymbol{\theta}_0)$, $P_{\boldsymbol{\theta}^*}\{|T_n| > \varepsilon\} < \delta$. We shall prove the second requirement first and in the process shall prove the first requirement. We note that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}_1 \in N_n(\boldsymbol{\theta}_0)} \left(\frac{-1}{n_i n_j}\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_1} - [\mathbf{J}]_{ij}\right) \\
&\quad = \sup_{\boldsymbol{\theta}_1 \in N_n(\boldsymbol{\theta}_0)} \left\{\frac{-1}{n_i n_j}\left[\left(\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_1}\right) - \left(\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_2}\right)\right]\right\} \\
&\qquad - \frac{1}{n_i n_j}\left(\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_2} - \mathscr{E}_{\boldsymbol{\theta}_2}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_2}\right]\right) \\
&\qquad - \frac{1}{n_i n_j}\left(\mathscr{E}_{\boldsymbol{\theta}_2}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_2}\right] - \mathscr{E}_{\boldsymbol{\theta}_2}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right]\right) \\
&\qquad - \frac{1}{n_i n_j}\left(\mathscr{E}_{\boldsymbol{\theta}_2}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right] - \mathscr{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right]\right) \\
&\qquad + \left(\frac{-1}{n_i n_j}\mathscr{E}_{\boldsymbol{\theta}_0}\left[\frac{\partial^2\lambda}{\partial\theta_i\,\partial\theta_j}\bigg|_{\boldsymbol{\theta}_0}\right] - [\mathbf{J}]_{ij}\right).
\end{aligned} \tag{16}$$

We may then bound $q_i q_j$ times each term of the right-hand side of (16) separately. Denote the five terms as $\phi_1 - \phi_5$ respectively. ϕ_3, ϕ_4 and ϕ_5 are non-stochastic and will only involve limiting arguments. For ϕ_2 it suffices to prove $q_i^2 q_j^2 \operatorname{Var}_{\boldsymbol{\theta}_2}(\phi_2) \to 0$. For ϕ_1 we appeal to Lemma A.4. We will give an example of the proof of convergence to zero of each term. Then we will show how to assemble all these proofs into a proof of the theorem. The remaining details are very similar to those given in Miller (1973).

For the remainder of this proof we shall represent each q_i as q and will call $n_{p_1} \equiv n_f$, (f for fixed). Suppose Conditions A.1 and A.2 are true. Consider $q^2|\phi_5| \leqq \kappa^{-\frac{1}{2}}\kappa = \kappa^{\frac{1}{2}} \to 0$ by (15), (7) and (8). For ϕ_4 take the $\partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\sigma_i$ form; from (13) it suffices to bound for any $p_0 \times 1$ $\boldsymbol{\xi}$ such that $\boldsymbol{\xi}'\boldsymbol{\xi} = 1$ the quantity $\psi_4 \equiv [q^2/(n_i n_f)]\boldsymbol{\xi}'\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{X}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)$. (A subscript $a = 0, 1, 2$ on $\boldsymbol{\alpha}$ or $\boldsymbol{\Sigma}$ means it is formed from the appropriate $\boldsymbol{\theta}_a' = (\boldsymbol{\alpha}_a', \boldsymbol{\sigma}_a')$ where $\boldsymbol{\sigma}_a' = (\sigma_{a0}, \sigma_{a1}, \cdots, \sigma_{ap_1})$, and $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ are any elements of $N_n(\boldsymbol{\theta}_0)$.) Now

$$\begin{aligned}
\psi_4^2 &\leqq [q^4/(n_i^2 n_f^2)][\boldsymbol{\xi}'\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{X}\boldsymbol{\xi}][(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)] \\
&\leqq [q^4/(n_i^2 n_f^2)][\boldsymbol{\xi}'\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X}\boldsymbol{\xi}][\lambda_{\max}(\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i)^2][\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)][\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})] \\
&\leqq [q^4/(n_i^2 n_f^2)]\boldsymbol{\xi}'\boldsymbol{\xi}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)[\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})]^2[\lambda_{\max}(\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i)^2] \\
&= [q^4 n_f^2/n_i^2]\boldsymbol{\xi}'\boldsymbol{\xi}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)[\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})/n_f^2]^2[\lambda_{\max}(\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i)^2]
\end{aligned}$$

where the first inequality follows from an application of the Cauchy–Schwarz inequality and the next two by definition of characteristic root. But $\boldsymbol{\xi}'\boldsymbol{\xi} = 1$; $(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0) \leqq p_0 q^2/n_f^2$ by Lemma A.2.i; $\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})/n_f^2$ is bounded because the matrix converges to the constant matrix $\mathbf{C}_0$; the last term is bounded by $1/\sigma_{0i}^2$ by A.2.iii and A.2.iv. Thus ψ_4^2 is bounded by a constant times q^6/n_i^2 which converges to zero by definition of q.

For ϕ_3 consider the $\partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\boldsymbol{\alpha}$ term; from (12) it suffices to bound for any $p_0 \times 1$ vectors $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ such that $\boldsymbol{\xi}_1'\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2'\boldsymbol{\xi}_2 = 1$ the quantity $\psi_3 = [q^2/n_f^2]\boldsymbol{\xi}_1'\mathbf{X}'(\boldsymbol{\Sigma}_2^{-1} - \boldsymbol{\Sigma}_0^{-1})\mathbf{X}\boldsymbol{\xi}_2$. As above we may bound $\psi_3^2 \leqq q^4\boldsymbol{\xi}_1'\boldsymbol{\xi}_1\boldsymbol{\xi}_2'\boldsymbol{\xi}_2[\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})/n_f^2]^2[\max|\lambda_k(\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_2))|]^2$. The last term is bounded by $4q^2/\min(n_i\sigma_{0i})^2$ via A.2.v; thus ψ_3^2 is bounded by a constant times $q^6/\min(n_i\sigma_{0i})^2 \to 0$ by definition of q.

Now consider the $\partial^2\lambda/\partial\sigma_i\,\partial\sigma_j$ term for ϕ_2. It suffices to prove that $\mathrm{Var}_{\boldsymbol{\theta}_2}\psi_4 \to 0$ where $\psi_4 \equiv [q^2/(n_i n_j)][\mathrm{tr}\,\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_j - 2(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)'\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_j\boldsymbol{\Sigma}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)]/2$. Then using rules for variances of quadratic forms we find that $\mathrm{Var}_{\boldsymbol{\theta}_2}(\psi_4) = 2[q^4/(n_i^2 n_j^2)]\,\mathrm{tr}\,(\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_j\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\Sigma}_2)^2 \leqq 2q^4[\min(m_i, m_j)/n_i^2]\lambda_{\max}^2[\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_j]/n_j^2$ because there are at most $\min(m_i, m_j)$ nonzero characteristic roots of $(\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}G_j)^2$ each of which is bounded by $\lambda_{\max}^2[\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_j]$. This term is in turn bounded by $16/(\sigma_{0i}^2\sigma_{0j}^2)$ via Lemma A.2.iv. $\min(m_i, m_j)/n_i^2$ is bounded by definition of the $n_i^2 = \nu_i$ in (6) and Assumption 3.3. Thus $\mathrm{Var}_{\boldsymbol{\theta}_2}(\psi_4)$ is bounded by a constant times $q^4/n_j^2 \to 0$. Note that proving $\mathrm{Var}_{\boldsymbol{\theta}_2}(q^2\phi_2) \to 0$ will also prove the first condition of Theorem 2.1. $[1/(n_i n_j)](-\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0}) - [\mathbf{J}]_{ij} = [1/(n_i n_j)][-\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0} - \mathscr{E}_{\boldsymbol{\theta}_0}(-\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0})] + [(1/[n_i n_j])\mathscr{E}_{\boldsymbol{\theta}_0}(-\partial^2\lambda/\partial\theta_i\,\partial\theta_j|_{\boldsymbol{\theta}_0}) - [\mathbf{J}]_{ij}]$. The second term is bounded by κ and the variance of the first under $\boldsymbol{\theta}_0$ is covered above; $\boldsymbol{\theta}_0 \in N_n(\boldsymbol{\theta}_0)$ and leaving out q^2 will only increase the convergence to zero.

The convergence of $q^2\phi_1$ to zero is the subject of Lemma A.4. Now we assemble all these steps together. Given $\boldsymbol{\theta}_2 \in N_n(\boldsymbol{\theta}_0)$, $\varepsilon > 0$ and $\delta > 0$ we wish to find n_0 such that for all $n > n_0$ the probability under $\boldsymbol{\theta}_2$ that q^2 times the absolute value of the left-hand side of (16) is greater than δ is less than ε. First choose n_1 such that for all $n \geqq n_1$, $P_{\boldsymbol{\theta}_2}$ (Conditions A.1 and A.2 are false) $< \varepsilon/2$, which is possible by Lemma A.1. Now choose $n_2 \geqq n_1$ such that for $n \geqq n_2$, $P_{\theta_2}\{q^2|\phi_2| > \delta/5\} < \varepsilon/2$, which is possible because $q^2|\phi_2| \rightarrow_{P_{\theta_2}} 0$. Now choose $n_3 \geqq n_2$ such that for $n > n_3$, $q^2|\phi_3|$, $q^2|\phi_4|$ and $q^2|\phi_5|$ are all less than $\delta/5$, which can be done by definition of limits. Finally choose $n_0 \geqq n_3$ such that for $n > n_0$, $q^2|\phi_1| < \delta/5$ when Conditions A.1 and A.2 are true, which can be done by Lemma A.4. Then for $n > n_0$ $P_{\theta_2}\{q^2|\text{LHS (16)}| > \delta\} < \varepsilon$. This proves the final condition of Theorem 2.1 and hence proves Theorem 3.1.

4. Two simple examples. To illustrate the asymptotic properties proved in Theorem 3.1 we first take the simplest possible case, the one-way balanced random effects model. $y_{ij} = \mu + b_i + e_{ij}$, $j = 1, 2, \cdots, J$, $i = 1, 2, \cdots, I$, where y_{ij} is the observation, μ is the unknown mean, the b_i are independent identically distributed as $\mathscr{N}(0, \sigma_1)$, the e_{ij} are independent identically distributed as $\mathscr{N}(0, \sigma_0)$ and the b_i and e_{ij} are independent. This may be written in matrix form as $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{U}_1\mathbf{b}_1 + \mathbf{e}$, where $\mathbf{y}$ is an $IJ \times 1$ vector of observations, $\mathbf{X}$ is an $IJ \times 1$ vector of ones, α is an unknown constant, $\mathbf{U}_1$ is an $IJ \times I$ standard design matrix for this model, $\mathbf{b}_1$ is an $I \times 1$ vector of random effects and $\mathbf{e}$ is an $IJ \times 1$ vector of random errors. In this case the likelihood equations can be explicitly solved to yield the following maximum likelihood estimates. (We use standard analysis of variance notation: $y_{\bullet\bullet} = (\sum_i \sum_j y_{ij})/IJ$, $y_{i\bullet} = (\sum_j y_{ij})/J$, $SS_1 = J \sum_i (y_{i\bullet} - y_{\bullet\bullet})^2$, $SS_0 = \sum_i \sum_j (y_{ij} - y_{i\bullet})^2$, $MS_0 = SS_0/I(J-1)$.)

$$\begin{aligned}
\hat\alpha &= y_{\bullet\bullet}\,, \\
\hat\sigma_1 &= (SS_1/I - MS_0)/J && \text{when} \quad SS_1/I > MS_0\,, \\
&= 0 && \text{otherwise}, \\
\hat\sigma_0 &= MS_0 && \text{when} \quad SS_1/I > MS_0\,, \\
&= (SS_0 + SS_1)/IJ && \text{otherwise}.
\end{aligned}$$

Now $y_{\bullet\bullet}$, SS_0, and SS_1 are independent and distributed as $\mathscr{N}[\alpha, (\sigma_0 + J\sigma_1)/(IJ)]$, $\sigma_0\chi^2_{I(J-1)}$ and $(\sigma_0 + J\sigma_1)\chi^2_{(I-1)}$ respectively. Thus the means, variances and covariances of $\hat\alpha$, $\hat\sigma_1$, and $\hat\sigma_0$ can be calculated. It is easily seen that only $\hat\alpha$ is unbiased. To consider asymptotic properties it is sufficient that $I \rightarrow \infty$. However, it is of some interest to observe the behavior of the estimates as I and J each increase to infinity. (It is not necessary that I and J be of the same order of magnitude.) If $\sigma_0 > 0$ and $\sigma_1 > 0$, truncation will be needed with probability approaching zero so that we find that as $I, J \rightarrow \infty$, $\mathscr{E}(\hat\alpha) = \alpha$; $\mathscr{E}(\hat\sigma_1) \doteq (1 - 1/I)\sigma_1 - \sigma_0/IJ$; $\mathscr{E}(\hat\sigma_0) \doteq \sigma_0$; $\operatorname{Var}(\hat\alpha) = \sigma_0/IJ + \sigma_1/I$; $\operatorname{Var}(\hat\sigma_1) \doteq 2\sigma_0^2(IJ - J + 1)/I^2J^2(J-1) + 4\sigma_0\sigma_1(I-1)/I^2J + 2\sigma_1^2(I-1)/I^2$; $\operatorname{Var}(\hat\sigma_0) \doteq 2\sigma_0^2/I(J-1)$; $\operatorname{Cov}(\hat\alpha, \hat\sigma_0) = 0$; $\operatorname{Cov}(\hat\sigma_1, \hat\sigma_0) \doteq -2\sigma_0^2/IJ(J-1)$. Thus the estimates are consistent because each expected value converges to the true value and each variance converges to zero. However, a joint asymptotic normal distribution will not be obtained when each estimate is normalized by $I^{\frac12}J^{\frac12}$. The normalized variances of $\hat\alpha$ and $\hat\sigma_1$ do not converge to finite values. The correct normalizing sequences for $\hat\alpha$, $\hat\sigma_1$, and $\hat\sigma_0$ are $I^{\frac12}$, $I^{\frac12}$ and $I^{\frac12}J^{\frac12}$ respectively, in which case the asymptotic covariance matrix is $\mathbf{J} = \text{diag}(\sigma_1, 2\sigma_1^2, 2\sigma_0^2)$. This situation illustrates the need for normalizing sequences of different orders of magnitude. The example may be criticized by pointing out that if J does not become infinite normalization of all estimates by $I^{\frac{1}{2}}J^{\frac{1}{2}}$ does lead to an asymptotic normal distribution. However, there are many examples where normalizing sequences of different orders of magnitude cannot be avoided, for instance, any crossed model that is at least partially balanced.

Now consider the two-way balanced random effects model. $y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}$, $i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, $k = 1, 2, \cdots, K$, where the observed value y_{ijk} is the sum of μ, the unknown mean and a_i, b_j, c_{ij} and e_{ijk} all of which are independently normally distributed with mean zero and variances σ_2, σ_3, σ_1 and σ_0 respectively. This may be written as $\mathbf{y} = \mathbf{X}\alpha + \mathbf{U}_1\mathbf{b}_1 + \mathbf{U}_2\mathbf{b}_2 + \mathbf{U}_3\mathbf{b}_3 + \mathbf{e}$, where $\mathbf{y}$ is $IJK \times 1$, $\mathbf{X}$ is an $IJK \times 1$ vector of ones, $\mathbf{b}_1$ contains the c_{ij}'s, $\mathbf{b}_2$ the a_i's, $\mathbf{b}_3$ the b_j's and $\mathbf{e}$ the e_{ijk}'s. The $\mathbf{U}$'s may be written as $\mathbf{U}_1 = [\mathbf{I}_I \otimes \mathbf{I}_J \otimes \mathbf{e}_K]$, $\mathbf{U}_2 = [\mathbf{I}_I \otimes \mathbf{e}_J \otimes \mathbf{e}_K]$ and $\mathbf{U}_3 = [\mathbf{e}_I \otimes \mathbf{I}_J \otimes \mathbf{e}_K]$ where $\otimes$ signifies left Kronecker product and $\mathbf{I}$ and $\mathbf{e}$ are identity matrices and vectors of ones of appropriate dimension. Some correspondence between the various items defined for the proof of Theorem 3.1 and this model are $n = IJK$; $p_0 = 1$; $p_1 = 3$; $p = 5$; $m_1 = IJ$; $m_2 = I$; $m_3 = J$; $\nu_0 = n - \text{rank}(\mathbf{U}_1 : \mathbf{U}_2 : \mathbf{U}_3) = IJ(K-1)$; $\nu_1 = \text{rank}(\mathbf{U}_1 : \mathbf{U}_2 : \mathbf{U}_3) - \text{rank}(\mathbf{U}_2 : \mathbf{U}_3) = (I-1)(J-1)$; $\nu_2 = \text{rank}(\mathbf{U}_2 : \mathbf{U}_3) - \text{rank}(\mathbf{U}_3) = I - 1$; $\nu_3 = J - 1$; $S_0 = \{0\}$; $S_1 = \{1\}$; $S_2 = \{2, 3\}$. (See below for relation of I and J.) Note that the ν_i correspond to the degrees of freedom of the various sums of squares in the ANOVA table.

We shall only illustrate certain asymptotic results for this model. A complete discussion of the asymptotics and calculations of the MLE's is given in Miller (1973), Sections 6.1 and 6.2. (Hartley and Rao (1967) discuss another ANOVA model at length.) For asymptotic theory of this paper to be applicable it is necessary that $I \to \infty$ and $J \to \infty$. (K may or may not $\to \infty$.) In setting up the S_s we have assumed that $\lim (I/J) = \rho$ with $0 < \rho < \infty$. It then may be shown that one choice for ν_4 is I and that $\mathbf{X}'\mathbf{\Sigma}_0^{-1}\mathbf{X}/\nu_4 = IJK/[(\sigma_{00} + K\sigma_{01} + JK\sigma_{02} + IK\sigma_{03})] \to (\sigma_{02} + \rho\sigma_{03})^{-1}$, which is the only element of $\mathbf{C}_0$. The elements of $\mathbf{C}_1$ are found to depend on whether or not $K \to \infty$. Suppose it does not. Then the 0, 0 term is one-half the limit of $[IJ(K-1)]^{-1}[(\sigma_{00} + K\sigma_{01} + JK\sigma_{02} + IK\sigma_{03})^{-2} + (I-1)(\sigma_{00} + K\sigma_{01} + JK\sigma_{02})^{-2} + (J-1)(\sigma_{00} + K\sigma_{01} + IK\sigma_{03})^{-2} + (I-1)(J-1)(\sigma_{00} + K\sigma_{01})^{-2} + IJ(K-1)\sigma_{00}^{-2}] \to \sigma_{00}^{-2} + (K-1)^{-1}(\sigma_{00} + K\sigma_{01})^{-2}$. The limit of the 0, 1 term is found to be $\frac{1}{2}K(K-1)^{-\frac{1}{2}}(\sigma_{00} + K\sigma_{01})^{-2}$; the (1, 1), (2, 2), (3, 3) terms have respective limits $\frac{1}{2}K^2(\sigma_{00} + K\sigma_{01})^{-2}$, $\frac{1}{2}\sigma_{02}^{-2}$ and $\frac{1}{2}\sigma_{03}^{-2}$; all other limits are zero. (If $K \to \infty$, $\mathbf{C}_1 = \frac{1}{2}\,\text{diag}(\sigma_{00}^{-2}, \sigma_{01}^{-2}, \sigma_{02}^{-2}, \sigma_{03}^{-2})$.) If one then inverts $\mathbf{C}_1$ one obtains the asymptotic variances and covariances of the maximum likelihood estimates. These are found to be identical to the asymptotic variances and covariances that arise if the usual ANOVA estimators of the variance components are normalized by the same sequences.

The above represents but a brief glance at the two-way model. The points to be noted are that the normalizing sequences are necessarily of different orders of magnitude and that there is a close relation between the MLE and ANOVA estimates at least asymptotically.

5. Comments on asymptotic efficiency and on likelihood ratio tests. Since the maximum likelihood estimates are asymptotically normally distributed, it is of interest to discover whether they are asymptotically efficient in the sense of attaining the Cramér–Rao lower bound for the covariance matrix. This bound, the inverse of the Fisher information matrix, cannot be defined in the usual sense in this problem because there is not a sequence of independent observations having a common density from which to compute such a matrix. However, if we define attaining the bound for a sequence of experiments to mean that the difference between the covariance matrix of the estimates in a particular experiment and the inverse of the Fisher information matrix for that experiment converges to zero as we pass through the sequence, which seems reasonable, then the maximum likelihood estimates attain the bound and are thus asymptotically efficient (when properly normalized).

Likelihood ratio tests of hypotheses of the form $\sigma_i = 0$ (or $\alpha_j = 0$) for i (or j) belonging to certain sets are easily calculated because the reduced model (in either case) is another model of the same form. The test statistic is then a ratio of determinants which can be easily computed. Unfortunately the distribution of the test statistic cannot be easily calculated. Under the alternative hypotheses $\sigma_i \neq 0$ Weiss' (1975) result may be applied to yield an asymptotic χ^2 distribution for $-2\log L$. However, under the null hypothesis of $\sigma_i = 0$ the asymptotic distribution is not a χ^2. Consider the model of the first example of Section 4. It is easily shown (see, for instance, Miller (1976)) that $-2\log L = -I\{(J-1)\log[I/(I-1)] + J\log J + \log F - J\log[I(J-1)/(I-1)+F]\}$ when $F > I/(I-1)$ and $-2\log L = 0$ otherwise, where $F = MS_1/MS_0$, $MS_1 = SS_1/(I-1)$ is the usual F statistic. By the change of variable $G = [(I-1)/I]F - 1$ we may rewrite $-2\log L = I\{J\log(1+G/J) - \log(1+G)\}$ when $G > 0$ and $-2\log L = 0$ when $G \leqq 0$. Note that for any fixed $G > 0$, $-2\log L \to \infty$ as $I \to \infty$ whether or not $J \to \infty$. Thus we need only consider the limit as $G \to 0$. Expanding $-2\log L$ and keeping terms up to G^2 we have $-2\log L \doteqdot IG^2(J-1)/2J$. Then for any $X_0 > 0$ we have $P\{-2\log L > X_0\} \doteqdot P\{IG^2(J-1)/2J > X_0\} = P\{G > [2JX_0/(I(J-1))]^{\frac{1}{2}}\}$ because $G < 0$ implies $-2\log L = 0$. But $P\{G > G_0\} = P\{F > F_0\}$, where G_0 and F_0 are functions of X_0, I, and J. Now by considering the distributions of MS_1 and MS_0 derived in Section 4 and recalling that MS_1 and MS_0 are independent it can be shown that $P\{F > F_0\} \doteqdot 1 - \Phi\{[-\sigma_0(1-F_0) - J\sigma_1]/(2[(\sigma_0 + J\sigma_1)^2/(I-1) + (\sigma_0 F_0)^2/I(J-1)]^{\frac{1}{2}})\}$, where Φ is the standard normal cumulative distribution function. But if $\sigma_1 > 0$ the argument of $\Phi(\ )$ converges to $-\infty$ as $I \to \infty$ again whether or not $J \to \infty$ (i.e., the test is consistent). If $\sigma_1 = 0$ the argument of $\Phi(\ )$ becomes after transforming back to F equal to a quantity which tends as $I \to \infty$ (whether or not $J \to \infty$) to $X_0^{\frac{1}{2}}$.

Thus in the limit $P\{-2\log L > X_0\} \to 1 - \Phi(X_0^{\frac{1}{2}})$ if $X_0 > 0$ and $\to \frac{1}{2}$ if $X_0 = 0$ when $\sigma_1 = 0$. Under the null hypothesis then $-2\log L$ is asymptotically a $\frac{1}{2}$, $\frac{1}{2}$ mixture of a χ_1^2 and a χ_0^2 (point mass at zero) random variable. This conforms to Chernoff's (1954) result for the standard case. Further research may generalize this finding to more complex situations.

Thus we see that although the likelihood ratio tests are easy to compute, their usefulness is limited because the distribution under the null hypothesis is generally not known.

Acknowledgment. This paper is based on the author's Ph. D. dissertation at Stanford University. I would sincerely like to thank my advisor, T. W. Anderson, for his suggestions and guidance in its preparation. His insight into this topic was most helpful. I would also like to thank the referees and the associate editor of an earlier version of this paper for their most helpful comments. Their remarks pointed out the applicability of Weiss' result instead of a more cumbersome proof used in the previous version. Their comments also led to many improvements in the presentation of the manuscript.

REFERENCES

ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

ANDERSON, T. W. (1969). Statistical inference for covariance matrices with linear structure. *Proc. Second Internat. Symp. Multivariate Anal.* (P. R. Krishnaiah, ed.), 55–66, Academic Press, New York.

ANDERSON, T. W. (1970). Estimation of covariance matrices which are one linear combination or whose inverses are linear combinations of given matrices. *Essays in Probability and Statist.*, 1–24, Univ. of North Carolina Press, Chapel Hill.

ANDERSON, T. W. (1971). Estimation of covariance matrices with linear structure and moving average processes of finite order. Technical Report No. 6, Department of Statistics, Stanford Univ.

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

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25** 573–578.

HARTLEY, H. O. and RAO, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54** 93–108.

HARTLEY, H. O. and VAUGHN, W. K. (1972). A computer program for the mixed analysis of variance model based on maximum likelihood. *Statistical Papers in Honor of George W. Snedecor* (T. A. Bancroft, ed.), Iowa State Univ. Press, Ames.

MILLER, J. (1973). Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report No. 12, Department of Statistics. Stanford Univ.

MILLER, J. (1976). An example of nonstandard results for large-sample theory of maximum likelihood. Technical Report, Rutgers Univ.

RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.

RUDIN, W. (1964). *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

WEISS, L. (1971). Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *J. Amer. Statist. Assoc.* **66** 345–350.

WEISS, L. (1973). Asymptotic properties of maximum likelihood estimators in some nonstandard cases II. *J. Amer. Statist. Assoc.* **68** 428–430.

WEISS, L. (1975). The asymptotic distribution of the likelihood ratio in some nonstandard cases. *J. Amer. Statist. Assoc.* **70** 204-208.

WHITBY, O. (1971). Estimation of parameters in the generalized beta distribution. Technical Report No. 29, Department of Statistics, Stanford Univ.

APPENDIX A

Details from the proof of Theorem 3.1. The two conditions referred to in the proof of Theorem 3.1 are straightforward; the first is a requirement that the (q_i/n_i) be small and the second is a bounding requirement on the $\mathbf{y}$ vector.

CONDITION A.1. Let q_1, n_i, σ_{0i} be as in Theorem 3.1. Then $(\sigma_{0i}/2) > (q_i/n_i)$, $i = 0, 1, \cdots, p_1$. Now for each n define several matrices (as usual dependence on n is suppressed in the notation); consider $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_2$ to be two fixed points in Θ. Define the sets S_s, $s = 1, 2, \cdots, c$ by (5). Let $\mathbf{H}_c$ be an orthonormal basis for $\mathscr{L}(\mathbf{U}_{i_c}: \cdots : \mathbf{U}_{p_1})$; for $s = 1, 2, \cdots, c-1$, let $\mathbf{H}_s$ be an orthonormal basis for the part of $\mathscr{L}(\mathbf{U}_{i_s}: \cdots : \mathbf{U}_{p_1})$ orthogonal to $\mathscr{L}(\mathbf{U}_{i_{s+1}}: \cdots : \mathbf{U}_{p_1})$; let $\mathbf{H}_0$ be an orthonormal basis for the orthogonal complement of $\mathscr{L}(\mathbf{U}_1: \cdots : \mathbf{U}_{p_1})$. Let the dimension of $\mathbf{H}_s$ be $n \times \tilde{m}_s$, $s = 0, 1, \cdots, c$. Then $\mathbf{P} \equiv [\mathbf{H}_0 : \mathbf{H}_1 : \cdots : \mathbf{H}_c]$ is an $n \times n$ orthogonal matrix. Furthermore, $\mathbf{U}_j'\mathbf{H}_s = \mathbf{0}$ (and hence $\mathbf{G}_j\mathbf{H}_s = \mathbf{0}$) for $i_{s+1} \leqq j \leqq p_1$, $s = 0, 1, \cdots, c$, because $\mathbf{H}_s$ spans a space orthogonal to $\mathscr{L}(\mathbf{U}_{i_{s+1}}: \cdots : \mathbf{U}_{p_1})$. It follows that $(\sum_{i=0}^{p_1} b_i\mathbf{G}_i)\mathbf{H}_s = (\sum_{i=0}^{i_{s+1}-1} b_i\mathbf{G}_i)\mathbf{H}_s$.

Since $\boldsymbol{\Sigma}_2 = \sum_{i=0}^{p_1} \sigma_{2i}\mathbf{G}_i$ is positive definite, there exists a lower triangular matrix $\mathbf{A}_2$ such that $\boldsymbol{\Sigma}_2 = \mathbf{A}_2\mathbf{A}_2'$. But $\mathbf{P}'\boldsymbol{\Sigma}_2\mathbf{P}$ is also positive definite, so there exists $\mathbf{T}$ upper triangular such that $\mathbf{T}'\mathbf{P}'\boldsymbol{\Sigma}_2\mathbf{P}\mathbf{T} = \mathbf{T}'\mathbf{P}'\mathbf{A}_2\mathbf{A}_2'\mathbf{P}\mathbf{T} = \mathbf{I}$. Thus the $n \times n$ matrix $\mathbf{Q} = \mathbf{A}_2'\mathbf{P}\mathbf{T}$ is orthogonal and can be written as $\mathbf{Q} = [\mathbf{Q}_0 : \mathbf{Q}_1 : \cdots : \mathbf{Q}_c] \equiv A'[\mathbf{H}_0^* : \mathbf{H}_1^* : \cdots : \mathbf{H}_c^*]$ where $\mathbf{H}_s^* = \sum_{t=0}^{s} \mathbf{H}_t\mathbf{T}_{ts}$. ($\mathbf{H}_s^*$ is $n \times \tilde{m}_s$ and $\mathbf{T}_{ts}$ is $\tilde{m}_t \times \tilde{m}_s$.) Then $(\sum_{i=0}^{p_1} \tau_i\mathbf{G}_i)\mathbf{H}_s^* = \sum_{i=0}^{p_1}\sum_{t=0}^{s} \tau_i\mathbf{G}_i\mathbf{H}_t\mathbf{T}_{ts} = (\sum_{i=0}^{i_{s+1}-1} \tau_i\mathbf{G}_i)\mathbf{H}_s^*$ because $\mathbf{H}_s^*$ only involves $\mathbf{H}_t$ for $t \leqq s$ and $t \leqq s$, $i \geqq i_{s+1}$ implies $i \geqq i_{t+1}$ so that $\mathbf{G}_i\mathbf{H}_t = \mathbf{0}$.

The vector $\mathbf{z} \equiv \mathbf{A}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2) \sim \mathscr{N}_n(\mathbf{0}, \mathbf{I})$ under $\boldsymbol{\theta}_2$ because $\mathbf{y} \sim \mathscr{N}_n(\mathbf{X}\boldsymbol{\alpha}_2, \boldsymbol{\Sigma}_2)$ under $\boldsymbol{\theta}_2$. Let $\mathbf{w} \equiv \mathbf{Q}'\mathbf{z}$ so that $\mathbf{w} \sim \mathscr{N}_n(\mathbf{0}, \mathbf{I})$ and write $\mathbf{w}' = (\mathbf{w}_0', \mathbf{w}_1', \cdots, \mathbf{w}_c')$ where $\mathbf{w}_s = \mathbf{Q}_s'\mathbf{z} = \mathbf{Q}_s'\mathbf{A}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)$.

CONDITION A.2. For $\mathbf{w}_s$ defined above, $(\mathbf{w}_s'\mathbf{w}_s)/\tilde{m}_s \leqq \frac{11}{10}$, $s = 0, 1, \cdots, c$.

LEMMA A.1. *Under the conditions of Theorem* 3.1, $P_{\boldsymbol{\theta}_2}\{$*Conditions* A.1 *and* A.2 *are true*$\} \to 1$ *as* $n \to \infty$.

PROOF. Each $\mathbf{w}_s'\mathbf{w}_s \sim \chi^2_{\tilde{m}_s}$ under $\boldsymbol{\theta}_2$; each $\tilde{m}_s \to \infty$ because $\tilde{m}_s \geqq \nu_i$ for some $i \in S_s$ (ν_i defined by (6)); $q/n_i \to 0$. The lemma follows.

LEMMA A.2. *Given* $\boldsymbol{\theta}_0$ *an interior point of* Θ *and given* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$ *each in* $N_n(\boldsymbol{\theta}_0)$, *if Condition* A.1 *is true then the following statements are true.*

i) $(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_0) \leqq p_0q^2/n_f^2$.
$(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)'(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) \leqq 4p_0q^2/n_f^2$.

ii) $\lambda_{\max}(\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i) \leqq 1/\sigma_{0i}$.

iii) $\lambda_{\max}(\mathbf{\Sigma}_1^{-1}\mathbf{\Sigma}_0) \leqq 2.$
$\lambda_{\max}(\mathbf{\Sigma}_2^{-1}\mathbf{\Sigma}_0) \leqq 2.$
iv) $\lambda_{\max}(\mathbf{\Sigma}_1^{-1}\mathbf{G}_i) \leqq 2/\sigma_{0i}.$
$\lambda_{\max}(\mathbf{\Sigma}_2^{-1}\mathbf{G}_i) \leqq 2/\sigma_{0i}.$
v) $\max_{1\leqq k\leqq n} |\lambda_k[\mathbf{\Sigma}_0^{-1}(\mathbf{\Sigma}_0 - \mathbf{\Sigma}_2)]| \leqq q/\min_{0\leqq i\leqq p_1}(n_i\sigma_{0i}).$
$\max |\lambda_k[\mathbf{\Sigma}_0^{-1}(\mathbf{\Sigma}_1 - \mathbf{\Sigma}_2)]| \leqq 2q/\min(n_i\sigma_{0i}).$
vi) $\lambda_{\max}(\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{\Sigma}_0^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s) \leqq 6/\sigma_{0i}^2 \; 0 \leqq i < i_{s+1} \qquad s = 0, 1, \cdots, c$
$= 0$ *otherwise.*

PROOF. Observe that by the definition of $N_n(\boldsymbol{\theta}_0)$, $|\sigma_{0i} - \sigma_{2i}| \leqq q/n_i < \sigma_{0i}/2$ by Condition A.1. This implies $\sigma_{0i}/2 < \sigma_{2i} < 3\sigma_{0i}/2$; the same is true for σ_{1i}. Statement (i) follows from the definition of $N_n(\boldsymbol{\theta}_0)$ for $\boldsymbol{\alpha}$ and from the triangle inequality (which also gives $|\sigma_{1i} - \sigma_{2i}| < 2q/n_i$). Consider that for matrices of the form $\mathbf{C} = \mathbf{D}^{-1}\mathbf{E}$ ($\mathbf{D}$ positive definite, $\mathbf{E}$ symmetric) every characteristic root is of the form $\mathbf{x}'\mathbf{E}\mathbf{x}/\mathbf{x}'\mathbf{D}\mathbf{x}$ for some vector $\mathbf{x}$. Consider also that $(\sum_{i=0}^{p_1} a_i z_i)/(\sum_{i=0}^{p_1} b_i z_i) \leqq \max(a_i/b_i)$ provided $b_i > 0$, $z_i \geqq 0$, $i = 0, 1, \cdots, p_1$, and some $z_i > 0$. It then follows that (sup is taken over $\mathbf{x} \neq \mathbf{0}$) $\lambda_{\max}(\mathbf{\Sigma}_0^{-1}\mathbf{G}_i) = \sup(\mathbf{x}'\mathbf{G}_i\mathbf{x}/\mathbf{x}'\mathbf{\Sigma}_0\mathbf{x}) = \sup(\mathbf{x}'\mathbf{G}_i\mathbf{x}/\sum_{j=0}^{p_1}\sigma_{0j}\mathbf{x}'\mathbf{G}_j\mathbf{x}) \leqq \max(0, 1/\sigma_{0i}) = 1/\sigma_{0i}$ because $\sigma_{0j} > 0$, each $\mathbf{G}_j$ is positive semidefinite, and $\mathbf{G}_0$ is positive definite. Statements (iii) and (iv) follow by analogous arguments. Furthermore, $\max|\lambda_k[\mathbf{\Sigma}_0^{-1}(\mathbf{\Sigma}_0 - \mathbf{\Sigma}_2)]| \leqq \sup|\mathbf{x}'(\mathbf{\Sigma}_0 - \mathbf{\Sigma}_2)\mathbf{x}|/\mathbf{x}'\mathbf{\Sigma}_0\mathbf{x} \leqq \sup[\sum_{i=0}^{p_1}|\sigma_{0i} - \sigma_{2i}|\mathbf{x}'\mathbf{G}_i\mathbf{x}]/[\sum_{i=0}^{p_1}\sigma_{0i}\mathbf{x}'\mathbf{G}_i\mathbf{x}] \leqq \max[|\sigma_{0i} - \sigma_{2i}|/\sigma_{0i}] \leqq \max[q/n_i\sigma_{0i}] = q/\min(n_i\sigma_{0i})$ by the same argument.

Now observe that $\lambda_{\max}(\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{\Sigma}_0^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s) = \lambda_{\max}(\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{A}_2'\mathbf{\Sigma}_0^{-1} \times \mathbf{A}_2\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s) \leqq \lambda_{\max}(\mathbf{A}_2'\mathbf{\Sigma}_0^{-1}\mathbf{A}_2)\lambda_{\max}(\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s)$ by elementary properties of characteristic roots. The first term is $\lambda_{\max}(\mathbf{\Sigma}_0^{-1}\mathbf{\Sigma}_2)$ and is bounded by $\frac{3}{2}$ as above. The second term equals (sup is over $\boldsymbol{\gamma} \neq \mathbf{0}$ and then over $\mathbf{x} \neq \mathbf{0}$) $\sup[\boldsymbol{\gamma}'\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s\boldsymbol{\gamma}/\boldsymbol{\gamma}'\boldsymbol{\gamma}] = \sup[\boldsymbol{\gamma}'\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t} \times \mathbf{Q}_s\boldsymbol{\gamma}/\boldsymbol{\gamma}'\mathbf{Q}_s'\mathbf{Q}_s\boldsymbol{\gamma}] \leqq \sup[\mathbf{x}'\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{A}_2^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{x}/\mathbf{x}'\mathbf{x}] = \lambda_{\max}(\mathbf{\Sigma}_2^{-1}\mathbf{G}_i)^2 \leqq 4/\sigma_{0i}^2$. Thus the bound is derived as $(4/\sigma_{0i}^2)(\frac{3}{2}) = 6/\sigma_{0i}^2$. However, if $i \geqq i_{s+1}$ then $\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s = \mathbf{G}_i\mathbf{H}_s^* = \mathbf{0}$ as was shown above. The matrix in question is then the zero matrix and has all characteristic roots equal to zero.

LEMMA A.3. *If Condition* A.2 *is true then* $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)'\mathbf{F}'\mathbf{\Sigma}_0^{-1}\mathbf{F}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2) \leqq \frac{11}{10}[\sum_{s=0}^{c}\tilde{m}_s^{\frac{1}{2}}\lambda_{\max}^{\frac{1}{2}}(\mathbf{Q}_s'\mathbf{A}_2'\mathbf{F}'\mathbf{\Sigma}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_s)]^2$ *for any* $n \times n$ *matrix* $\mathbf{F}$.

PROOF. Observe that $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2) = \mathbf{A}_2\mathbf{Q}\mathbf{Q}'\mathbf{A}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2) = \mathbf{A}_2\mathbf{Q}\mathbf{w} = \mathbf{A}_2\sum_{s=0}^{c}\mathbf{Q}_s\mathbf{w}_s$, where $\mathbf{w}$ is defined above. This yields $(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)'\mathbf{F}'\mathbf{\Sigma}_0^{-1}\mathbf{F}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2) = \sum_{s=0}^{c}\sum_{t=0}^{c}\mathbf{w}_s'\mathbf{Q}_s'\mathbf{A}_2'\mathbf{F}'\mathbf{A}_0^{-t}\mathbf{A}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_t\mathbf{w}_t$, where $\mathbf{\Sigma}_0 = \mathbf{A}_0\mathbf{A}_0'$. But the square of any term of the sum is bounded by $(\mathbf{w}_s'\mathbf{Q}_s'\mathbf{A}_2'\mathbf{F}'\mathbf{A}_0^{-t}\mathbf{A}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_s\mathbf{w}_s) \times (\mathbf{w}_t'\mathbf{Q}_t'\mathbf{A}_2'\mathbf{F}'\mathbf{A}_0^{-t}\mathbf{A}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_t\mathbf{w}_t)$ by the Cauchy–Schwarz inequality. But $\mathbf{w}_s'\mathbf{Q}_s'\mathbf{A}_2'\mathbf{F}'\mathbf{A}_0^{-t}\mathbf{A}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_s\mathbf{w}_s \leqq \mathbf{w}_s'\mathbf{w}_s\lambda_{\max}(\mathbf{Q}_s'\mathbf{A}_2'\mathbf{F}'\mathbf{\Sigma}_0^{-1}\mathbf{F}\mathbf{A}_2\mathbf{Q}_s)$ and $\mathbf{w}_s'\mathbf{w}_s \leqq (\frac{11}{10})\tilde{m}_s$ by Condition A.2. The lemma follows immediately.

Using Lemmas A.2 and A.3 and similar lemmas we can prove the details needed in the proof of Theorem 3.1 and also prove the following lemma which completes the proof.

LEMMA A.4. *For $\boldsymbol{\theta}_0$ an interior point of Θ and for a fixed $\boldsymbol{\theta}_2 \in N_n(\boldsymbol{\theta}_0)$, if Conditions* A.1 *and* A.2 *are true then*

$$\frac{q_i q_j}{n_i n_j} \sup_{\boldsymbol{\theta}_1 \in N_n(\boldsymbol{\theta}_0)} \left| \frac{\partial^2 \lambda}{\partial \theta_i \, \partial \theta_j}\bigg|_{\boldsymbol{\theta}_1} - \frac{\partial^2 \lambda}{\partial \theta_i \, \partial \theta_j}\bigg|_{\boldsymbol{\theta}_2} \right| \to 0$$

as $n \to \infty$, for all i, j.

PROOF. The proof of this lemma is quite tedious. We give an example which illustrates the method of proof. Consider derivatives of the form $\partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\sigma_i$ defined by (13). It is sufficient to prove that for $p_0 \times 1$ vector $\boldsymbol{\xi}$ such that $\boldsymbol{\xi}'\boldsymbol{\xi} = 1$, $q^2/(n_i n_f)$ times the difference $\boldsymbol{\xi}'[\partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\sigma_i|_{\boldsymbol{\theta}_1} - \partial^2\lambda/\partial\boldsymbol{\alpha}\,\partial\sigma_i|_{\boldsymbol{\theta}_2}]$ converges to zero independent of $\boldsymbol{\theta}_1$ provided $\boldsymbol{\theta}_1 \in N_n(\boldsymbol{\theta}_0)$. The quantity in question can be written as $|[q^2/(n_i n_f)]\boldsymbol{\xi}'\mathbf{X}'[\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_1^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_1) - \boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)]| \leqq |[q^2/(n_i n_f)]\boldsymbol{\xi}'\mathbf{X}'[\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}](\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)| + |[q^2/(n_i n_f)]\boldsymbol{\xi}'\mathbf{X}'\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_1^{-1} \times \mathbf{X}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)|$. The square of the second term is bounded by $[(q^4 n_f^2)/n_i^2]\boldsymbol{\xi}'\boldsymbol{\xi}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)'(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)[\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})/n_f^2]^2\lambda_{\max}[\mathbf{A}_0'\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_0]$ by several applications of the Cauchy–Schwarz inequality and the definition of characteristic root. But $\boldsymbol{\xi}'\boldsymbol{\xi} = 1$; $(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1) \leqq 4p_0 q^2/n_f^2$ by Lemma A.2.i; the first characteristic root term in brackets is bounded because the matrix converges to $\mathbf{C}_0$, a constant matrix, by Assumption 3.4; the last characteristic root is easily bounded by $16/\sigma_{0i}^2$ by A.2.iii and A.2.iv. Thus the square of the second term is bounded by a constant times $q^6/n_i^2 \to 0$ by the definition of q.

The first term can be broken into three terms by writing $\boldsymbol{\Sigma}_1^{-1} = \boldsymbol{\Sigma}_2^{-1} + (\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1})$ and observing $\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_2^{-1} = \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_1^{-1}$. One such term is $[q^2/(n_i n_f)]|\boldsymbol{\xi}'\mathbf{X}'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_1^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)| = [q^2/(n_i n_f)]|\boldsymbol{\xi}'\mathbf{X}'\mathbf{A}_0^{-t}\mathbf{A}_0'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_0\mathbf{A}_0^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)|$. Again we use the Cauchy–Schwarz inequality and definition of characteristic root to bound the square of this term by $q^4\boldsymbol{\xi}'\boldsymbol{\xi}[\lambda_{\max}(\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})/n_f^2]\lambda_{\max}(\mathbf{A}_0'\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_1^{-1}\mathbf{A}_0\mathbf{A}_0'\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Sigma}_2 - \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_2^{-1}\mathbf{A}_0)(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)'\boldsymbol{\Sigma}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_2^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}_2)/n_i^2$. The first three terms after q^4 are bounded as above. The last term is bounded using Lemma A.3 by $(\frac{1}{16})[\sum_{s=0}^{c}(\tilde{m}_s/n_i^2)^{\frac{1}{2}} \times \lambda_{\max}^{\frac{1}{2}}(\mathbf{Q}_s'\mathbf{A}_2^{-1}\mathbf{G}_i\boldsymbol{\Sigma}_0^{-1}\mathbf{G}_i\mathbf{A}_2^{-t}\mathbf{Q}_s)]^2$. But $\tilde{m}_s/n_i^2 = \tilde{m}_s/\nu_i$ is bounded so long as $i < i_{s+1}$. (Both $\tilde{m}_s$ and ν_i are of the same magnitude if $i \in S_s$ and $\tilde{m}_s$ is of smaller order if $i < i_s$.) In this case the characteristic root is bounded by $2/\sigma_{0i}^2$ by Lemma A.2. ii and iv.

When $i \geqq i_{s+1}\tilde{m}_s/\nu_i$ is not bounded but the characteristic root is zero. (This argument illustrates the necessity of partitioning by using the $\mathbf{Q}$ matrix. At crucial points in the proof, the n_i term in the denominator cannot overwhelm the numerator unless manipulations with $\mathbf{Q}$ are used.) Thus the last term is bounded. The third term is bounded by $16q^2/\min(n_i\sigma_{0i})^2$ using the definition of characteristic root and A.2.v. Then the entire term is bounded by a constant times $q^6/\min(n_i\sigma_{0i})^2 \to 0$ by definition of q. Thus the lemma is true for these terms. The complete proof of the lemma and of Theorem 3.1 is similar to arguments given in Miller (1973).

APPENDIX B

A condition sufficient for positive definiteness of $\mathbf{C}_1$. We first note that to show $\mathbf{C}_1$ is positive definite requires us to show that for any $(p_1+1)\times 1$ vector $\mathbf{b}\neq\mathbf{0}$ $\mathbf{b}'\mathbf{C}_1\mathbf{b}>0$. But $2\mathbf{b}'\mathbf{C}_1\mathbf{b}=\sum_{i=0}^{p_1}\sum_{j=0}^{p_1} b_i b_j \lim \operatorname{tr} \mathbf{\Sigma}_0^{-1}\mathbf{G}_i\mathbf{\Sigma}_0^{-1}\mathbf{G}_j/n_i n_j = \lim \operatorname{tr} [\mathbf{\Sigma}_0^{-1}(\sum_{i=0}^{p_1}(b_i/n_i)\mathbf{G}_i)]^2 \geqq 0$. We must prove the inequality is strict in the limit. One condition which is sufficient to guarantee this is the following. For each $\mathbf{U}_i$, $i=1,2,\cdots,p_1$, let the columns of $\mathbf{U}_i$ be represented by $\mathbf{U}_i=[\mathbf{u}_1^{(i)},\mathbf{u}_2^{(i)},\cdots,\mathbf{u}_{m_i}^{(i)}]$; the sets S_s are defined by (5).

CONDITION B.1. For every i and every $j\in S_s$, $j\neq i$, where $i\in S_s$, there exist two nonnegative constants, R_1 and R_2, both less than or equal to one, such that $\sum_{l=1}^{m_j}[(\mathbf{u}_l^{(j)\prime}\mathbf{u}_k^{(i)})/(\mathbf{u}_k^{(i)\prime}\mathbf{u}_k^{(i)})]\leqq R_2$ for all but $R_1 m_i$ values of k in the set $\{1,2,\cdots,m_i\}$. Furthermore, R_1 and R_2 are such that $R_1+(1-R_1)R_2<(N(S_s)+1)^{-1}$, where $N(S_s)$ is the number of indices in the set S_s. The proof that Condition B.1 is sufficient for the positive definiteness of $\mathbf{C}_1$ is given in full detail in Miller (1973; pages 180–193). We illustrate here how one proceeds.

Let $\mathbf{B}=\sum_{i=0}^{p_1}(b_i/n_i)\mathbf{G}_i$. Suppose $b_0\neq 0$. Then we show that $\lim \operatorname{tr}(\mathbf{\Sigma}_0^{-1}\mathbf{B})^2>0$ by showing that a certain number of characteristic roots of $\mathbf{\Sigma}_0^{-1}\mathbf{B}$ are large enough. In particular there is a space of dimension ν_0 orthogonal to $\mathscr{L}(\mathbf{U}_1:\mathbf{U}_2:\cdots:\mathbf{U}_{p_1})$ via (6) and hence for any vector $\mathbf{x}$ in this space $\mathbf{\Sigma}_0\mathbf{x}=\sigma_{00}\mathbf{x}$ and $\mathbf{B}\mathbf{x}=(b_0/n_0)\mathbf{x}$. Thus there are ν_0 $(=n_0^2)$ independent characteristic vectors of $\mathbf{\Sigma}_0^{-1}\mathbf{B}$ with characteristic root equal to $b_0/(\sigma_{00}n_0)$. But then $\operatorname{tr}(\mathbf{\Sigma}_0^{-1}\mathbf{B})^2 \geqq n_0^2[b_0/(\sigma_{00}n_0)]^2 = b_0^2/\sigma_{00}>0$. The situation is much more complicated for $b_0=0$. Let $s>0$ be the least index for which $b_i\neq 0$ for some $i\in S_s$. Observe $|b_i|/n_i$ for each $i\in S_s$; one of these must be the largest in the sense that $\lim[(|b_j|/n_j)/(|b_i|/n_i)]\leqq 1$ for $j\in S_s$, $j\neq i$. Consider this i fixed and bound characteristic roots of $\mathbf{\Sigma}_0^{-1}\mathbf{B}$ by considering forms $(\gamma'\mathbf{U}_i'\mathbf{B}\mathbf{U}_i\gamma)/(\gamma'\mathbf{U}_i'\mathbf{\Sigma}_0\mathbf{U}_i\gamma)$. By placing appropriate restrictions on γ and by using Condition B.1 we can bound enough characteristic roots far enough from zero for $\lim \operatorname{tr}(\mathbf{\Sigma}_0^{-1}\mathbf{B})^2$ to be positive.

A short comment about this assumption. It does seem unwieldy (although it does occur naturally in the above proof) and probably is too strong. However, most design sequences will meet this condition, which might be called "asymptotic near orthogonality." Simpler conditions (for instance Assumption 3.3 above) have not been proved sufficient by this author.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK
NEW JERSEY 08903