*The Annals of Statistics*
1997, Vol. 25, No. 2, 897–916

# GENERALIZATION OF LIKELIHOOD RATIO TESTS UNDER NONSTANDARD CONDITIONS

BY H. T. V. VU AND S. ZHOU

*University of Western Australia*

In this paper, we analyze the statistic which is the difference in the values of an estimating function evaluated at its local maxima on two different subsets of the parameter space, assuming that the true parameter is in each subset, but possibly on the boundary. Our results extend known methods by covering a large class of estimation problems which allow sampling from nonidentically distributed random variables. Specifically, the existence and consistency of the local maximum estimators and asymptotic properties of useful hypothesis tests are obtained under certain law of large number and central limit-type assumptions. Other models covered include those with general log-likelihoods and/or covariates. As an example, the large sample theory of two-way nested random variance components models with covariates is derived from our main results.

**1. Introduction.** The purpose of this paper is to derive large sample properties of estimators obtained from a certain class of estimating functions. In order to include models involving covariates in statistics, we allow the sample to be collected from nonidentically distributed random variables. The true parameters are allowed to be on the boundary of the parameter space. The results are stated in terms of properties of a maximum estimator (ME) which maximizes an estimating function $\mathscr{L}_n(\theta)$ on the intersection between an open neighborhood of the true parameter and a given subset of the parameter space. Sufficient conditions are derived for the existence and consistency of a maximum estimator on a given region, and the large sample distribution of the deviance statistic $d_n = 2[\mathscr{L}_n(\hat{\theta}_n^2) - \mathscr{L}_n(\hat{\theta}_n^1)]$, where $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ are consistent ME's on two different subsets $\Omega$ and $\tau$ of the parameter space, is obtained. Especially, explicit expressions for the asymptotic distribution of $d_n$ are given when the parameter spaces are the product of intervals.

An ME is called a maximum likelihood estimator (MLE) if the estimating function is the log-likelihood. Thorough investigations of consistent MLE's for a general sample space have been done by Chernoff (1954), Feder (1968), Moran (1971) and Chant (1974), when the sample is of independent random variables having a common density function $f(x, \theta)$. Crowder (1990) considered the same setup with Weibull random variables. We refer to "interior" and "boundary" problems according to whether the true parameter is in the

Received January 1994; revised January 1995.

*AMS* 1991 *subject classifications*. Primary 62E20, 62F03, 62F05, 62H15; secondary 62F05, 62J10.

*Key words and phrases.* Maximum estimators, boundary hypothesis tests.

interior or on the boundary of the parameter space. Self and Liang (1987) gave a general approach for both problems when the sample is of independently and identically distributed random variables.

More recently, Geyer (1994) provides conditions under which asymptotics of global or local maximum estimators of a general estimating function $\mathscr{L}_n(\theta)$ are obtained for a sequence of observations. Geyer proves that the asymptotic distribution of $\mathscr{L}_n(\hat{\theta}_n) - \mathscr{L}_n(\theta_0)$ is a projection of a normal random vector on the tangent cone for a consistent sequence $\{\hat{\theta}_n\}$ of global maximum estimators under the Chernoff regularity of a subset of the parameter space, and for a $\sqrt{n}$-consistent sequence $\{\hat{\theta}_n\}$ of local maximum estimators under the Clarke regularity of a subset of the parameter space. The Clarke regularity is not needed in our formulation since there always exists a global maximizer on a neighborhood of the true parameter with probability approaching 1 under our conditions. Thus the results in Self and Liang (1987) still hold with the maximum estimators considered in this paper. The local maximum estimator $\hat{\theta}_n$ in Geyer (1994) maximizes the estimating function on the intersection between a subset of the parameter space and a neighborhood of $\hat{\theta}_n$ which does not necessarily contain the true parameter.

Geyer (1994) assumes a sampling model that is essentially a stationary process. Our model has no such restrictions. In particular, we allow general nonidentically distributed sampling so that models with covariances can be included. Moreover, Geyer (1994) uses a $\sqrt{n}$ scaling, as one would expect under stationary assumptions, whereas we scale more generally by a square root of the observed negative Hessian of the objective function. This enables us to obtain results when the convergence rate is not $n^{-1/2}$ or when different components of the parameter vector converge at different rates. This is needed for models involving covariates. We do require an extra condition (A3) which is shown to be necessary in Remark 3.1. Our results hold in fact under a generalized version of Chernoff regularity stated in Remark 2.2. For a stationary process, our generalized Chernoff regularity reduces to the Chernoff regularity stated in Geyer (1994).

It is revealed in this paper that in order to ensure that the asymptotic distribution of the deviance $d_n$ exists, the parameter subsets $\Omega$ and $\tau$ must settle down to a fixed cone possibly after certain transformations, as $n$ tends to infinity. (Recall that a subset $C$ of $\mathbb{R}^k$ is a cone with vertex at 0 if $x \in C$ implies that $\lambda x \in C$ for all $\lambda > 0$.) This requirement is described by condition (A3) in Section 2. The effect of (A3) is shown by an example in Remark 3.1 where (A3) is violated and the asymptotic distribution of $d_n$ does not exist. Furthermore, our results in Theorem 2.3 show that the existence and the form of the asymptotic distribution of $d_n$ depend on the asymptotic behavior of the expected information matrix and/or the forms of $\Omega$ and $\tau$. Such effects of the information matrix and the forms of $\Omega$ and $\tau$ can only be revealed by the use of the observed negative Hessian of the objective function. For convenience, the regions $\Omega$ and $\tau$ in this paper are assumed to coincide with a closed cone near the true parameter $\theta_0$, as specified in Assumption (A2) in Section 2. However, our results are still valid if (A2) is relaxed to requiring only that $\Omega$ (and/or $\tau$) can be approximated by a cone with vertex at $\theta_0$ in the sense described in Remark 2.2.

Our method is to combine the approach in Self and Liang (1987) for a general parameter space with the approach in Fahrmeir and Kaufmann (1985) for a sequence of observations drawn from nonidentically distributed random variables. It copes, for example, with general log-likelihoods for a sample of observations drawn from random variables with improper and/or censored distribution functions. (In survival analysis, failure time is said to be "censored" if it is longer than follow-up time.)

In fact one motivation for this work came from a need to fit mixture models to survival data in which not all individuals are subject to death or failure. Such data sets occur, for example, in reliability analysis, where failure time may be the time for a device to malfunction in a certain way, if this occurs, in recividism studies in criminology, and in medical studies, where there may be an immune or cured proportion in the population consisting of those who never catch the same disease again [see Ghitany, Maller and Zhou (1994)]. In other words, "immune" individuals are those who never fail. We allow improper failure distributions so as to allow for a proportion of immunes in the model, and a question of great interest in medical or criminological studies, for example, is whether there is indeed a component of immune individuals present. This boundary testing problem falls within the scope of our methods. Furthermore, covariates such as age, race and so on may be included to account for differences between observations.

This paper concentrates on the properties of hypothesis tests for both interior and boundary problems for models involving covariates. As a substantial example, we derive the nonstandard asymptotic distribution of the likelihood ratio (LR) tests for the two-way nested random variance components model. Searle, Casella and McCulloch (1992) gave the exact distribution of the log-likelihood ratio to test the hypothesis that the variance component is equal to zero for a one-way random model. However, neither exact nor asymptotic distributions of the log-likelihood ratio to test the hypotheses that one or both variance components are equal to zero for the above two-way nested random model are mentioned by them. Suppose that we have $I$ classes where each class has $J_i$ members. We select a random sample of $K_i$ observations from the $j$th member of the $i$th class, $i = 1, \dots, I$, $j = 1, \dots, J_i$. Suppose that the $k'$th observation of the $j$th member from the $i$th class has the form

$$(1.1) \qquad y_{ijk'} = \varepsilon_{ijk'} + B_{ij} + A_i, \quad i = 1, \dots, I, j = 1, \dots, J_i, k' = 1, \dots, K_i,$$

where

$$(1.2) \qquad \varepsilon_{ijk'} \sim N(0, \sigma^2), \qquad B_{ij} \sim N(0, \sigma_B^2) \quad \text{and} \quad A_i \sim N(\mu, \sigma_A^2).$$

Define $\theta_1 = \mu$, $\theta_2 = \sigma^2$, $\theta_3 = \sigma_B^2$ and $\theta_4 = \sigma_A^2$. Then $\theta = (\theta_1 \;\; \theta_2 \;\; \theta_3 \;\; \theta_4)^T \in \Theta = \mathbb{R} \times (0, \infty) \times [0, \infty) \times [0, \infty)$ is the parameter to be estimated. Suppose that we wish to test the hypothesis that the variance components $\sigma_A^2$ and $\sigma_B^2$ are both zero. Then we let the true parameter be $\theta_0 = (\theta_{10} \;\; \theta_{20} \;\; \theta_{30} \;\; \theta_{40})^T =$

$(\mu_0\ \sigma_0^2\ \sigma_{B0}^2\ \sigma_{A0}^2)^T = (\mu_0\ \sigma_0^2\ 0\ 0)^T$. Under this hypothesis and other assumptions set out in Section 3, we will derive the asymptotic distribution of

$$d_I = 2\left[\sup_{\theta \in N_I(A)} \mathscr{L}_I(\theta) - \sup_{\theta \in N_I(A),\, \theta_3 = \theta_4 = 0} \mathscr{L}_I(\theta)\right] \tag{1.3}$$

where $\mathscr{L}_I(\theta)$ is the log-likelihood and $N_I(A)$ is a neighborhood of $\theta_0$ defined by (2.6) with $n = I$ and $k = 4$. This distribution is given in Theorem 3.1, and it is not a chi-squared distribution or even a mixture of chi-squared distributions. Furthermore, we may drop the normality assumptions on $\varepsilon_{ijk'}$, $B_{ij}$ and $A_i$ in (1.2) and Theorem 3.1 remains valid if we use (3.6) as an estimating function, provided that $\varepsilon_{ijk'}$, $B_{ij}$ and $A_i$ have bounded fifth moments.

It should be noted that the assumptions required on the estimating functions in this paper do not involve any specific forms for the sample distributions, unlike ordinary likelihood methods where the specifications of the distributions are crucial. In some models such as quasi-likelihood models, or least squares procedures, the appropriate estimating functions may arise naturally. In other cases, we may use the log-likelihood from distributions which are not necessarily the distributions of the observations, such as in the above example.

In the next section we state the assumptions under which we can derive the asymptotic properties of local maxima of $\mathscr{L}_n(\theta)$ and of hypothesis tests based on them. In Section 3, we state and discuss the result for the two-way nested random variance components model mentioned above. All proofs are relegated to Section 4.

**2. The main results.** Consider a sample of $n$ observations on random variables $Y_1, \dots, Y_n$. Suppose that the distribution function of $Y_i$ is drawn from the family $\mathscr{F}_i(y; \theta)$, where $\theta \in \mathbb{R}^k$ is the parameter to be estimated. The true distribution function of $Y_i$ is $\mathscr{F}_i(y; \theta_0)$, where $\theta_0 = (\theta_{10} \cdots \theta_{k0})^T$ is called the true parameter. Consider an estimating function of the form

$$\mathscr{L}_n(\theta) = \sum_{i=1}^{n} g(Y_i, \theta),$$

where $g$ is a function from $\mathbb{R}^{k+1}$ to $\mathbb{R}$. The parameter $\theta$ will be restricted to lie in a parameter space $\Theta \subseteq \mathbb{R}^k$, which is assumed to be a cone of the form

$$\Theta = \left\{\theta = \theta_0 + \tilde{\theta}_1 u_1 + \cdots + \tilde{\theta}_k u_k : \tilde{\theta}_j \in I_j,\ j = 1, \dots, k\right\} \tag{2.1}$$

where $u_j$ are $k$ linearly independent unit vectors, $j = 1, \dots, k$, and $I_j$'s are either closed, half open or open intervals containing 0.

We will need to define derivatives in $\Theta$. This is done as follows. Let $\theta = \theta_0 + \tilde{\theta}_1 u_1 + \cdots + \tilde{\theta}_k u_k$, $\tilde{\theta}_j \in I_j$, $j = 1, \dots, k$. For each $j = 1, \dots, k$, let $D_{u_j}\mathscr{L}_n(\theta)$ be the usual directional derivative of $\mathscr{L}_n(\theta)$ in the direction $u_j$ if $\tilde{\theta}_j$ is in the interior of $I_j$. If $\tilde{\theta}_j$ is on the boundary of $I_j$, define

$$D_{u_j} g(\theta) = \lim_{h \to 0,\, h + \tilde{\theta}_j \in I_j} D_{u_j} g(\theta + h u_j).$$

If $\tilde{\theta}_j$ is in the interior of $I_j$, denote by $D_{u_j} D_{u_l} \mathscr{L}_n(\theta)$ the usual directional derivative of $D_{u_l} \mathscr{L}_n(\theta)$ in the direction $u_j$. If $\tilde{\theta}_j$ is on the boundary of $I_j$, define

$$D_{u_j} D_{u_l} g(\theta) = \lim_{h \to 0,\, h + \tilde{\theta}_j \in I_j} D_{u_j} D_{u_l} g(\theta + h u_j).$$

Basic properties such as one-sided Taylor expansions of $g(\theta)$ can be easily derived using these definitions. We make the following assumptions on the function $\mathscr{L}_n$ and the parameter spaces we consider.

(A1) For a neighborhood $\mathscr{N}$ of $\theta_0$, the function $\mathscr{L}_n(\theta)$ is continuous on $\Theta \cap \mathscr{N}$, and the first and second directional derivatives $D_{u_j} \mathscr{L}_n(\theta)$ and $D_{u_j} D_{u_l} \mathscr{L}_n(\theta)$, $j, l = 1, \dots, k$, exist, are finite and are continuous on $\Theta \cap \mathscr{N}$.

(A2) A subset $\Omega$ and $\Theta$ is said to satisfy (A2) if there is a closed cone $C_\Omega$ with vertex at $\theta_0$ such that

$$C_\Omega \subseteq \Theta \quad \text{and} \quad C_\Omega \cap \mathscr{N} = \Omega \cap \mathscr{N}, \tag{2.2}$$

where $\mathscr{N}$ is a closed neighborhood of $\theta_0$.

For any positive definite matrix $\mathbf{A}$, let $\mathbf{A}^{1/2}(\mathbf{A}^{T/2})$ be a left (the corresponding right) square root of $\mathbf{A}$, that is, any matrices satisfying $\mathbf{A}^{1/2}\mathbf{A}^{T/2} = \mathbf{A}$, where $\mathbf{A}^{T/2} = (\mathbf{A}^{1/2})^T$. In addition, let $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$ and $\mathbf{A}^{-T/2} = (\mathbf{A}^{T/2})^{-1}$. Usual versions of the square root are the Cholesky square root and the symmetric positive definite square root. The left and right Cholesky square roots $\mathbf{A}^{1/2}$ and $\mathbf{A}^{T/2}$ are defined as the lower and upper triangular matrices with positive diagonal elements satisfying $\mathbf{A}^{1/2}\mathbf{A}^{T/2} = \mathbf{A}$ and $\mathbf{A}^{T/2} = (\mathbf{A}^{1/2})^T$. Denote by $\|\cdot\|_1$ the sum of the absolute values of the elements of a matrix. Also denote by $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ the minimum and the maximum eigenvalues of a symmetric matrix.

Let

$$\mathbf{T} = [u_1 \cdots u_k] \tag{2.3}$$

denote the $k \times k$ matrix of directions defining $\Theta$ [see (2.1)]. We now define the derivative of $\mathscr{L}_n(\theta)$ with respect to $\theta$ to be the $k$-vector

$$S_n(\theta) = \mathbf{T}^T \big[ D_{u_1} \mathscr{L}_n(\theta) \cdots D_{u_k} \mathscr{L}_n(\theta) \big]^T, \tag{2.4}$$

where $\mathbf{T}^T$ denotes the transpose of $\mathbf{T}$. Similarly, we define the negative of the second derivative of $\mathscr{L}_n(\theta)$ to be the $k \times k$ symmetric matrix

$$\mathbf{F}_n(\theta) = -\mathbf{T}^T \big[ D_{u_j} D_{u_l} \mathscr{L}_n(\theta) \big] \mathbf{T}. \tag{2.5}$$

Define $\mathbf{D}_n = \mathbb{E}\{S_n(\theta_0) S_n^T(\theta_0)\}$ and $\mathbf{G}_n = \mathbb{E}\{\mathbf{F}_n(\theta_0)\}$. For any fixed $A > 0$, define subsets of $\mathbb{R}^k$ by

$$N_n(A) = \big\{ \theta : (\theta - \theta_0)^T \mathbf{G}_n (\theta - \theta_0) \le A^2,\ \theta \in \Theta \big\}, \tag{2.6}$$

$$M_n(A) = \big\{ \theta : (\theta - \theta_0)^T \mathbf{G}_n (\theta - \theta_0) = A^2,\ \theta \in \Theta \big\}. \tag{2.7}$$

To obtain the existence, consistency and the asymptotic distribution of an ME for the model, we need the following assumptions on the asymptotic behavior of the first and second derivative matrices and their expectations. (Convergences are as $n \to \infty$ unless otherwise stated.)

(B1) $\mathbb{E}\{S_n(\theta_0)\} = 0$, and the matrices $\mathbf{D}_n$ and $\mathbf{G}_n$ are finite, where the expectations are taken with respect to the true distributions.

(B2) $\lambda_{\min}\{\mathbf{G}_n\} \to \infty$. (When (B2) holds, $\mathbf{G}_n$ is positive definite for $n$ large enough, so we assume it to be so in general.)

(B3) $\sup_{\theta \in N_n(A)} \|\mathbf{G}_n^{-1/2}\mathbf{F}_n(\theta)\mathbf{G}_n^{-T/2} - \mathbf{I}_k\|_1 \to_P 0$.

(B4) For some positive definite matrix $\mathbf{V}$, $\|\mathbf{G}_n^{-1/2}\mathbf{D}_n\mathbf{G}_n^{-T/2} - \mathbf{V}\|_1 \to 0$. (When (B2) and (B4) hold, $\mathbf{D}_n$ is positive definite for $n$ large enough, so we assume it to be so in general.)

(B5) $\mathbf{D}_n^{-1/2}S_n(\theta_0) \to_D N(0, \mathbf{I}_k)$.

Denote by $|y|$ the modulus of a vector $y \in \mathbb{R}^k$. We say that a sequence of events $\{A_n\}$ occurs with probability approaching 1 (WPA1) if $\mathbb{P}\{A_n\} \to 1$ as $n \to \infty$. We wish to define maximum estimates (ME's) with respect to a fixed subset $\Omega$ of $\Theta$. An estimate $\hat{\theta}_n$ of $\theta_0$ is called a maximum estimate on $\Omega$ if $\mathscr{L}_n(\hat{\theta}_n)$ is the maximum of $\mathscr{L}_n(\theta)$ on an intersection between $\Omega$ and an open (possibly depending on $n$) neighborhood of $\theta_0$. Such an estimator will be said to be locally unique WPA1 if the event that there exists a unique maximum of $\mathscr{L}_n(\theta)$ on this intersection occurs WPA1. Specifically, we will show that the event that there exists a unique maximum of $\mathscr{L}_n(\theta)$ on $[N_n(A) \cap \Omega] \setminus M_n(A)$ occurs WPA1 for $A$ sufficiently large. For our first theorem, it suffices to replace (A2), (B3) and (B4) by the following weaker conditions.

(A2′) A subset $\Omega$ of $\Theta$ is said to satisfy (A2′) if $\Omega$ contains $\theta_0$, and if the intersection between $\Omega$ and a closed neighborhood $\mathscr{N}$ of $\theta_0$ is a closed subset of $\mathbb{R}^k$.

(B3′) There exists a constant $c > 0$ such that for each $A > 0$,

$$\mathbb{P}\Big\{\inf_{\theta \in N_n(A)} \lambda_{\min}\big\{\mathbf{G}_n^{-1/2}\mathbf{F}_n(\theta)\mathbf{G}_n^{-T/2}\big\} \le c\Big\} \to 0, \qquad n \to \infty.$$

(B4′) $\mathbf{G}_n^{-1/2}S_n(\theta_0)$ is tight, that is, $\lim_{A \to \infty} \limsup_{n \to \infty} \mathbb{P}\{|\mathbf{G}_n^{-1/2}S_n(\theta_0)| > A\} = 0$.

Theorem 2.1. *Let* $\Omega$ *be a subset of* $\Theta$ *satisfying* (A2′). *If conditions* (A1), (B1), (B2) *and* (B3′), (B4′) *hold, then a* ME $\hat{\theta}_n$ *of* $\theta_0$ *on* $\Omega$ *exists, is locally unique* WPA1, *and is consistent for* $\theta_0$.

Remark 2.1. It is possible that there are many maximum estimators. Theorem 2.1 says that among these maximum estimators there is an ME which is consistent and locally unique WPA1. This particular ME is in fact a global maximizer within a neighborhood of $\theta_0$.

Let $\Omega$ and $\tau$ be two fixed subsets of $\Theta$ which satisfy (A2) with corresponding $C_\Omega$ and $C_\tau$. Let $\mathbf{T}_n$ be arbitrary nonstochastic orthogonal matrices and define

$$\tilde{C}_{\Omega_n} = \left\{ \tilde{\theta} \colon \tilde{\theta} = \mathbf{T}_n \mathbf{G}_n^{T/2}(\theta - \theta_0),\ \theta \in C_\Omega \right\} \tag{2.8}$$

and similarly for $\tilde{C}_{\tau_n}$. Note that the orthogonal matrix $\mathbf{T}_n$ in the definition of $\tilde{C}_{\Omega_n}$ can be different from that of $\tilde{C}_{\tau_n}$. We need one more assumption on the behavior of the sets $\Omega$ and $\tau$.

(A3) A subset $\Omega$ of $\Theta$ is said to satisfy (A3) if there exists a closed cone $\tilde{C}_\Omega$ with vertex at 0, not depending on $n$, such that the sets $\tilde{C}_{\Omega_n}$ asymptotically coincide with $\tilde{C}_\Omega$ in the sense that as $n \to \infty$,

$$\sup_{|\beta|=1} \left| \inf_{\theta \in \tilde{C}_{\Omega_n}} |\beta - \theta|^2 - \inf_{\theta \in \tilde{C}_\Omega} |\beta - \theta|^2 \right| \to 0.$$

Let $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ be local maxima of $\mathscr{L}_n(\theta)$ on $\Omega$ and $\tau$ as obtained in Theorem 2.1. Define

$$d_n = 2\left[ \mathscr{L}_n\left(\hat{\theta}_n^2\right) - \mathscr{L}_n\left(\hat{\theta}_n^1\right) \right]. \tag{2.9}$$

Denote by $N = (N_1 \cdots N_k)^T$ a random vector which has a multivariate normal distribution with mean zero and covariance matrix $\mathbf{V}$.

THEOREM 2.2. *Suppose that* (A1) *holds and* (A2) *and* (A3) *hold for* $\Omega$ *and* $\tau$. *Suppose also that* (B1)–(B5) *hold. Then the asymptotic distribution of* $d_n$ *exists and is the same as the distribution of*

$$\inf_{\theta \in \tilde{C}_\Omega} |N - \theta|^2 - \inf_{\theta \in \tilde{C}_\tau} |N - \theta|^2. \tag{2.10}$$

REMARK 2.2. Although $\Omega$ and $\tau$ are assumed to satisfy (A2) in Theorems 2.1 and 2.2, the results still hold if (A2) is relaxed to the following assumption.

(A2$''$) A subset $\Omega$ of $\Theta$ is said to satisfy (A2$''$) if there exists a closed cone $C_\Omega$ with vertex at $\theta_0$ such that

$$\inf_{x \in C_\Omega} |\mathbf{G}_n^{T/2}(x - y)| \le u(y) |\mathbf{G}_n^{T/2}(y - \theta_0)|$$

and

$$\inf_{y \in \Omega} |\mathbf{G}_n^{T/2}(x - y)| \le v(x) |\mathbf{G}_n^{T/2}(x - \theta_0)|,$$

where the real functions $u(y)$ on $\Omega$ and $v(x)$ on $C_\Omega$ satisfy $u(y) \to 0$ as $y \to \theta_0$ and $v(x) \to 0$ as $x \to \theta_0$.

We omit here the proof that (A2) can be replaced by (A2$''$) and also the fact that (A2$''$) is equivalent to Chernoff regularity in the sense defined by Geyer (1994) if

$$\liminf_{n \to \infty} \frac{\lambda_{\min}\{\mathbf{G}_n\}}{\lambda_{\max}\{\mathbf{G}_n\}} > 0. \tag{2.11}$$

Thus Chernoff regularity is sufficient for our results, under the assumption that (2.11) holds [which is the case in Geyer (1994)]. Geyer provides counterexamples which show that Chernoff regularity is not sufficient to guarantee asymptotic results similar to those in Theorem 2.2 for some local maxima. But these examples do not apply to our case, as Theorem 2.2 only guarantees the asymptotic results for $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ as obtained in Theorem 2.1 (see Remark 2.1); other local maxima are not covered by Theorem 2.2.

Suppose for the remainder of this section that $\Theta = \Theta_1 \times \cdots \times \Theta_k$ where the $\Theta_i$'s are either closed, half open or open intervals. We also assume for the remainder of this section that $\mathbf{G}_n^{1/2}$ is the left Cholesky square root of $\mathbf{G}_n$ and that (A1) and (B1)–(B5) hold. Let $\chi_r^2$ be the chi-squared distribution on $r$ degrees of freedom.

In Theorem 2.3, we illustrate how to calculate the asymptotic distribution of $d_n$ when two components of $\theta_0$ are on the boundary of the intervals $\Theta_j$ and $\mathbf{G}_n$ is not diagonal. It will be applied in the next section to the two-way nested random variance components model. Suppose that the components $\theta_{j0}$, $j = k-1, k$, are on the boundaries of $\Theta_j$, which now have the form $(a_j, \theta_{j0}]$ or $[a_j, \theta_{j0}]$, say, with $a_j < \theta_{j0}$ for $j = k-1, k$ (the case $[\theta_{j0}, b_j)$ or $[\theta_{j0}, b_j]$ with $b_j > \theta_{j0}$ for $j = k-1$ or $j = k$ is similar), while the remaining components $\theta_{j0}$, $j = 1, \dots, k-2$, are interior points of $\Theta_j$. Suppose also that the components $\theta_{j0}$, $j = k-1, k$, are known while the components $\theta_{j0}$, $j = 1, \dots, k-2$, are to be estimated. In this setup, $C_\Omega = \mathbb{R}^{k-2} \times \{\theta_{(k-1)0}\} \times \{\theta_{k0}\}$ and $C_r = \mathbb{R}^{k-2} \times (-\infty, \theta_{(k-1)0}] \times (-\infty, \theta_{k0}]$. Let

$$\mathbf{G}_n^{T/2} = \begin{bmatrix} \mathbf{U}_n & \mathbf{V}_n \\ \mathbf{0} & \mathbf{W}_n \end{bmatrix}, \tag{2.12}$$

where $\mathbf{U}_n$ is a $(k-2) \times (k-2)$ upper triangular matrix, $\mathbf{V}_n$ is a $(k-2) \times 2$ matrix and

$$\mathbf{W}_n = \begin{bmatrix} a_n & c_n \\ 0 & b_n \end{bmatrix} \tag{2.13}$$

for some $a_n > 0$, $b_n > 0$. Suppose that

$$c_n / b_n \to x_0 \in [-\infty, \infty], \qquad n \to \infty. \tag{2.14}$$

When $x_0 \in (-\infty, \infty)$ let

$$\begin{aligned} f(N_{k-1}, N_k) = {} & (N_{k-1}^2 + N_k^2) 1_{\{N_{k-1} \geq 0,\, x_0 N_{k-1} + N_k \geq 0\}} \\ & + N_k^2 1_{\{N_{k-1} < 0,\, N_k \geq 0\}} \\ & + \frac{(N_{k-1} - x_0 N_k)^2}{1 + x_0^2} 1_{\{x_0 N_{k-1} + N_k < 0,\, N_{k-1} - x_0 N_k \geq 0\}}. \end{aligned} \tag{2.15}$$

When $x_0 = \infty$ or $x_0 = -\infty$ let

$$f(N_{k-1}, N_k) = \begin{cases} N_{k-1}^2 1_{\{N_{k-1} \geq 0\}} + N_k^2, & \text{if } x_0 = \infty; \\ N_k^2 1_{\{N_k \geq 0\}}, & \text{if } x_0 = -\infty. \end{cases} \tag{2.16}$$

THEOREM 2.3. *Suppose* (2.14) *holds. The asymptotic distribution of* $d_n$ *is the same as the distribution of* $(N_{k-1}^2 + N_k^2) - f(N_{k-1}, N_k)$.

**3. Variance component analysis—the two-way nested random model.** Recall from the introduction that we have $I$ classes where each class has $J_i$ members. We select a random sample of $K_i$ observations from the $j$th member of the $i$th class, $i = 1, \dots, I$, $j = 1, \dots, J_i$. Suppose that $J_i$ and $K_i$ are positive integers and that the $k'$th observation of the $j$th member from the $i$th class has the form defined by (1.1) and (1.2). It is assumed that the random vectors $(A_i, (B_{ij})_{1 \leq j \leq J_i}, (\varepsilon_{ijk'})_{1 \leq j \leq J_i, 1 \leq k' \leq K_i})$, $1 \leq i \leq I$, are independent. For each $i$, conditional on $A_i = a_i$, the random vectors $(B_{ij}, (\varepsilon_{ijk'})_{1 \leq k' \leq K_i})$, $1 \leq j \leq J_i$, are assumed to be independent, and for each pair $(i, j)$, conditional on $A_i = a_i$ and $B_{ij} = b_{ij}$, the random variables $\varepsilon_{ijk'}$, $1 \leq k' \leq K_i$, are assumed to be independent. Recall that

$$\theta = (\theta_1 \;\; \theta_2 \;\; \theta_3 \;\; \theta_4)^T = (\mu \;\; \sigma^2 \;\; \sigma_B^2 \;\; \sigma_A^2)^T \in \mathbb{R} \times (0, \infty) \times [0, \infty) \times [0, \infty)$$

is the parameter to be estimated. The log-likelihood ratio test that either $\sigma_A^2 = 0$ or $\sigma_B^2 = 0$ has as asymptotic distribution a 50–50 mixture between a chi-squared distribution on 1 degree of freedom and a point mass at zero. But suppose that we wish to test the hypothesis that the variance components $\sigma_A^2$ and $\sigma_B^2$ are both zero. Then let the true parameter be $\theta_0 = (\theta_{10} \;\; \theta_{20} \;\; \theta_{30} \;\; \theta_{40})^T = (\mu_0 \;\; \sigma_0^2 \;\; \sigma_{B0}^2 \;\; \sigma_{A0}^2)^T = (\mu_0 \;\; \sigma_0^2 \;\; 0 \;\; 0)^T$. Under this hypothesis, we will derive the asymptotic distribution of the log-likelihood ratio $d_I$ defined by (1.3).

Under the normality assumptions in Section 1, the observed likelihood is

$$L_I = \prod_{i=1}^{I} L_i, \tag{3.1}$$

where

$$\begin{aligned} L_i = \int \Bigg[ \prod_{j=1}^{J_i} \int \Bigg( \prod_{k'=1}^{K_i} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ \frac{-(y_{ijk'} - a_i - b_{ij})^2}{2\sigma^2} \right\} \Bigg) \\ \times \frac{1}{\sqrt{2\pi}\,\sigma_B} \exp\left\{ \frac{-b_{ij}^2}{2\sigma_B^2} \right\} db_{ij} \Bigg] \\ \times \frac{1}{\sqrt{2\pi}\,\sigma_A} \exp\left\{ \frac{-(a_i - \mu)^2}{2\sigma_A^2} \right\} da_i. \end{aligned} \tag{3.2}$$

Define

$$\bar{y}_{ij} = \left( \sum_{k'=1}^{K_i} y_{ijk'} \right) \Big/ K_i \quad \text{and} \quad \bar{y}_i = \left( \sum_{j=1}^{J_i} \bar{y}_{ij} \right) \Big/ J_i. \tag{3.3}$$

Then conditional on $B_{ij} = b_{ij}$ and $A_i = a_i$, the random variables $y_{ijk'}$, $1 \le k' \le K_i$, are distributed as $N(a_i + b_{ij}, \sigma^2)$. Conditional on $A_i = a_i$, the random variables $\bar{y}_{ij}$, $1 \le j \le J_i$, are distributed as $N(a_i, \sigma_B^2 + \sigma^2/K_i)$. Finally each of the random variables $\bar{y}_i$, $1 \le i \le I$, has the distribution $N(\mu, \sigma_A^2 + \sigma_B^2/J_i + \sigma^2/(J_i K_i))$. Define

$$W_{ij} = \sum_{k'=1}^{K_i} \left( y_{ijk'} - \bar{y}_{ij} \right)^2, \qquad \Phi_i^2 = \sigma^2 + K_i \sigma_B^2 \tag{3.4}$$

and

$$W_i = \sum_{j=1}^{J_i} \left( \bar{y}_{ij} - \bar{y}_i \right)^2 \quad \text{and} \quad \Psi_i^2 = \sigma^2 + K_i \sigma_B^2 + J_i K_i \sigma_A^2. \tag{3.5}$$

Then the observed log-likelihood can be written as

$$\begin{aligned} \mathscr{L}_I(\theta) = -\frac{1}{2} \sum_{i=1}^{I} \Bigg\{ & J_i(K_i - 1)\log \theta_2 + \frac{1}{\theta_2} \sum_{j=1}^{J_i} W_{ij} \\ & + (J_i - 1)\log(\theta_2 + K_i \theta_3) \\ & + \frac{K_i W_i}{\theta_2 + K_i \theta_3} + \log(\theta_2 + K_i \theta_3 + J_i K_i \theta_4) \\ & \qquad + \frac{J_i K_i (\bar{y}_i - \theta_1)^2}{\theta_2 + K_i \theta_3 + J_i K_i \theta_4} \Bigg\} + \text{const.} \end{aligned} \tag{3.6}$$

The derivative of $\mathscr{L}_i(\theta)$ with respect to $\theta$ is the vector

$$\begin{aligned} S_I(\theta) &= \frac{\partial}{\partial \theta} \mathscr{L}_I(\theta) = \sum_{i=1}^{I} s_i(\theta) \\ &= \sum_{i=1}^{I} \left[ s_{i1}(\theta) \;\; s_{i2}(\theta) \;\; s_{i3}(\theta) \;\; s_{i4}(\theta) \right]^T, \end{aligned} \tag{3.7}$$

where

$$s_{i1}(\theta) = \frac{J_i K_i (\bar{y}_i - \theta_1)}{\Psi_i^2}, \tag{3.8}$$

$$s_{i2}(\theta) = \frac{1}{2}\Bigg(\frac{-J_i(K_i-1)}{\theta_2} + \frac{\sum_{j=1}^{J_i} W_{ij}}{\theta_2^2} - \frac{J_i - 1}{\Phi_i^2} + \frac{K_i W_i}{\Phi_i^4} - \frac{1}{\Psi_i^2} + \frac{J_i K_i (\bar{y}_i - \theta_1)^2}{\Psi_i^4}\Bigg), \tag{3.9}$$

$$s_{i3}(\theta) = \frac{1}{2}\left(\frac{-(J_i-1)K_i}{\Phi_i^2} + \frac{K_i^2 W_i}{\Phi_i^4} - \frac{K_i}{\Psi_i^2} + \frac{J_i K_i^2 (\bar{y}_i - \theta_1)^2}{\Psi_i^4}\right) \tag{3.10}$$

and

$$s_{i4}(\theta) = \frac{1}{2}\left(\frac{-J_i K_i}{\Psi_i^2} + \frac{J_i^2 K_i^2 (\bar{y}_i - \theta_1)^2}{\Psi_i^4}\right). \tag{3.11}$$

Similarly, by differentiating again, one obtains

$$\mathbf{F}_I(\theta) = -\frac{\partial^2 \mathscr{L}_I(\theta)}{\partial \theta^2} = \sum_{i=1}^{I} [f_i^{rs}(\theta)],$$

where $[f_i^{rs}(\theta)]$ is the negative of the second derivative of the log-likelihood of the $i$th observation. Then it can be checked directly that

$$\mathbb{E}\{S_I(\theta_0)\} = 0 \quad \text{and} \quad \mathbb{E}\{S_I(\theta_0) S_I^T(\theta_0)\} = \mathbb{E}\{\mathbf{F}_I(\theta_0)\}. \tag{3.12}$$

Furthermore, defining $\Phi_{i0}^2 = \theta_{20} + K_i \theta_{30}$ and $\Psi_{i0}^2 = \theta_{20} + K_i \theta_{30} + J_i K_i \theta_{40}$, it can be seen that

$$\mathbf{D}_I = \mathbb{E}\{\mathbf{F}_I(\theta_0)\} = \sum_{i=1}^{I} \mathscr{D}_i = \sum_{i=1}^{I} [d_i^{rs}], \tag{3.13}$$

where $\mathscr{D}_i = \mathbb{E}\{[f_i^{rs}(\theta_0)]\}$ is the following matrix:

$$\begin{bmatrix} \frac{J_i K_i}{\Psi_{i0}^2} & 0 & 0 & 0 \\ 0 & \frac{1}{2}\left(\frac{J_i(K_i-1)}{\theta_{20}^2} + \frac{J_i-1}{\Phi_{i0}^4} + \frac{1}{\Psi_{i0}^4}\right) & \frac{K_i}{2}\left(\frac{J_i-1}{\Phi_{i0}^4} + \frac{1}{\Psi_{i0}^4}\right) & \frac{J_i K_i}{2\Psi_{i0}^4} \\ 0 & \frac{K_i}{2}\left(\frac{J_i-1}{\Phi_{i0}^4} + \frac{1}{\Psi_{i0}^4}\right) & \frac{K_i^2}{2}\left(\frac{J_i-1}{\Phi_{i0}^4} + \frac{1}{\Psi_{i0}^4}\right) & \frac{J_i K_i^2}{2\Psi_{i0}^4} \\ 0 & \frac{J_i K_i}{2\Psi_{i0}^4} & \frac{J_i K_i^2}{2\Psi_{i0}^4} & \frac{J_i^2 K_i^2}{2\Psi_{i0}^4} \end{bmatrix}. \tag{3.14}$$

[This agrees with (104) in Searle, Casella and McCulloch (1992), page 157.] Now suppose that $\theta_{30} = \theta_{40} = 0$. Then $\Phi_{i0}^2 = \Psi_{i0}^2 = \theta_{20}$ and the left Cholesky square root of $\mathbf{D}_I$ is

$$\mathbf{D}_I^{1/2} = \begin{bmatrix} \sqrt{2\theta_{20}}\, z_I & 0 & 0 & 0 \\ 0 & z_I & 0 & 0 \\ 0 & z_I & a_I & 0 \\ 0 & z_I & c_I & b_I \end{bmatrix}, \tag{3.15}$$

where $z_I = (\sum_{i=1}^I J_i K_i / [2\theta_{20}^2])^{1/2}$,

$$a_I = c_I = \left( \frac{\sum_{i=1}^I J_i (K_i - 1) K_i}{2\theta_{20}^2} \right)^{1/2}, \qquad b_I = \left( \frac{\sum_{i=1}^I (J_i - 1) J_i K_i^2}{2\theta_{10}^2} \right)^{1/2}. \tag{3.16}$$

Suppose that

$$\lim_{n \to \infty} \left( \frac{\sum_{i=1}^I J_i (K_i - 1) K_i}{\sum_{i=1}^I (J_i - 1) J_i K_i^2} \right)^{1/2} = x_0 \in [0, \infty]. \tag{3.17}$$

THEOREM 3.1. *Suppose that* (3.17) *holds*, $J_i > 1$, $K_i > 1$ *and* $\{J_i\}$ *and* $\{K_i\}$ *are bounded above. Then the asymptotic distribution of* $d_I$, *as* $I \to \infty$, *is* $(N_3^2 + N_4^2) - f(N_3, N_4)$ *where* $f(N_3, N_4)$ *is given by* (2.15) *with* $k = 4$.

The above model corresponds to a design which may be unbalanced if the integers $J_i$ (or $K_i$) are not the same for all classes. Designs like this have been used often in agriculture. For example, one can consider $I$ litters with $J_i$ pigs in each litter and $K_i$ observations on each pig. Then the litters and the pigs in each litter correspond to the classes and members in each class. Suppose for example that the classes are divided into $m$ groups, and the classes in each group have the same numbers $J_i$ and $K_i$ of members and of observations on each member. Then (3.17) holds if, as $I$ tends to infinity, the limits of the proportions of classes belonging to a group out of the $I$ classes exist. The test given by Theorem 3.1 is the test of whether there is no variation among observations of members of each class. The percentage points of the asymptotic distribution given by Theorem 3.1 are easily approximated by simulation with either the exact value of $x_0$ obtained from the design or its approximated value obtained from the experiment.

REMARK 3.1. We give an example where the limiting proportions in classes do not exist. Suppose again that there are $I$ classes with $J_i$ members in each class and $K_i$ observations on each member. Also suppose that there are two groups of classes with $K_i = 2$ if the $i$th class belongs to the first group, and $K_i = 3$, otherwise, and that $J_i = 2$ for $1 \le i \le I$. Let $p_I$ be the proportion of the classes belonging to the first group out of the $I$ classes. Then

$$\left(\frac{\sum_{i=1}^{I} J_i(K_i - 1)K_i}{\sum_{i=1}^{I}(J_i - 1)J_i K_i^2}\right)^{1/2} = \left(\frac{2p_I + 6(1 - p_I)}{4p_I + 9(1 - p_I)}\right)^{1/2}. \tag{3.18}$$

One can easily construct a sequence $\{p_I\}$ with the property that there are two subsequences $\{p_{I'}\}$ and $\{p_{I''}\}$ of $\{p_I\}$ such that the right-hand side of (3.18) tends to two different limits $x_0'$ and $x_0''$. Then if the observations are assumed to satisfy (1.1) and (1.2) and the other assumptions given at the beginning of this section, the deviances $d_{I'}$ and $d_{I''}$ tend in distribution to two different limits as $I' \to \infty$ and $I'' \to \infty$ respectively. Thus the asymptotic distribution of $d_I$ does not exist. Note that (A3) does not hold in this case.

## 4. Proofs.

PROOF OF THEOREM 2.1. Assume (A1), (A2′), (B1), (B2) and (B3′), (B4′) hold. Under (A1), $\mathscr{L}_n(\theta)$ is continuous on $\Omega \cap \mathscr{N}$ for a neighborhood $\mathscr{N}$ of $\theta_0$. Fix $A > 0$. By (B2), $N_n(A) \subseteq \mathscr{N}$ for $n$ large enough, and since $\Omega \cap N_n(A)$ is closed by (A2′), $\mathscr{L}_n(\theta)$ must have a maximum on $\Omega \cap N_n(A)$. We will prove that

$$\lim_{A\to\infty} \liminf_{n\to\infty} \mathbb{P}\{\mathscr{L}_n(\theta) < \mathscr{L}_n(\theta_0) \quad \text{for all } \theta \in M_n(A)\} = 1. \tag{4.1}$$

Since $\inf_{\theta \in N_n(A)} \lambda_{\min}(\mathbf{F}_n(\theta)) \to_P \infty$ by (B2) and (B3′), $\mathscr{L}_n(\theta)$ is concave on $\Theta \cap N_n(A)$ WPA1. Thus it follows from (4.1) that there exists a unique maximum of $\mathscr{L}_n(\theta)$ on $[N_n(A) \cap \Omega] \setminus M_n(A)$ WPA1. To prove (4.1), let $\theta \in M_n(A)$. It follows from Taylor expansion that there exists some $\lambda \in [0, 1]$, depending on $\theta$, such that

$$\mathscr{L}_n(\theta) - \mathscr{L}_n(\theta_0) = (\theta - \theta_0)^T S_n(\theta_0) - \tfrac{1}{2}(\theta - \theta_0)^T \mathbf{F}_n(\tilde{\theta}_n)(\theta - \theta_0), \tag{4.2}$$

where $\tilde{\theta}_n = \lambda\theta_0 + (1 - \lambda)\theta$. Define $Q(\theta) = \frac{1}{2}(\theta - \theta_0)^T \mathbf{F}_n(\tilde{\theta}_n)(\theta - \theta_0)$. Observe that

$$\begin{aligned} &\mathbb{P}\{\mathscr{L}_n(\theta) \ge \mathscr{L}_n(\theta_0) \text{ for some } \theta \in M_n(A)\} \\ &\quad \le \mathbb{P}\{(\theta - \theta_0)^T S_n(\theta_0) \ge Q(\theta), Q(\theta) > cA^2/2 \text{ for some } \theta \in M_n(A)\} \\ &\qquad + \mathbb{P}\{Q(\theta) \le cA^2/2 \text{ for some } \theta \in M_n(A)\}, \end{aligned} \tag{4.3}$$

where $c$ is defined in (B3′). Denote $v_n(\theta) = (1/A)\mathbf{G}_n^{T/2}(\theta - \theta_0)$. Then $v_n(\theta)$ is a unit vector for each $\theta \in M_n(A)$. For the first term of (4.3), it follows from (B4′) that

$$\begin{aligned} &\limsup_{n\to\infty} \mathbb{P}\{(\theta - \theta_0)^T S_n(\theta_0) \ge Q(\theta), Q(\theta) > cA^2/2 \text{ for some } \theta \in M_n(A)\} \\ &\quad \le \limsup_{n\to\infty} \mathbb{P}\{v_n^T(\theta)\mathbf{G}_n^{-1/2} S_n(\theta_0) > cA/2 \text{ for some } \theta \in M_n(A)\} \\ &\quad \le \limsup_{n\to\infty} \mathbb{P}\{|\mathbf{G}_n^{-1/2} S_n(\theta_0)| > cA/2\} \to 0, \qquad A \to \infty. \end{aligned} \tag{4.4}$$

For the second term of (4.3), it follows from (B3′) that

$$
\begin{aligned}
&\mathbb{P}\{Q(\theta) \le cA^2/2 \text{ for some } \theta \in M_n(A)\}\\
&\quad = \mathbb{P}\Big\{\exists\, \theta \in M_n(A) \colon (\theta - \theta_0)^T \mathbf{G}_n^{1/2}\big[\mathbf{G}_n^{-1/2}\mathbf{F}_n(\tilde{\theta}_n)\mathbf{G}_n^{-T/2}\big]\\
&\qquad\qquad\qquad\qquad \times \mathbf{G}_n^{T/2}(\theta - \theta_0) \le cA^2\Big\} \qquad (4.5)\\
&\quad = \mathbb{P}\big\{v_n^T(\theta)\mathbf{G}_n^{-1/2}\mathbf{F}_n(\tilde{\theta}_n)\mathbf{G}_n^{-T/2}v_n(\theta) \le c \text{ for some } \theta \in M_n(A)\big\}\\
&\quad \le \mathbb{P}\Big\{\inf_{\theta \in N_n(A)} \lambda_{\min}\{\mathbf{G}_n^{-1/2}\mathbf{F}_n(\theta)\mathbf{G}_n^{-T/2}\} \le c\Big\} \to 0, \qquad n \to \infty.
\end{aligned}
$$

Hence (4.1) follows from (4.3)–(4.5). □

PROOF OF THEOREM 2.2. Suppose that (A1) and (B1)–(B5) hold and (A2) and (A3) hold for $\Omega$ and $\tau$. Let $\varepsilon > 0$ be given. Recall that $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ are local maxima of $\mathscr{L}_n(\theta)$ on $\Omega$ and $\tau$ respectively as obtained in Theorem 2.1. Note that (B4) implies (B4′), because by the Markov inequality,

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}\{|\mathbf{G}_n^{1/2}S_n(\theta_0)| > A\}\\
&\qquad \le \limsup_{n\to\infty} \frac{1}{A^2}\mathbb{E}\{S_n^T(\theta_0)\mathbf{G}_n^{-1}S_n(\theta_0)\}\\
&\qquad \le \limsup_{n\to\infty} \frac{1}{A^2}\mathbb{E}\{\mathrm{tr}[\mathbf{G}_n^{-1/2}S_n(\theta_0)S_n^T(\theta_0)\mathbf{G}_n^{-T/2}]\}\\
&\qquad = \frac{1}{A^2}\,\mathrm{tr}\{\mathbf{V}\} \to 0, \qquad A \to \infty.
\end{aligned}
$$

Also it is obvious that (B3) implies (B3′). Therefore, by (B4) and (B5) and Theorem 2.1, there exists a constant $A$, depending on $\varepsilon$, such that for $n$ large enough,

$$
\mathbb{P}\big\{\hat{\theta}_n^1 \in N_n(A),\ \hat{\theta}_n^2 \in N_n(A), \quad \text{and} \quad |\mathbf{G}_n^{-1/2}S_n(\theta_0)| \le A/2\big\} > 1 - \varepsilon. \qquad (4.6)
$$

Suppose that $\hat{\theta}_n^1$ and $\hat{\theta}_n^2$ are in $N_n(A)$ and $|\mathbf{G}_n^{1/2}S_n(\theta_0)| \le A/2$. By Taylor expansion, there exists $\lambda \in [0,1]$ such that

$$
\begin{aligned}
&2\big[\mathscr{L}_n(\hat{\theta}_n^1) - \mathscr{L}_n(\theta_0)\big]\\
&\qquad = 2(\hat{\theta}_n^1 - \theta_0)^T S_n(\theta_0) - (\hat{\theta}_n^1 - \theta_0)^T \mathbf{F}_n(\tilde{\theta}_n)(\hat{\theta}_n^1 - \theta_0),
\end{aligned} \qquad (4.7)
$$

where $\tilde{\theta}_n = \lambda\theta_0 + (1-\lambda)\hat{\theta}_n^1$. The expression (4.7) can be rewritten as

$$
2\big[\mathscr{L}_n(\hat{\theta}_n^1) - \mathscr{L}_n(\theta_0)\big] = h_n(\hat{\theta}_n^1) + r_n(\hat{\theta}_n^1), \qquad (4.8)
$$

where

$$
h_n(\theta) = -|\mathbf{G}_n^{-1/2}S_n(\theta_0) - \mathbf{G}_n^{T/2}(\theta - \theta_0)|^2 + S_n^T(\theta_0)\mathbf{G}_n^{-1}S_n(\theta_0)
$$

and

$$r_n(\theta) = (\theta - \theta_0)^T \left[\mathbf{G}_n - \mathbf{F}_n(\tilde{\theta}_n)\right](\theta - \theta_0).$$

Denote by $\tilde{\theta}_n^1$ the value that maximizes the quadratic function $h_n(\theta)$ on $N_n(A) \cap \Omega$. Since $\hat{\theta}_n^1$ maximizes $\mathscr{L}_n(\theta)$ on $N_n(A) \cap \Omega$, and $h_n(\hat{\theta}_n^1) \leq h_n(\tilde{\theta}_n^1)$, it follows from (4.8) that

$$0 \leq 2\left[\mathscr{L}_n(\hat{\theta}_n^1) - \mathscr{L}_n(\tilde{\theta}_n^1)\right] \leq r_n(\hat{\theta}_n^1) - r_n(\tilde{\theta}_n^1). \tag{4.9}$$

By (B3), there exists a $k \times k$ real symmetric matrix $\mathbf{V}_n(\tilde{\theta}_n) = o_p(1)$ such that

$$\mathbf{F}_n(\tilde{\theta}_n) = \mathbf{G}_n + \mathbf{G}_n^{1/2}\mathbf{V}_n(\tilde{\theta}_n)\mathbf{G}_n^{T/2}.$$

Since $\hat{\theta}_n^1 \in N_n(A)$,

$$\begin{aligned} |r_n(\hat{\theta}_n^1)| &\leq |\lambda|_{\max}\left\{\mathbf{V}_n(\tilde{\theta}_n)\right\}(\hat{\theta}_n^1 - \theta_0)^T \mathbf{G}_n(\hat{\theta}_n^1 - \theta_0) \\ &\leq |\lambda|_{\max}\left\{\mathbf{V}_n(\tilde{\theta}_n)\right\}A^2, \end{aligned}$$

where $|\lambda|_{\max}(\cdot)$ is the maximum absolute eigenvalue of a symmetric matrix. This implies that $r_n(\hat{\theta}_n^1) = o_p(1)$. Similarly, $r_n(\tilde{\theta}_n^1) = o_p(1)$ as $\tilde{\theta}_n^1 \in N_n(A)$. Therefore, it follows from (4.9) that

$$2\left[\mathscr{L}_n(\hat{\theta}_n^1) - \mathscr{L}_n(\theta_0)\right] = o_p(1) + 2\left[\mathscr{L}_n(\tilde{\theta}_n^1) - \mathscr{L}_n(\theta_0)\right]. \tag{4.10}$$

By the Taylor expansion, there exists $\beta \in [0, 1]$ such that

$$2\left[\mathscr{L}_n(\tilde{\theta}_n^1) - \mathscr{L}_n(\theta_0)\right] = h_n(\tilde{\theta}_n^1) + (\tilde{\theta}_n^1 - \theta_0)^T \left[\mathbf{G}_n - \mathbf{F}_n(\breve{\theta}_n)\right](\tilde{\theta}_n^1 - \theta_0),$$

where $\breve{\theta}_n = \beta\theta_0 + (1 - \beta)\tilde{\theta}_n^1$. Since $\tilde{\theta}_n^1$ is in $N_n(A)$, it can be easily verified using the same argument as above that

$$\begin{aligned} 2\left[\mathscr{L}_n(\tilde{\theta}_n^1) - \mathscr{L}_n(\theta_0)\right] &= h_n(\tilde{\theta}_n^1) + o_p(1) \\ &= -\inf_{\theta \in N_n(A) \cap \Omega} \left|\mathbf{G}_n^{-1/2} S_n(\theta_0) - \mathbf{G}_n^{T/2}(\tilde{\theta}_n^1 - \theta_0)\right| \\ &\quad + S_n^T(\theta_0)\mathbf{G}_n^{-1} S_n(\theta_0) + o_p(1). \end{aligned} \tag{4.11}$$

Transforming from $\theta$ to $\theta' = \mathbf{T}_n \mathbf{G}_n^{T/2}(\theta - \theta_0)$ [so that $\theta \in N_n(A)$ if and only if $|\theta'| \leq A$], it follows from (A2) and (B2) that for all $n$ large enough,

$$\begin{aligned} &\inf_{\theta \in N_n(A) \cap \Omega} |\mathbf{G}_n^{-1/2} S_n(\theta_0) - \mathbf{G}_n^{T/2}(\theta - \theta_0)|^2 \\ &\quad = \inf_{|\theta'| \leq A,\ \theta' \in \tilde{C}_{\Omega_n}} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta'|^2. \end{aligned} \tag{4.12}$$

Recall that we assume $|\mathbf{G}_n^{-1/2} S_n(\theta_0)| \leq A/2$, so $|\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0)| \leq A/2$. Since $\tilde{C}_{\Omega_n}$ contains the origin, we have

$$\inf_{\theta \in \tilde{C}_{\Omega_n}} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2 \leq |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0)|^2 \leq A^2/4. \tag{4.13}$$

There also exists $\dot{\theta} \in \tilde{C}_{\Omega_n}$ such that

$$|\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \dot{\theta}|^2 = \inf_{\theta \in \tilde{C}_{\Omega_n}} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2. \tag{4.14}$$

Since (4.13) and (4.14) imply that $|\dot{\theta}| \le A$, we have

$$\inf_{|\theta| \le A,\, \theta \in \tilde{C}_{\Omega_n}} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2 = \inf_{\theta \in \tilde{C}_{\Omega_n}} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2. \tag{4.15}$$

Thus it is derived from (4.10)–(4.12), (4.15) and (A3) that

$$\begin{aligned} 2\left[\mathscr{L}_n\left(\hat{\theta}_n^1\right) - \mathscr{L}_n(\theta_0)\right] = {} & - \inf_{\theta \in \tilde{C}_\Omega} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2 \\ & + S_n^T(\theta_0) \mathbf{G}_n^{-1} S_n(\theta_0) + o_p(1). \end{aligned} \tag{4.16}$$

Similarly, we have

$$\begin{aligned} 2\left[\mathscr{L}_n\left(\hat{\theta}_n^2\right) - \mathscr{L}_n(\theta_0)\right] = {} & - \inf_{\theta \in \tilde{C}_\tau} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2 \\ & + S_n^T(\theta_0) \mathbf{G}_n^{-1} S_n(\theta_0) + o_p(1). \end{aligned} \tag{4.17}$$

It follows from (4.16) and (4.17) that on a set whose probability exceeds $1 - 2\varepsilon$,

$$\begin{aligned} 2\left[\mathscr{L}_n\left(\hat{\theta}_n^2\right) - \mathscr{L}_n\left(\hat{\theta}_n^1\right)\right] = {} & \inf_{\theta \in \tilde{C}_\Omega} |\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0) - \theta|^2 \\ & - \inf_{\theta \in \tilde{C}_\tau} |\mathbf{T}_n \mathbf{G}_n^{1/2} S_n(\theta_0) - \theta|^2 + o_p(1). \end{aligned} \tag{4.18}$$

Since $\mathbf{T}_n \mathbf{G}_n^{-1/2} S_n(\theta_0)$ is asymptotically normally distributed with mean 0 and covariance matrix $\mathbf{V}$ by (B4) and (B5), Theorem 2.2 follows from (4.18) and the continuous mapping theorem. □

REMARK 4.1. If $C_\Omega$ and $C_\tau$ are the approximating cones for $\Omega$ and $\tau$ in the sense defined in Remark 2.2, our results are proved to be valid by verifying that (4.12) still holds in this case.

PROOF OF THEOREM 2.3. Take the matrices $\mathbf{T}_n$ to be identity $k$-dimensional matrices. We have $\tilde{C}_\Omega = \tilde{C}_{\Omega_n} = \mathbf{T}_n \mathbf{G}_n^{T/2}(C_\Omega - \theta_0) = \mathbb{R}^{k-2} \times \{0\} \times \{0\}$ and $\tilde{C}_{\tau_n} = \mathbf{T}_n \mathbf{G}_n^{T/2}(C_\tau - \theta_0) = \{\theta = (\theta_1 \cdots \theta_k)^T \in \mathbb{R}^k\colon b_n \theta_{k-1} - c_n \theta_k \le 0,\ \theta_k \le 0\}$. Hence (A3) holds for $\Omega$ and

$$\inf_{\theta \in \tilde{C}_\Omega} |N - \theta|^2 = N_{k-1}^2 + N_k^2. \tag{4.19}$$

Suppose that $c_n / b_n \to x_0 \in (-\infty, \infty)$ as $n \to \infty$. Then we will prove that the sets $\tilde{C}_{\tau_n}$ asymptotically coincide with

$$\tilde{C}_\tau = \left\{\theta = (\theta_1 \cdots \theta_k)^T \in \mathbb{R}^k\colon \theta_{k-1} - x_0 \theta_k \le 0,\ \theta_k \le 0\right].$$

Let $\alpha_n$, $\alpha_0 \in (0, \pi)$ be such that $\cot(\alpha_n) = c_n / b_n$ and $\cot(\alpha_0) = x_0$. Since $c_n / b_n \to x_0$ as $n \to \infty$, we have $\alpha_n \to \alpha_0$ as $n \to \infty$. Since $\tilde{C}_{\tau_n}$ and $\tilde{C}_\tau$ contain the origin, it can be easily verified that as $n \to \infty$,

$$\sup_{|\beta|=1} \left| \inf_{\theta \in \tilde{C}_{\tau_n}} |\beta - \theta|^2 - \inf_{\theta \in \tilde{C}_{\tau}} |\beta - \theta|^2 \right|$$

$$\leq \max\{|\sin(\alpha_0 - \alpha_n)|, (1 - \cos(\alpha_0 - \alpha_n))\} \to 0.$$

Thus the sets $\tilde{C}_{\tau_n}$ asymptotically coincide with $\tilde{C}_\tau$ and (A3) holds for $\tau$.

Now the projection of the set $\tilde{C}_\tau$ onto the plane whose coordinates are $(\theta_{k-1}\theta_k)^T$ is given as Region 4 in Figure 1. In the figure we have

$$\inf_{\theta \in \tilde{C}_\tau} |N - \theta|^2 = \begin{cases} N_{k-1}^2 + N_k^2, & \text{if } (N_{k-1}, N_k) \text{ is in Region 1}, \\ N_k^2, & \text{if } (N_{k-1}, N_k) \text{ is in Region 2}, \\ (N_{k-1} - x_0 N_k)^2/(1 + x_0^2), & \text{if } (N_{k-1}, N_k) \text{ is in Region 3}, \\ 0, & \text{if } (N_{k-1}, N_k) \text{ is in Region 4}, \end{cases}$$

where

$$\text{Region 1} = \{(\theta_{k-1}\theta_k)^T : \theta_{k-1} \geq 0,\ x_0\theta_{k-1} + \theta_k \geq 0\};$$

$$\text{Region 2} = \{(\theta_{k-1}\theta_k)^T : \theta_{k-1} < 0,\ \theta_k \geq 0\};$$

$$\text{Region 3} = \{(\theta_{k-1}\theta_k)^T : x_0\theta_{k-1} + \theta_k < 0,\ \theta_{k-1} - x_0\theta_k \geq 0\};$$

$$\text{Region 4} = \{(\theta_{k-1}\theta_k)^T : \theta_{k-1} - x_0\theta_k \leq 0,\ \theta_k \leq 0\}.$$

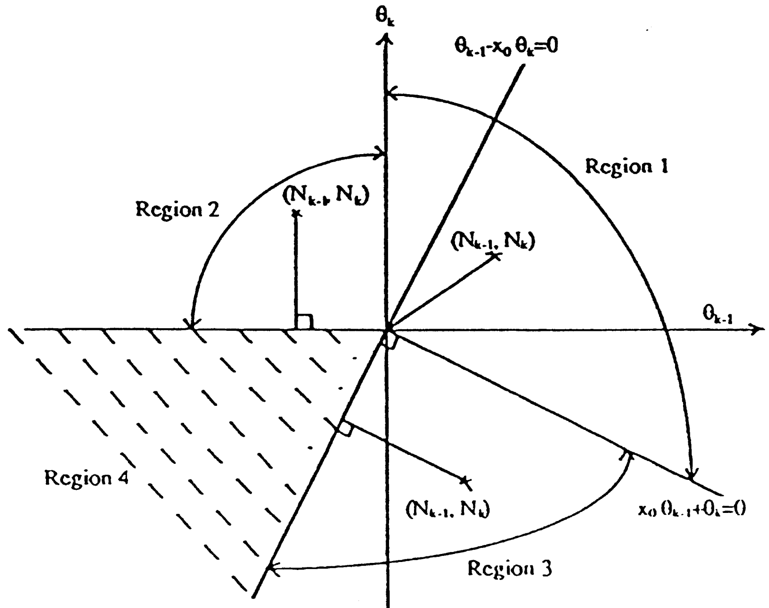


FIG. 1. *The distance of* $(N_{k-1}, N_k)$ *to Region* 4.

It can be calculated that

$$\inf_{\theta \in \tilde{C}_\tau} |N - \theta|^2 = f(N_{k-1}, N_k) \tag{4.20}$$

where $f(N_{k-1}, N_k)$ is given by (2.15). Thus it follows from (4.19) and (4.20) and Theorem 2.2 that the asymptotic distribution of $d_n$ is $N_{k-1}^2 + N_k^2 - f(N_{k-1}, N_k)$.

Suppose that $c_n/b_n \to x_0$ as $n \to \infty$ with $|x_0| = \infty$. Then the sets $\tilde{C}_{\tau_n}$ asymptotically coincide with

$$\tilde{C}_\tau = \begin{cases} \{\theta \in \mathbb{R}^k \colon \theta_{k-1} \le 0,\ \theta_k = 0\}, & \text{if } x_0 = \infty; \\ \{\theta \in \mathbb{R}^k \colon \theta_k \le 0\}, & \text{if } x_0 = -\infty. \end{cases}$$

Thus (A3) holds and $\inf_{\theta \in \tilde{C}_\tau} |N - \theta|^2 = f(N_{k-1}, N_k)$ where $f(N_{k-1}, N_k)$ is given by (2.16). Hence it follows from Theorem 2.2 that the asymptotic distribution of $d_n$ is again $N_{k-1}^2 + N_k^2 - f(N_{k-1}, N_k)$. □

REMARK 4.2. In the above proof, it is assumed that $\theta_{(k-1)0}$ and $\theta_{k0}$ are the right end points of their admissible intervals. If $\Theta_j = [\theta_{j0}, b_j)$ or $\Theta_j = [\theta_{j0}, b_j]$ with $\theta_{j0} < b_j$ for some $j = k-1, k$, then replace $(-\infty, \theta_{j0}]$ by $[\theta_{j0}, \infty)$ in the expression for $C_\tau$ and let $(\mathbf{T}_n)_{jj} = -1$.

PROOF OF THEOREM 3.1. Suppose that (3.17) holds, $J_i > 1$, $K_i > 1$ and $\{J_i\}$ and $\{K_i\}$ are bounded above. Then $x_0$ must be finite. Thus Theorem 2.3 applies to give the asymptotic distribution for the log-likelihood ratio $d_I$, defined by (1.3), with $a_I$, $b_I$ and $c_I$ given by (3.16), $n = I$, $k = 4$ and $f(N_{k-1}, N_k)$ given by (2.15), provided that (A1)–(A3) and (B1)–(B5) are verified. Here $\Omega = \mathbb{R} \times (0, \infty) \times \{0\} \times \{0\}$ and $\tau = \mathbb{R} \times (0, \infty) \times [0, \infty) \times [0, \infty)$. So (A1)–(A3) hold. Since $\mathbf{D}_n = \mathbf{G}_n$, (B4) is trivially satisfied. Since $\det(\mathscr{D}_i) = J_i^4 K_i^5 (J_i - 1)(K_i - 1)/(8\theta_{20}^7) \ne 0$, there exists $c_0 > 0$ such that

$$\lambda_{\min}(\mathscr{D}_i) \ge c_0, \qquad 1 \le i \le I. \tag{4.21}$$

Thus $\lambda_{\min}\{\mathbf{D}_I\} \ge c_0 I$, and so (B1) and (B2) hold. It remains to verify (B3) and (B5). This is done as follows. Let $k = 4$. Let $\theta \in N_I(A)$ and write

$$\begin{aligned} &\mathbf{D}_I^{-1/2} \mathbf{F}_I(\theta) \mathbf{D}_I^{-T/2} \\ &\quad = \mathbf{I}_k + \mathbf{D}_I^{-1/2}\{\mathbf{F}_I(\theta_0) - \mathbf{D}_I\}\mathbf{D}_I^{-T/2} + \mathbf{D}_I^{-1/2}\{\mathbf{F}_I(\theta) - \mathbf{F}_I(\theta_0)\}\mathbf{D}_I^{-T/2}. \end{aligned}$$

If $\mathbf{A}_I$ is a $k \times k$ matrix such that $\|I^{-1}\mathbf{A}_I\|_1 = o_p(1)$ as $I \to \infty$, then it follows from (4.21) that for any unit vector $u$,

$$\begin{aligned} |u^T \mathbf{D}_I^{-1/2} \mathbf{A}_I \mathbf{D}_I^{-T/2} u| &\le |\lambda|_{\max}\{\mathbf{A}_I\} u^T \mathbf{D}_I^{-1} u \\ &\le |\lambda|_{\max}\{\mathbf{A}_I\} \lambda_{\max}\{\mathbf{D}_I^{-1}\} = |\lambda|_{\max}\{\mathbf{A}_I\} \lambda_{\min}^{-1}\{\mathbf{D}_I\} \\ &\le (c_0 I)^{-1} |\lambda|_{\max}\{\mathbf{A}_I\} = o_p(1). \end{aligned} \tag{4.22}$$

Thus (B3) holds if

$$\operatorname{tr}\left\{\frac{\mathbf{F}_I(\theta_0) - \mathbf{D}_I}{I}\right\} = o_p(1),$$

(4.23)

$$\sup_{\theta \in N_I(A)} \operatorname{tr}\left\{\frac{\mathbf{F}_I(\theta) - \mathbf{F}_I(\theta_0)}{I}\right\} = o_p(1).$$

Since $\{J_i\}$ and $\{K_i\}$ are bounded above, $\{f_i^{rs}(\theta)\}$ is equicontinuous at $\theta_0$ and $\{\mathbb{E}\{(f_i^{rs}(\theta_0) - d_i^{rs})^2\}\}$ is uniformly bounded for $r, s = 1, 2$. Thus (4.23) holds and so does (B3).

Let $\xi_I$ be any unit vector in $\mathbb{R}^k$. For $1 \le i \le I$, define $Y_{iI} = \xi_I^T \mathbf{D}_I^{-1/2} s_i(\theta_0)$ and $\sigma_{iI}^2 = \operatorname{Var}\{Y_{iI}\} = \xi_I^T \mathbf{D}_I^{-1/2} \mathscr{D}_i \mathbf{D}_I^{-T/2} \xi_I$. Then $Y_{iI}$, $1 \le i \le I$, are mutually independent for each $I$, $\mathbb{E}\{Y_{iI}\} = 0$, and $\sigma_I^2 = \sigma_{1I}^2 + \cdots + \sigma_{II}^2 = 1$. Since $\mathbb{E}\{|s_i(\theta_0)|^\delta\} \le L$ with $\delta = 5/2$, $1 \le i \le I$, for some constant $L$, it follows from (4.21) that as $I \to \infty$,

$$\begin{aligned}
\sum_{i=1}^{I} \mathbb{E}\left\{|Y_{iI}|^\delta\right\} &= \sum_{i=1}^{I} \mathbb{E}\left\{\left(\xi_I^T \mathbf{D}_I^{-1/2} s_i(\theta_0) s_i^T(\theta_0) \mathbf{D}_I^{-T/2} \xi\right)^{\delta/2}\right\} \\
&\le \sum_{i=1}^{I} \mathbb{E}\left\{\lambda_{\max}^{\delta/2}\left\{s_i(\theta_0) s_i^T(\theta_0)\right\}\left(\xi^T \mathbf{D}_I^{-1} \xi\right)^{\delta/2}\right\} \\
&\le \sum_{i=1}^{I} \mathbb{E}\left\{\left(\operatorname{tr}\left\{s_i(\theta_0) s_i^T(\theta_0)\right\}\right)^{\delta/2} \lambda_{\max}^{\delta/2}\left\{\mathbf{D}_I^{-1}\right\}\right\} \\
&= \sum_{i=1}^{I} \frac{\mathbb{E}\left\{|s_i(\theta_0)|^\delta\right\}}{\lambda_{\min}^{\delta/2}\{\mathbf{D}_I\}} \le \frac{LI}{(c_0 I)^{\delta/2}} \to 0.
\end{aligned}$$

Thus, by Lyapounov's theorem in Billingsley [(1968), page 44], $\xi_I^T \mathbf{D}_I^{-1/2} S_I(\theta_0) = Y_{1I} + \cdots + Y_{II}$ converges in distribution to the standard normal random variable. This verifies (B5). □

**Acknowledgments.** We thank Professor Murray Aitkin for his useful advice on variance component analysis, Professor Terry Speed for helpful advice and a referee for drawing our attention to Geyer (1995). We also thank Associate Professor Ross Maller for advice and assistance above and beyond the call of duty.

## REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

CHANT, D. (1974). On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika* **61** 291–298.

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

CROWDER, M. (1990). On some nonregular tests for a modified Weibull model. *Biometrika* **77** 449–506.

FAHRMEIR, L. and KAUFMANN, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.* **13** 342–368.

FEDER, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is 'near' the boundaries of the hypothesis region. *Ann. Math. Statist.* **39** 2044–2055.

GEYER, C. J. (1994). On the asymptotics of constrained M-estimation. *Ann. Statist.* **22** 1993–2010.

GHITANY, M. E., MALLER, R. A. and ZHOU, S. (1994). Exponential mixture models with long-term survivors and covariates. *J. Multivariate Anal.* **49** 218–241.

MORAN, A. P. (1971). Maximum-likelihood estimation in non-standard conditions. *Proc. Cambridge Philos. Soc.* **70** 441–450.

SEARLE, S. R., CASELLA, G. and MCCULLOCH, C. E. (1992). *Variance Components*. Wiley, New York.

SELF, S. G. and LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82** 605–610.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WESTERN AUSTRALIA
NEDLANDS 6907
WESTERN AUSTRALIA
E-MAIL: hien@maths.uwa.edu.au