



Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions

Steven G. Self; Kung-Yee Liang

Journal of the American Statistical Association, Vol. 82, No. 398. (Jun., 1987), pp. 605-610.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28198706%2982%3A398%3C605%3AAPOMLE%3E2.0.CO%3B2-2>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions

STEVEN G. SELF and KUNG-YEE LIANG*

Large sample properties of the likelihood function when the true parameter value may be on the boundary of the parameter space are described. Specifically, the asymptotic distribution of maximum likelihood estimators and likelihood ratio statistics are derived. These results generalize the work of Moran (1971), Chant (1974), and Chernoff (1954). Some of Chant's results are shown to be incorrect.

The approach used in deriving these results follows from comments made by Moran and Chant. The problem is shown to be asymptotically equivalent to the problem of estimating the restricted mean of a multivariate Gaussian distribution from a sample of size 1. In this representation the Gaussian random variable corresponds to the limit of the normalized score statistic and the estimate of the mean corresponds to the limit of the normalized maximum likelihood estimator. Thus the limiting distribution of the maximum likelihood estimator is the same as the distribution of the projection of the Gaussian random variable onto the region of admissible values for the mean.

A variety of examples is provided for which the limiting distributions of likelihood ratio statistics are mixtures of chi-squared distributions. One example is provided with a nuisance parameter on the boundary for which the asymptotic distribution is not a mixture of chi-squared distributions.

KEY WORD: Boundary.

1. INTRODUCTION

The purpose of this article is to derive large sample properties of the likelihood surface under conditions similar to Cramér's (1946) but allowing the true parameter value to be on the boundary of the parameter space. The results are stated in terms of properties of maximum likelihood estimators in the loose sense. By analogy to Kullback (1957), we define any point in the parameter space at which a local maximum of the likelihood function occurs to be a maximum likelihood estimator in the loose sense. The results presented include the existence of a consistent maximum likelihood estimator, the large sample distribution of that estimator, and the large sample distribution of likelihood ratio statistics.

Previous work in this area includes that done by Moran (1971) and Chant (1974), who investigated the large sample distribution of maximum likelihood estimators. Moran considered the special case of a rectangular parameter space with at most two coordinates of the parameter on the boundary. Chant described the relationship between the $C(\alpha)$ test and tests based on maximum likelihood estimators. In discussing their results, both Moran and Chant mentioned the asymptotic equivalence of this problem with the problem of estimating the restricted mean of a multivariate Gaussian distribution. The approach taken

here may be considered a rigorous development of these comments.

Let $f(x; \theta)$ be the probability density function of a random variable X indexed by $\theta = (\theta_1, \dots, \theta_p)$, where θ takes values in the parameter space Ω , a subset of R^p . We assume that distinct values of θ correspond to distinct probability distributions. Let X_1, \dots, X_N be N independent observations on X and denote the log-likelihood function, $\sum \log f(X_i; \theta)$, by $l_N(\theta)$. We assume the almost sure existence of the first three derivatives of $l_N(\theta)$ with respect to θ on the intersection of neighborhoods of the true parameter value and Ω . If the true parameter value, denoted by θ_0 , is on the boundary of Ω , the derivatives of $l_N(\theta)$ are taken from the appropriate side. Moreover, on the intersection of neighborhoods of θ_0 and Ω , N^{-1} times the absolute value of the third derivative of $l_N(\theta)$ is bounded by a function of X_1, \dots, X_N whose expectation exists. We denote the first two derivatives of $l_N(\theta)$ by $U_N(\theta)$ and $-I_N(\theta)$, respectively. Finally, the expectation of $N^{-1}I_N(\theta)$, denoted by $I(\theta)$, is assumed to be positive definite on neighborhoods of θ_0 and at θ_0 is equal to the variance-covariance matrix of $N^{-1/2}U_N(\theta_0)$.

2. CONSISTENCY AND WEAK CONVERGENCE OF THE MAXIMUM LIKELIHOOD ESTIMATOR

Following Chernoff (1954), we offer the following definition:

Definition. The set $\Omega \subset R^p$ is approximated at θ_0 by a cone with vertex at θ_0 , C_Ω , if

$$(1) \inf_{x \in C_\Omega} \|x - y\| = o(\|y - \theta_0\|) \text{ for all } y \in \Omega$$

and

$$(2) \inf_{y \in \Omega} \|x - y\| = o(\|x - \theta_0\|) \text{ for all } x \in C_\Omega.$$

Recall that a cone with vertex at θ_0 , C , is a set of points such that if $x \in C$ then $a(x - \theta_0) + \theta_0 \in C$, where a is any real, nonnegative number. Let $C - \theta_0$ denote the set obtained by translating a cone, C , with vertex at θ_0 so that its vertex lies at the origin. From now on we will assume that Ω is regular enough to be approximated by a cone with vertex at θ_0 . This condition is mild enough to encompass a wide variety of shapes for Ω . For instance, it is easy to show that a sphere may be approximated at a point by the plane tangent to the sphere at that point.

We first give a consistency result for the maximum like-

* Steven G. Self is Associate Member, Fred Hutchinson Cancer Research Center, Seattle, WA 98104. Kung-Yee Liang is Associate Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205. This work was supported in part by National Institutes of Health Grant GM-32502.

likelihood estimator when θ_0 is on the boundary of Ω . For this result we require that near θ_0 , Ω behave like a closed set. Specifically, we will assume that the intersection of Ω and the closure of neighborhoods centered about θ_0 constitute closed subsets of R^p .

Theorem 1. If the foregoing conditions hold, then with probability tending to 1 as $N \rightarrow \infty$ there exists a sequence of points in Ω , $\hat{\theta}_N$, at which local maxima of $l_N(\theta)$ occur, and that converges to θ_0 in probability. Moreover, $N^{1/2}(\hat{\theta}_N - \theta_0) = O_p(1)$.

Proof. Pick $\delta > 0$. Since the intersection of Ω and the closure of a δ neighborhood about θ_0 is closed, $l_N(\theta)$ must have a local maximum on this set. If it can be shown that this maximum occurs at a point in Ω at a distance from θ_0 less than δ with probability tending toward 1, then the existence and consistency of $\hat{\theta}_N$ will follow immediately. This may be shown by proving that $l_N(\theta) < l_N(\theta_0)$ with probability tending toward 1 for all θ in Ω that are at a distance δ from θ_0 . The same argument used by Lehmann (1983, pp. 429–432), which relies on a Taylor series expansion of $l_N(\theta)$ about θ_0 , may be used to establish this inequality. The root- N consistency then follows from arguments in Chernoff's (1954) lemma 1.

To show weak convergence of $\hat{\theta}_N$, we proceed in two steps. First, a quadratic approximation to $l_N(\theta)$ is made and the value of θ that maximizes this quadratic function, $\tilde{\theta}_N$, is shown to be asymptotically equivalent to $\hat{\theta}_N$. The large sample distribution of $\tilde{\theta}_N$ is then derived by arguments similar to those of Chernoff (1954).

Lemma 1. Under the conditions stated in Section 1 and for all θ such that $\theta - \theta_0 = O_p(N^{-1/2})$, two times $N^{-1}l_N(\theta) - N^{-1}l_N(\theta_0)$ is equal to

$$-(Z_N - (\theta - \theta_0))' I(\theta_0) (Z_N - (\theta - \theta_0)) + N^{-2} U_N'(\theta_0) I^{-1}(\theta_0) U_N(\theta_0) + O_p(1) \|\theta - \theta_0\|^3, \quad (2.1)$$

where $Z_N = N^{-1} I^{-1}(\theta_0) U_N(\theta_0)$. Moreover, if Ω is convex in a neighborhood of θ_0 then $N^{1/2}|\hat{\theta}_N - \tilde{\theta}_N| = o_p(1)$, where $\tilde{\theta}_N$ is the value of θ in Ω at which the first term in (2.1) is maximized.

Proof. Expression (2.1) follows from a Taylor series expansion of $l_N(\theta)$ about θ_0 and a reexpression in terms of Z_N . To show that $N^{1/2}|\hat{\theta}_N - \tilde{\theta}_N| = o_p(1)$, let $g_N(\theta)$ denote the first two terms in expression (2.1) and let $r_N(\theta)$ denote the last term. Because g_N is a positive definite quadratic function, the desired result will follow from the convexity of Ω upon showing that $|g_N(\hat{\theta}_N) - g_N(\tilde{\theta}_N)|$ is $o_p(N^{-1})$. By using (2.1) we have

$$0 \leq 2N^{-1}[l_N(\hat{\theta}_N) - l_N(\tilde{\theta}_N)] = g_N(\hat{\theta}_N) - g_N(\tilde{\theta}_N) + r_N(\hat{\theta}_N) - r_N(\tilde{\theta}_N).$$

Since $g_N(\hat{\theta}_N) - g_N(\tilde{\theta}_N)$ is negative, it follows that it must be bounded in absolute value by $r_N(\hat{\theta}_N) - r_N(\tilde{\theta}_N)$. This term is seen to be $O_p(N^{-3/2})$ by the results of Theorem 1 and by noting that the same arguments used in Theorem 1 may be applied to demonstrate the root- N consistency of $\tilde{\theta}_N$.

We now derive a representation for the asymptotic distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$.

Theorem 2. Let Z be a random variable with a multivariate Gaussian distribution with mean θ and covariance matrix $I^{-1}(\theta_0)$, where θ is restricted to lie in $C_\Omega - \theta_0$. Let F denote the distribution of the maximum likelihood estimate of θ based on a single realization of Z when $\theta = 0$, hereafter denoted by $\hat{\theta}$. Then under the conditions given previously, the limiting distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$ equals F .

Proof. Two approximations are made. First, Ω is approximated by C_Ω as the set over which $l_N(\theta)$ is maximized. This is justified by the root- N consistency of $\hat{\theta}_N$ and the definition of approximating cones given previously. The second approximation, justified in Lemma 1, replaces $l_N(\theta)$ by the quadratic form

$$(N^{1/2}Z_N - N^{1/2}(\theta - \theta_0))' I(\theta_0) (N^{1/2}Z_N - N^{1/2}(\theta - \theta_0)).$$

This expression is maximized over values of θ in C_Ω or, equivalently, values of $N^{1/2}(\theta - \theta_0)$ in $C_\Omega - \theta_0$. The result follows upon noting that the limit law of $N^{1/2}Z_N$ is multivariate Gaussian with mean 0 and covariance matrix $I^{-1}(\theta_0)$.

Although the representation given in Theorem 2 reduces the general problem of computing the limiting distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$ to a single class of considerably simpler problems, the solution to the simpler problem in any given case might still be rather arduous. We now offer several special cases in which the representation given in Theorem 2 may be used to compute the limiting distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$.

Case 1. Suppose that θ_0 is an interior point of Ω . Then $C_\Omega - \theta_0 = R^p$. The maximum likelihood estimator, $\hat{\theta}$, in the previous representation equals Z , so F is multivariate Gaussian with mean 0 and covariance matrix $I^{-1}(\theta_0)$.

Case 2. Suppose that $\Omega = \Omega_1 \times \cdots \times \Omega_p$, where Ω_i are closed intervals in R^1 and θ_{01} is a left endpoint of Ω_1 and θ_{0i} are interior points of Ω_i for $2 \leq i \leq p$. Then $C_\Omega - \theta_0 = [0, \infty) \times R^{p-1}$ and $\hat{\theta}$ has the representation

$$\begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{bmatrix} I\{Z_1 > 0\} + \begin{bmatrix} 0 \\ Z_2 - (I^{21}/I^{11})Z_1 \\ \vdots \\ Z_p - (I^{p1}/I^{11})Z_1 \end{bmatrix} I\{Z_1 < 0\}, \quad (2.2)$$

where $I^{ij} = I^{ij}(\theta_0)$ are elements of the matrix $I^{-1}(\theta_0)$. Upon noting that the elements of the vector in the second term in (2.2) and the event $\{Z_1 < 0\}$ are independent, we see that this representation corresponds to the result described in Moran's (1971) theorem 1 and Chant's (1974) case (i). The vector in the second term in (2.2) is simply the projection of Z onto the half-plane $\{Z_1 \geq 0\}$, where the projection is taken according to the metric defined by $I(\theta_0)$.

In the previous example, the fact that $\hat{\theta}$ is multivariate Gaussian conditional on $\hat{\theta}_1 = 0$ stems from the fact that the event $\{Z_1 < 0\}$ and components of $\hat{\theta}$ in the second term of (2.2) are independent. This independence between elements of $\hat{\theta}$ and events characterizing which coordinates

of $\hat{\theta}$ are zero does not obtain when there is more than one coordinate of θ_0 on the boundary. For example, let $p = 2$, $C_\Omega = [0, \infty) \times [0, \infty)$, and $\theta_{01} = \theta_{02} = 0$. Then $\hat{\theta}'$ may be written as $(0, 0)$, $(0, Z_2 - (I^{21}/I^{11})Z_1)$, $(Z_1 - (I^{12}/I^{22})Z_2, 0)$, or (Z_1, Z_2) depending in which of four regions (Z_1, Z_2) falls. The regions are given by $\{Z_1 - (I^{12}/I^{22})Z_2 < 0, Z_2 - (I^{21}/I^{11})Z_1 < 0\}$, $\{Z_2 - (I^{21}/I^{11})Z_1 > 0, Z_1 < 0\}$, $\{Z_1 - (I^{12}/I^{22})Z_2 > 0, Z_2 < 0\}$, and $\{Z_1 > 0, Z_2 > 0\}$, respectively. Simple calculation of the distribution of $Z_2 - (I^{21}/I^{11})Z_1$, say, conditional on (Z_1, Z_2) falling in the second region demonstrates the nonnormality of the components of the distribution of $\hat{\theta}$, although if $I^{21} = 0$, the components are half-normal. A counterexample to Chant's theorem 2, in which he states that all components of the mixture are multivariate Gaussian, may be constructed by taking the original estimation problem to correspond to the asymptotic representation given previously. Then for every N , the maximum likelihood estimator has the representation given previously and, therefore, in the limit does not have a distribution corresponding to a mixture of multivariate Gaussians.

The error in Chant's proof appears in the specifications of the inequalities that characterize the configuration of $\hat{\theta}$'s equal to zero. In the counterexample, for instance, Chant computed the second component of the mixture as the distribution of $Z_2 - (I^{21}/I^{11})Z_1$ conditional on $\{Z_1 < 0\}$ rather than conditional on $\{Z_1 < 0, Z_2 - (I^{21}/I^{11})Z_1 > 0\}$.

Case 3. Suppose that Ω is as in Case 2, $\theta_{01}, \dots, \theta_{0q}$ are left endpoints of $\Omega_1, \dots, \Omega_q$, and θ_{0i} are interior points of Ω_i ($q + 1 < i < p$). There are 2^q possible configurations of the first q coordinates of $\hat{\theta}$, where the configuration is taken to mean which coordinates of $\hat{\theta}$ are zero. The i th such configuration is characterized by a set of inequalities satisfied by q linear combinations of Z , denoted by $\{L_i Z > 0\}$. In addition, given the i th configuration, $\hat{\theta}$ is simply the projection of Z , $P_i Z$, onto the appropriate edge of $C_\Omega - \theta_0$. $P_i = I - I^{-1}(\theta_0)B_i[I^{-1}(\theta_0)B_i]^{-1}B_i'$, with B_i a diagonal matrix with ones as diagonal elements corresponding to coordinates of $\hat{\theta}$ that are zero, and zeros otherwise. Thus $\hat{\theta}$ may be written as $\sum P_i Z I\{L_i Z > 0\}$ and the distribution of $\hat{\theta}$, $F(\cdot)$ may be written as $\sum_i P[L_i Z_i > 0]P[P_i Z \leq (\cdot) | L_i Z_i > 0]$.

3. LARGE SAMPLE DISTRIBUTION OF LIKELIHOOD RATIO STATISTICS

In many problems, it is desired to test the hypothesis that θ_0 lies in a subset of Ω , denoted by Ω_0 , versus the alternative that θ_0 lies in the complement of Ω_0 in Ω , denoted by Ω_1 . When Ω_0 is an r -dimensional subset of Ω , θ_0 is a boundary point of both Ω_0 and Ω_1 but θ_0 is an interior point of Ω , and under suitable regularity conditions, the asymptotic distribution of the likelihood ratio test statistic, $-2 \ln \lambda_N$, is χ^2 on $p - r$ df. Chernoff (1954) provided a representation of the asymptotic distribution of $-2 \ln \lambda_N$ when Ω_0 and Ω_1 both have the same dimension as Ω , θ_0 is a boundary point of both Ω_0 and Ω_1 , and θ_0 is also interior point of Ω . The following theorem generalizes these results

to the case in which θ_0 is a boundary point of Ω . We will assume from now on that both Ω and Ω_0 are regular enough to be approximated by cones with vertices at θ_0 .

Theorem 3. Let Z be a random variable with a multivariate Gaussian distribution with mean θ and covariance matrix $I^{-1}(\theta_0)$, and let C_{Ω_0} and C_{Ω_1} be non-empty cones approximating Ω_0 and Ω_1 at θ_0 , respectively. Then under the regularity conditions given in Section 1, the asymptotic distribution of the likelihood ratio statistic, $-2 \ln \lambda_N$, is the same as the distribution of the likelihood ratio test of $\theta \in C_{\Omega_0}$ versus the alternative $\theta \in C_{\Omega_1}$ based on a single realization of Z when $\theta = \theta_0$.

The proof follows immediately from the approximations described in Section 2 and Chernoff's (1954) theorem 1.

The asymptotic representation of the likelihood ratio statistic given by Theorem 3 may be written as

$$\sup_{\theta \in C_{\Omega_0} - \theta_0} \{-(Z - \theta)' I(\theta_0)(Z - \theta)\} - \sup_{\theta \in C_{\Omega_0} - \theta_0} \{-(Z - \theta)' I(\theta_0)(Z - \theta)\}, \quad (3.1)$$

where Z has a multivariate Gaussian distribution with mean 0 and covariance matrix $I^{-1}(\theta_0)$. It is convenient to rewrite (3.1) as

$$\inf_{\tilde{\theta} \in \tilde{C}_0} \|\tilde{Z} - \tilde{\theta}\|^2 - \inf_{\tilde{\theta} \in \tilde{C}} \|\tilde{Z} - \tilde{\theta}\|^2 \quad (3.2)$$

with $\tilde{C} = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{1/2} P^T \theta \text{ for all } \theta \in C_\Omega - \theta_0\}$ and $\tilde{C}_0 = \{\tilde{\theta} : \tilde{\theta} = \Lambda^{1/2} P^T \theta \text{ for all } \theta \in C_{\Omega_0} - \theta_0\}$, where \tilde{Z} has a multivariate Gaussian distribution with mean 0 and identity covariance matrix and $P \Lambda P^T$ represents the spectral decomposition of $I(\theta_0)$. Once the problem is changed to \tilde{Z} and the new parameter sets, probabilities can be computed using the standard Gaussian distribution. Note that the probability structure is retained through orthogonal transformations of \tilde{Z} .

We now present some special cases in which the representation given by (3.2) is used to calculate the distribution of likelihood ratio statistics. We will assume from here on that $\Omega = \Omega_1 \times \dots \times \Omega_p$, where the Ω_i 's are either closed, half-open, or open intervals in R^1 . In these examples we partition the parameter vector into four components: the first q coordinates of θ will represent the parameters of interest, with true values on the boundary; the next s coordinates of θ are parameters of interest that have true values not on the boundary; the next t coordinates of θ will represent nuisance parameters with true values on the boundary; and finally the last $p - q - s - t$ coordinates of θ will represent nuisance parameters with true values not on the boundary. Thus θ may be written as $(\theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_{q+s}, \theta_{q+s+1}, \dots, \theta_{q+s+t}, \theta_{q+s+t+1}, \dots, \theta_p)$ and the parameter configuration for the special cases described subsequently can be characterized by the four-tuple $(q, s, t, p - q - s - t)$. We consider likelihood ratio tests of the hypothesis $\theta_1 = \theta_{01}$, $\theta_2 = \theta_{02}$, $\theta_{q+s} = \theta_{0q+s}$ with $\theta_{q+s+1}, \dots, \theta_p$ left unspecified.

Case 4. Suppose that the parameter configuration is

given by $(0, s, 0, p - s)$, so there are no parameters on the boundary. Then $\tilde{C} = R^p$ and \tilde{C}_0 is a $p - s$ dimensional subspace of R^p . By representing \tilde{Z} in terms of a new orthonormal basis for R^p in which the last $p - s$ basis vectors span \tilde{C}_0 , we have Equation (3.2) reducing to

$$\sum_{i=1}^s \tilde{Z}_i^2, \quad (3.3)$$

which has a chi-squared distribution on s df.

Case 5. Consider the parameter configuration $(1, 0, 0, p - 1)$. Then after an orthonormal transformation, we have $\tilde{C} = [0, \infty) \times R^{p-1}$ and $\tilde{C}_0 = \{0\} \times R^{p-1}$. Thus Equation (3.2) can be reduced to

$$\tilde{Z}_1^2 I(\tilde{Z}_1 > 0).$$

Consequently, the asymptotic distribution of $-2 \ln \lambda_N$ can easily be seen to be a 50:50 mixture of χ_0^2 and χ_1^2 . A typical example of this kind is the variance component problem (Miller 1977), in which we are interested in testing the existence of a single random effect.

Case 6. Consider the parameter configuration $(1, 1, 0, p - 2)$. Then, after an orthonormal transformation, we have $\tilde{C} = [0, \infty) \times R^{p-1}$ and $\tilde{C}_0 = \{0\} \times \{0\} \times R^{p-2}$. Equation (3.2) can be expressed as

$$\tilde{Z}_1^2 I(\tilde{Z}_1 > 0) + \tilde{Z}_2^2,$$

which has a 50:50 mixture of χ_1^2 and χ_2^2 as its distribution. By using the variance component problem as an illustration, we are testing in this example that the random effect is null and the mean is equal to μ_0 , a constant, simultaneously while leaving the other parameters, such as the error variance, totally unspecified.

In Cases 5 and 6 we take advantage of the fact that there is only one boundary parameter and it occurs as a parameter of interest. Cases 7 and 8 show that the asymptotic distribution of the likelihood ratio statistic may become considerably more complicated once the aforementioned conditions do not hold. This distribution is most conveniently expressed as a mixture, with the components of the mixture identified with different regions of R^p into which \tilde{Z} may fall.

Case 7. Consider the parameter configuration $(2, p, 0, 0)$. The shaded region in Figure 1 represents \tilde{C} , and \tilde{C}_0 is the origin. In the shaded region, the likelihood ratio test reduces to $\tilde{Z}_1^2 + \tilde{Z}_2^2$, which has a χ_2^2 distribution. In regions 1 and 3, (3.2) yields squared Gaussian variables and, therefore, has a χ_1^2 distribution. Finally, Equation (3.2) reduces to zero in region 2. Note that the angle in \tilde{C} is less than 180° . This is true because the convexity of $C_\Omega - \theta_0$ is preserved under the linear mapping into \tilde{C} . The mixing probability, p , for the shaded region can be calculated as

$$\cos^{-1} \left\{ \frac{(1 \ 0) P^T \Lambda^{1/2} \Lambda^{1/2} P \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\|\Lambda^{1/2} P \begin{pmatrix} 0 \\ 1 \end{pmatrix}\| \|\Lambda^{1/2} P \begin{pmatrix} 1 \\ 0 \end{pmatrix}\|} \right\} / 2\pi$$

$$= \cos^{-1} \{I_{12} / (\sqrt{I_{11} I_{22}})\} / 2\pi,$$

where the I_{ij} 's are the (i, j) entries of the information matrix $I(\theta_0)$. Thus the asymptotic distribution of $-2 \ln \lambda_N$ is a mixture of χ_0^2 , χ_1^2 , and χ_2^2 with mixing probabilities $\frac{1}{2} - p$, $\frac{1}{2}$, and p , respectively, where $p \leq \frac{1}{2}$ from the previous angle result.

Case 8. Consider the parameter configuration $(1, 0, 1, 0)$. There are six different regions that must be considered.

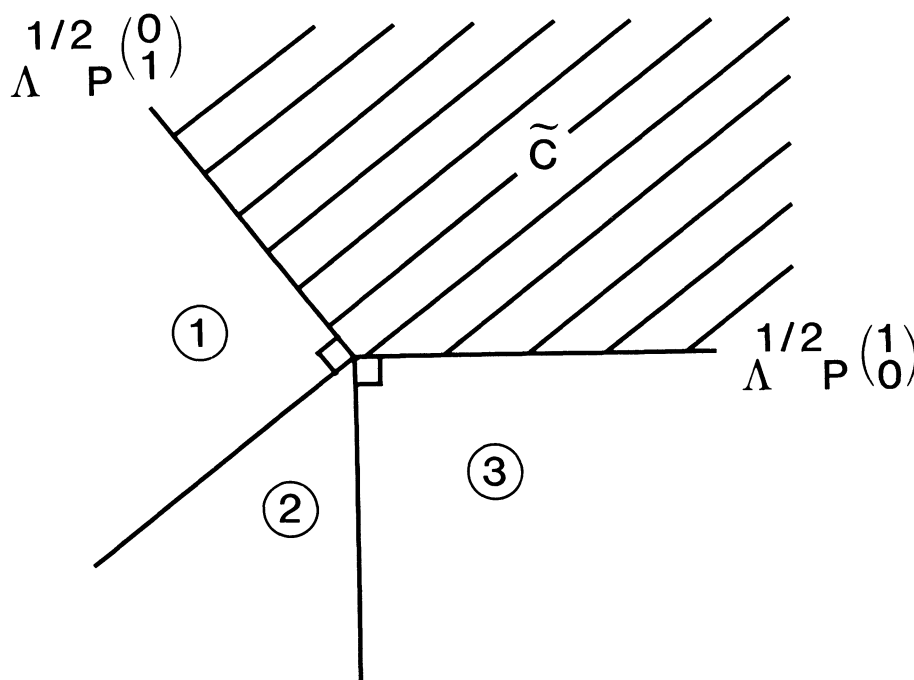


Figure 1. Diagram of the Parameter Space for Case 7. The shaded region, \tilde{C} , represents admissible parameter values under the alternative hypothesis. Under the null hypothesis, the parameter is located at the origin. The asymptotic distribution of the likelihood ratio test is a mixture of χ_0^2 , χ_1^2 , and χ_2^2 distributions with mixing probabilities depending on the angle in \tilde{C} .

The shaded region in Figure 2 represents \tilde{C} , and \tilde{C}_0 is the half-line in the same direction as the vector $\Lambda^{1/2}P^{(0)}$ with endpoint at the origin. In regions 1, 2, and 6, Equation (3.2) can be expressed as $\tilde{Z}_1^2 + \tilde{Z}_2^2$, a single squared normal random variable, and \tilde{Z}_2^2 , respectively. In regions 3, 4, and 5, (3.2) may be written as $\|Y^2\|$, $\|Y^2\| - \tilde{Z}_1^2$, respectively, where Y is the projection of \tilde{Z} onto $\Lambda^{1/2}P^{(0)}$. The distribution of these quantities conditional on \tilde{Z} being in the associated regions is not chi-squared. If α represents the angle in region 3, then by straightforward calculation the probability of (3.2) exceeding the value y^2 conditional on \tilde{Z} being in region 3 is seen to be

$$\frac{\left\{ \int_{y/\tan \alpha}^{\infty} \Phi(x \tan \alpha) d\Phi(x) - \Phi(y)[1 - \Phi(y \cot \alpha)] \right\}}{\left\{ \int_0^{\infty} \Phi(x \tan \alpha) d\Phi(x) - \frac{1}{4} \right\}},$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. The conditional distribution of (3.2) for region 5 has the same form. In a similar but more tedious calculation, the probability of (3.2) exceeding the value y^2 conditional on \tilde{Z} falling in region 4 is given by

$$\frac{\left\{ \frac{1}{2} - \int_0^{u \tan \alpha} \Phi(f(x, u, \alpha)) d\Phi(x) - \int_{u \tan \alpha}^{\infty} \Phi(x \cot \alpha) d\Phi(x) \right\}}{\left\{ \frac{1}{2} - \int_0^{\infty} \Phi(x \cot \alpha) d\Phi(x) \right\}},$$

where $f(x, u, \alpha) = [(x^2 + y^2)^{1/2} - x \sin \alpha]/\cos \alpha$ with α representing the angle in region 4. In principle, tests could be constructed from these distributions together with mix-

ing probabilities calculated from the information matrix, as in Case 8.

When the information matrix is diagonal, the orthogonal structure of $C_{\Omega} - \theta_0$ and $C_{\Omega_0} - \theta_0$ is retained upon transformation to \tilde{C} and \tilde{C}_0 , respectively. This makes explicit evaluation of Equation (3.2) much easier.

Case 9. Suppose that $I(\theta_0)$ is diagonal. Then $\tilde{C} = \tilde{\Omega}_1 \times \tilde{\Omega}_2 \times \cdots \times \tilde{\Omega}_p$, where $\tilde{\Omega}_i$ is either $[0, \infty)$ or $(-\infty, \infty)$ depending on whether the true value for the i th coordinate of the parameter vector is on the boundary or not. Similarly, $\tilde{C}_0 = \tilde{\Omega}_{01} \times \cdots \times \tilde{\Omega}_{0p}$, where $\Omega_{0i} = \{0\}$ for $i \leq q + s$ and $\tilde{\Omega}_{0i} = \tilde{\Omega}_i$ for $q + s + 1 \leq i \leq p$. Because of the orthogonal structure of these sets, the projection of \tilde{Z} onto these sets may be performed coordinatewise. Equation (3.2) may then be reduced to

$$\sum_{i=1}^q \tilde{Z}_i^2 I\{\tilde{Z}_i > 0\} + \sum_{i=q+s+1}^{q+s} \tilde{Z}_i^2. \quad (3.4)$$

Thus the asymptotic distribution of the likelihood ratio statistic is a mixture of χ^2 distributions on $s, \dots, q + s$ df, where the mixing probability for the χ_k^2 component is $\binom{q}{k-s} 2^{-q}$.

4. DISCUSSION

Recently, Shapiro (1985) investigated the asymptotic distribution of a class of test statistics (which includes likelihood ratio statistics) for the special case in which θ_0 is on the boundary of Ω_0 but is an interior point of Ω . By using virtually the same approach as in this work, Shapiro characterized this asymptotic distribution as a mixture of chi-squared distributions. When θ_0 is an interior point of

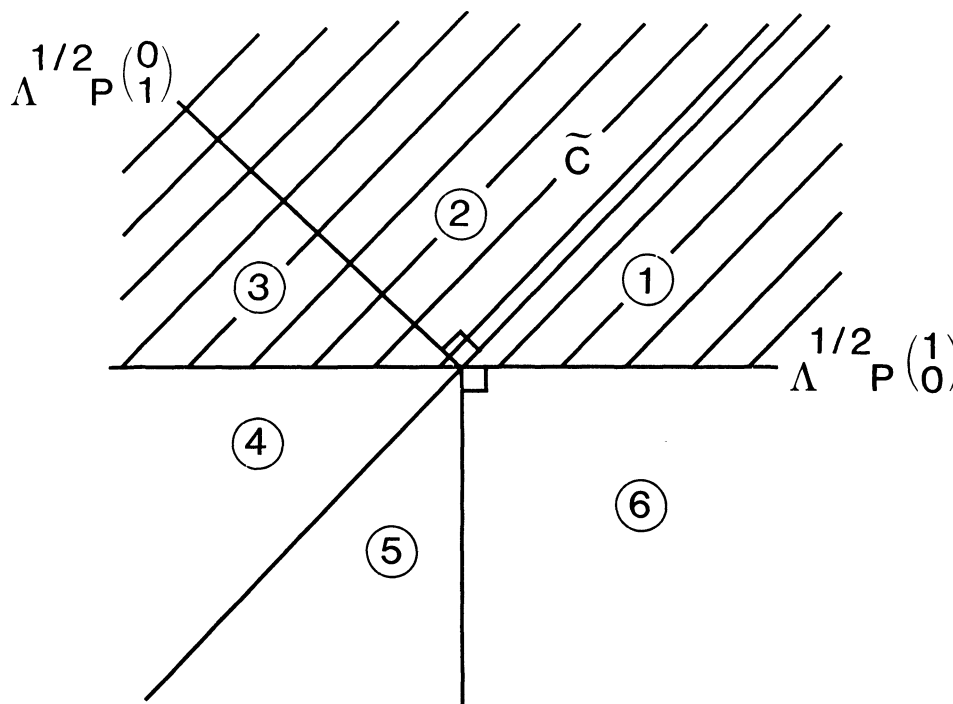


Figure 2. Diagram of the Parameter Space for Case 8. The shaded region, \tilde{C} , represents admissible parameter values under the alternative hypothesis. The half-line in the same direction as the vector $\Lambda^{1/2}P^{(0)}$ with endpoint at the origin represents admissible parameter values under the null hypothesis. The asymptotic distribution of the likelihood ratio test may be represented as a mixture of distributions. Some of the mixed distributions are χ^2 and others are not.

Ω , the second term in expression (3.2) is identically zero. This representation then corresponds to Shapiro's Lemma 2.2. It is interesting to note that our Cases 5–7 and Case 9 do not satisfy Shapiro's assumptions yet still yield a mixture of chi-squared distributions. Case 8, in which a nuisance parameter is on the boundary, demonstrates that the asymptotic distribution is not always a mixture of chi-squared distributions.

It has been suggested that it is possible to derive the previous results by a slightly different approach. If the family of probability distributions given in Section 1 can be embedded in a larger class of distributions such that θ_0 is an interior point of the parameter space within this larger class, then asymptotic properties of the maximum likelihood estimator within this larger class of models may be derived by standard methods. In fact, a referee has suggested that one can always "reflect" the parametric distributions across the boundary to create this larger problem. Projection of an asymptotic representation of this estimator onto the original parameter space may then be used to represent the maximum likelihood estimator in the original problem. This is fundamentally the same approach as the one used here in that the variable being projected onto $C_\Omega - \theta_0$ in Theorem 2, $N^{1/2}Z_N$, is asymptotically equivalent to the "unrestricted maximum likelihood estimator." The feature that characterizes the solution to the problem is the process of projection itself rather than the exact source of the Gaussian random variable that is being projected.

Although the distribution of a likelihood ratio statistic may be derived for any fixed θ_0 in Ω_0 , this distribution is generally different for different θ_0 in Ω_0 . These distributions typically vary over Ω_0 in a discontinuous way when some of the nuisance parameters may be on a boundary. For example, consider Case 5 with $p = 2$ and Case 8. The asymptotic distributions for these two cases are quite different, yet the only difference between the two cases is whether the true value of the nuisance parameter is on the boundary or not. This discontinuity can also affect the quality of the asymptotic approximation, since this de-

pends on the proximity of the nuisance parameter to the boundary relative to the precision with which the parameter can be estimated. One can imagine a situation in which the nuisance parameter is not on the boundary but is close enough so that the distribution for a given finite sample corresponds more closely to Case 8 than to Case 5.

Because the distribution may vary over Ω_0 , there may be some difficulty in determining a critical value that will give correct asymptotic size. One possibility is to compute critical values corresponding to every θ_0 in Ω_0 and then use the largest of these values. Another approach would be to use estimates of the nuisance parameters to identify the possible location of θ_0 within Ω_0 and use critical values associated with this location. One possibility for doing this is to use critical values corresponding to the analogous multivariate normal testing problem, where the true values of the nuisance parameters and covariance matrix are taken to be equal to the maximum likelihood estimates and observed information matrix, respectively. This approach might also enhance the asymptotic approximation to the finite sample problem.

[Received October 1984. Revised August 1986.]

REFERENCES

- 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," *Annals of Mathematical Statistics*, 25, 573–578.
- Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press.
- Kulldorf, G. (1957), "On the Conditions for Consistency and Asymptotic Efficiency of Maximum Likelihood Estimates," *Skandinavisk Aktuarietidskrift*, 40, 129–144.
- Lehmann, E. L. (1983), *Theory of Point Estimation*, New York: John Wiley.
- Miller, J. J. (1977), "Asymptotic Properties of Maximum Likelihood Estimates in the Mixed Model of the Analysis of Variance," *The Annals of Statistics*, 5, 746–762.
- Moran, P. A. P. (1971), "Maximum Likelihood Estimators in Non-Standard Conditions," *Proceedings of the Cambridge Philosophical Society*, 70, 441–450.
- Shapiro, A. (1985), "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints," *Biometrika*, 72, 133–144.