Statistical Inference Using Maximum Likelihood Estimation and the Generalized Likelihood Ratio When the True Parameter Is on the Boundary of the Parameter Space

by

Ziding Feng[1] and Charles E. McCulloch[2]

BU-1109-M December 1990

Summary

The classic asymptotic properties of the maximum likelihood estimator and generalized likelihood ratio statistic do not hold when the true parameter is on the boundary of the parameter space. An inferential procedure based on an enlarged parameter space is shown to have the classical asymptotic properties. Some other competing procedures are also examined.

1. Introduction

The purpose of this article is to derive an inferential procedure using maximum likelihood estimation and the generalized likelihood ratio statistic under the classic Cramer assumptions but allowing the true parameter value to be on the boundary of the parameter space. The results presented include the existence of a consistent local maxima in the neighborhood of the extended parameter space, the large-sample distribution of this maxima, the large-sample distribution of likelihood ratio statistics based on this maxima, and the construction of the confidence region constructed by what we call the Intersection Method. This method is easy to use and has asymptotically correct coverage probability. We illustrate the technique on a normal mixture problem.

---

[1] Ziding Feng was a graduate student in the Biometrics Unit and Statistics Center, Cornell University. He is now Assistant Member, Fred Hutchinson Cancer Research Center, Seattle, WA 98104.

[2] Charles E. McCulloch is an Associate Professor, Biometrics Unit and Statistics Center, Cornell University, Ithaca, NY 14853.

Keywords: restricted inference, extended parameter space, asymptotic maximum likelihood.

Previous work in this area has focused on the large-sample distribution of the maximum likelihood estimator and likelihood ratio statistics. This includes the work by Chernoff (1954), Moran (1971), Chant (1974), and Self and Liang (1987). Self and Liang's paper summarized all the earlier work and provided a uniform framework for the large-sample distribution of the above statistics. The exact limiting distributions are complicated by the number of unknown parameters, how many of them are on the boundary and the correlations between the components of $\hat{\theta}$. In some relatively simple cases the limiting distributions of the maximum likelihood estimator and likelihood ratio statistic are, respectively, mixtures of normals and mixtures of chi-squared distributions, while in the more complicated cases they are much more difficult to calculate. Our approach avoids the need for the exact limiting distributions of the maximum likelihood estimator and likelihood ratio statistics to construct confidence regions.

2. Asymptotic property of the maxima and the Intersection Method

Let $f(x,\theta)$ denote a probability density or mass function of a random variable X with parameter $\theta = (\theta_1, \theta_2, ..., \theta_p)$. $\theta \in \Omega \subset \mathbf{R}^p$ where Ω is the parameter space. Let $X = (X_1, X_2, ..., X_n)$ be a random sample of size n from $f(x,\theta)$. We assume that θ_0, the true parameter value, is on the boundary of Ω and denote the log-likelihood function $\sum_{i=1}^{n} \log f(\theta;x_i)$ by $\ell(\theta;\mathbf{x})$. $E_\theta(.)$ denotes the expectation evaulated at θ.

We first extend the definition of $\ell(\theta; \mathbf{x})$ to $\mathbf{R}^p$:

$$\ell^*(\theta; \mathbf{x}) \equiv \sum_{i=1}^{n} \log[f^*(\theta; x_i)1(f^*(\theta; x_i) > 0)] \tag{2.1}$$

where 1(.) is an indicator function and $f^*(\theta,x_i)$ is the extension of $f(\theta,x_i)$ to all $\theta \in \mathbf{R}^p$.

Lemma 2.1.

$\hat{\theta}$, the value which maximizes $\ell^*(\theta,\mathbf{x})$ in $\mathbf{R}^p$, is also the maximum for $\ell(\theta,\mathbf{x})$ in $\mathbf{R}^p$.

Remark: This lemma says that the definition of $\ell^*(\theta,\mathbf{x})$ is for mathematical convenience. The essence is to maximize $\ell(\theta,\mathbf{x})$ over $\mathbf{R}^p$ instead of over Ω but we need to explicitly define a rule to avoid the points where the density function is zero or negative.

One obvious but very useful fact is that:

$$\ell^*(\theta,\mathbf{x}) = \ell(\theta,\mathbf{x}) \quad \forall\, \theta \in \Omega\,. \tag{2.2}$$

This means that all the properties of $\ell^*(\theta,\mathbf{x})$ we will assume below are the same as the classic assumptions for $\theta \in \Omega$. Assumptions (modified from Lehmann, 1983, p.429) are:

(a) the parameter space Ω has finite dimension p, but we allow θ to be on the boundary of Ω.

(b) $f^*(x,\theta) = f^*(x,\theta')$ if and only if $\theta = \theta'$ for $\theta, \theta' \in \Omega$.

(c) there exists an open subset ω of $\mathbb{R}^p$ containing θ_0 such that for almost all x for which $f^*(\theta;\mathbf{x}) > 0$, $f^*(\theta;\mathbf{x})$ has all three derivatives w.r.t. (with respect to) all $\theta \in \omega$, and

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_\ell}\log f^*(\mathbf{x};\theta)\right| \le M_{jk\ell}(x)$$

for all $\theta \equiv \{\theta_1, \ldots, \theta_p\} \in \omega$ and $f^*(\mathbf{x},\theta) > 0$ and $m_{jk\ell} = E_{\theta_0}[M_{jk\ell}(X)] < \infty$ for all j, k, ℓ.

(d)
$$E_{\theta_0}\left[\frac{\partial}{\partial\theta_j}\log f(\mathbf{x},\theta)\right] = 0 \qquad \text{for } j = 1,\ldots, p$$

$$I_{jk}(\theta_0) \equiv E_{\theta_0}\left[\frac{\partial}{\partial\theta_j}\log f(\mathbf{x},\theta)\,\frac{\partial}{\partial\theta_k}\log f(\mathbf{x},\theta)\right] = E_{\theta_0}\left[-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\log f(\mathbf{x},\theta)\right];$$

also, $I_{jk}(\theta_0) < \infty$ and $I(\theta_0)$ is positive definite.

Theorem 2.2: Let $X_1,\ldots,X_n$ be independently identically distributed observations with density $f(x,\theta_0)$ satisfying assumptions (a)–(d) above but where the unknown θ_0 is on the boundary of Ω. Then with probability tending to 1 as $n\to\infty$, there exists a $\hat\theta \in \mathbb{R}^p$, a local maxima of the $\ell^*(\theta,\mathbf{x})$ as defined in (2.1), which has the property that

$$\text{(i) } \hat\theta \to \theta_0 \text{ w.p. 1} \qquad \text{and} \qquad \text{(ii) } n^{\frac{1}{2}}(\hat\theta-\theta_0) \xrightarrow{d} N\left[0,\, I(\theta_0)^{-1}\right].$$

To prove (i), we only need to show that for sufficiently small $\epsilon > 0$, $\ell^*(\theta,\mathbf{x}) < \ell^*(\theta_0,\mathbf{x})$ at all points θ on the surface of the ball $B_\epsilon(\theta_0)$, which defines a neighborhood centered at θ_0 with radius ϵ. Given $\ell^*(\theta,\mathbf{x})$, we can choose ϵ small enough such that $f^*(x_i,\theta) > 0$ for $\forall\, x_i$'s in the sample. This is true since (c) implies continuity of $f^*(\theta,\mathbf{x})$ with respect to θ and $f(x_i,\theta_0) > 0\ \forall\, x_i$'s. Then, by assumption (c), Taylor expansion of $\ell^*(\theta,\mathbf{x})$ about θ_0 leads to:

$$\tfrac{1}{n}\ell^*(\theta,\mathbf{x}) - \tfrac{1}{n}\ell^*(\theta_0,\mathbf{x}) = \tfrac{1}{n}\sum_{j=1}^{p}(\theta_j - \theta_{0j})\left[\frac{\partial}{\partial\theta_j}\ell^*(\theta,\mathbf{x})\Big|_{\theta=\theta_0}\right] +$$

$$+ \frac{1}{2n} \sum_{j=1}^{p} \sum_{k=1}^{p} (\theta_j - \theta_{0j})(\theta_k - \theta_{0k}) \left[\frac{\partial^2}{\partial\theta_j \partial\theta_k} \ell^*(\theta,\mathbf{x}) \Big|_{\theta=\theta_0} \right] \tag{2.3}$$

$$+ \frac{1}{6n} \sum_{j=1}^{p} \sum_{k=1}^{p} \sum_{\ell=1}^{p} (\theta_j - \theta_{0j})(\theta_k - \theta_{0k})(\theta_\ell - \theta_{0\ell}) \sum_{i=1}^{n} \gamma_{jk\ell}(x_i) M_{jk\ell}(x_i)$$

$$= S_1 + S_2 + S_3 \quad \text{and} \quad 0 \le \left| \gamma_{jk\ell}(x) \right| \le 1 \text{ by assumption (c)} .$$

By (2.2), the asterisk can be dropped in each term of the Taylor expansion. The proof then follows the classic one (Lehmann, 1983).

To prove (ii), first observe that $\ell^*(\hat{\theta},\mathbf{x}) > -\infty$; therefore, $f(x_i,\hat{\theta}) > 0$ for $\forall\, x_i$'s. This is also true for $\ell^*(\theta_0,\mathbf{x})$. By consistency of $\hat{\theta}$ and continuity of the likelihood, we can assume that $f^*(x_i,\hat{\theta}) > 0$ for $\forall\, x_i$'s and for $\forall\, \theta$ in the neighborhood of θ_0 containing $\hat{\theta}$. Then, by assumption (c), we can expand $\frac{\partial \ell^*(\hat{\theta},\mathbf{x})}{\partial\theta_j} \equiv \ell_j^{*\prime}(\hat{\theta},\mathbf{x})$ about θ_0 and we get

$$n^{\frac{1}{2}} \sum_{k=1}^{p} (\hat{\theta}_k - \theta_{0k}) \left[\tfrac{1}{n} \ell_{jk}^{*\prime\prime}(\theta_0,\mathbf{x}) + \frac{1}{2n} \sum_{\ell=1}^{p} (\hat{\theta}_1 - \theta_{01}) \ell_{jk\ell 1}^{*\prime\prime\prime}(\theta^*,X) \right]$$

$$= -n^{-\frac{1}{2}} \ell_j^{*\prime}(\theta_0,\mathbf{x}) \qquad j = 1,\ldots,p$$

where θ^* lies on the line segment connecting $\hat{\theta}$ and θ_0.

For each j, the R.H.S (right-hand side) converges in distribution to $N[0, I_{jj}(\theta_0)]$ by the Central Limit Theorem and $\frac{1}{n} \ell_{jk}^{*\prime\prime}(\theta_0,\mathbf{x}) \to I_{jk}(\theta_0)$ with probability one by the Law of Large Numbers. We claim that $\left| \ell_{jk\ell}^{*\prime\prime\prime}(\theta^*,\mathbf{x}) \right| < \infty$, so that the second term on the L.H.S (left-hand side) converges to 0 with probability one. Noticing that $\ell^*(\hat{\theta},\mathbf{x}) = \sup_{\theta} \ell^*(\theta,\mathbf{x}) > -\infty$ and $\ell^*(\theta_0,\mathbf{x}) > -\infty$, therefore $f^*(\theta^*;x_i) > 0$ by the fact that $\hat{\theta} \to \theta_0$ and θ^* is between them. The claim follows then from assumption (c). The asymptotic normality has been proved componentwise and (ii) follows from applying Lemma 4.1 of Lehmann (1983, p. 432). □

Theorem 2.2 says that the value which maximizes $\ell^*(\theta,\mathbf{x})$, is $0_p(n^{-1/2})$ consistent for θ_0 and more importantly has a limiting normal distribution.

Theorem 2.3. Let $X_1,\ldots,X_n$ be iid observations with a density function $f(x,\theta_0)$ satisfying assumptions (a)-(d) above but where the unknown θ_0 is on the boundary of Ω. Let $0 \le s < p$, where s is an

integer and $\theta = (\theta_1, \theta_2)$ where θ_1 has dimension p−s and θ_2 has dimension s. Consider testing H_0: $\theta_1 = \theta_{01}$ versus H_1: θ_1 is not specified. Then with probability tending to 1 as $n \to \infty$, there exists a $\hat{\theta} \in \mathbf{R}^p$, a local maxima of $\ell^*(\theta,\mathbf{x})$ as defined in (2.1), which has the property that:

$$-2\log\lambda^* \xrightarrow{d} \chi^2_{p-s}\,, \tag{2.4}$$

i.e., a chi-squared distribution with degree of freedom equal to p−s, where $\lambda^* = \ell^*[(\theta_{01}, \hat{\theta}_{02}); X] - \ell^*(\hat{\theta}; \mathbf{x})$, and $\hat{\theta}_{02}$ is the local maxima under H_0.

Proof. We only provide the proof for the case where s = 0, i.e., for the simple null hypothesis. The general case of composite null hypothesis is a direct extension of the simple case with the proof parallel to the classic proof (Cox and Hinkley, 1974, p.321-4).

By a Taylor expansion of $-2\log\lambda^*$ about $\hat{\theta}$, we get

$$\begin{aligned}
-2\log\lambda^* &= 2\Big\{\ell^*(\hat{\theta};\mathbf{x}) - \ell^*(\theta_0,\mathbf{x})\Big\} \\
&= 2\Big\{\ell^*(\hat{\theta},\mathbf{x}) - \ell^*(\hat{\theta},\mathbf{x}) + \sum_{j=1}^{p}(\hat{\theta}_j - \theta_{0j})\ell^{*\prime}_j(\hat{\theta},\mathbf{x}) \\
&\qquad -\tfrac{1}{2}\sum_{j=1}^{p}\sum_{k=1}^{p}(\hat{\theta}_j - \theta_{0j})(\hat{\theta}_k - \theta_{0k})\ell^{*\prime\prime}_{jk}(\theta^*,\mathbf{x})\Big\}\,,
\end{aligned}$$

where θ^* lies on the line segment between θ and $\hat{\theta}$. We have

$$\ell^{*\prime}_j(\hat{\theta},\mathbf{x}) = 0\,,$$

$$n^{\frac{1}{2}}(\hat{\theta}_j - \theta_{0j}) \xrightarrow{d} N[0, I(\theta_0)^{-1}]$$

by Theorem 2.3. and

$$\begin{aligned}
-\tfrac{1}{n}\ell^{*\prime\prime}_{jk}(\theta^*,\mathbf{x}) &= -\tfrac{1}{n}\ell^{*\prime\prime}_{jk}(\theta_0,\mathbf{x}) + \sum_{\ell=1}^{p}(\theta^*_\ell - \theta_{01})\,\tfrac{1}{n}\ell^{*\prime\prime\prime}_{jk\ell}(\theta^{**},\mathbf{x}) \\
&= I(\theta_0) + o_p(1)\,,
\end{aligned}$$

where $o_p(1)$ means that the remainder tends to 0 with probability 1 and θ^{**} lies between θ^* and θ_0.

Since $(\theta^*_1 - \theta_{01}) \to 0$ with probability 1 and $\left|\tfrac{1}{n}\ell^{*\prime\prime\prime}_{jk\ell}(\theta^{**},\mathbf{x})\right| < \infty$ and using the same arguments as in the proof of Theorem 2.2, it is obvious that the remainder goes to zero with probability 1. Therefore,

$$-2\,\ell\log\lambda^* = n(\hat{\theta}_j - \theta_{0j})^T I(\theta_0)(\hat{\theta}_j - \theta_{0j}) + 0_p(1) \xrightarrow{d} \chi^2_p\,,$$

since $n^{\frac{1}{2}}(\hat{\theta}_j - \theta_{0j}) \xrightarrow{d} N[0, I(\theta_0)^{-1}]$ by Theorem 2.2 and the quadratic form has a chi-squared distribution (Serfling, 1981, p.153). □

A simulation is presented to support the theoretical results. Figure 1 shows the simulated distribution of $-2\ell\log\lambda^*$ from 500 generated random samples of size 1000, all with an underlying distribution that is standard normal. We are testing H_0: $f(x,\theta) = N(0, 1)$ versus H_1: $f(x, \theta) = (1 - \pi)N(1, 1) + \pi N(0, 1)$. It is clear from Figure 1 that χ^2_1 fits the simulated distribution of $-2\ell\log\lambda^*$ extremely well when $n = 1000$. We also simulated samples of size $n = 100$ but the agreement was not as good, suggesting that the convergence is slow.

From Theorems 2.2 and 2.3, it is clear that an asymptotic $1-\alpha$ confidence interval or region, $\mathfrak{R}_\alpha$ about θ, can be easily constructed from $\hat{\theta}$, the maxima for $\ell^*(\theta;\mathbf{x})$, and $I(\hat{\theta})^{-1}$. We state the following procedure, which will be called the *Intersection Method*:

Step 1: Maximize $\ell^*(\theta,\mathbf{x})$ without restricting it to Ω. Call the maximizing value $\hat{\theta}$, and calculate $I(\hat{\theta})$, the Fisher information at $\hat{\theta}$.

Step 2: Calculate a confidence region based on the asymptotic chi-square distribution of Theorem 2.3:

$$\mathfrak{R}^*_\alpha \equiv \left\{\theta: -2\ell\log\lambda^* \leq \chi^2_{p,1-\alpha}\right\}. \tag{2.5}$$

Step 3:

Intersect the confidence region from Step 2 with the parameter space

$$\mathfrak{R}_\alpha \equiv \mathfrak{R}^*_\alpha \bigcap \Omega\,. \tag{2.6}$$

Theorem 2.4. Under assumptions (a)-(d) the confidence region $\mathfrak{R}_\alpha$, constructed by the above procedure, has asymptotic $1-\alpha$ coverage probability.

The proof is a simple consequence of Theorem 2.2 and 2.3. The asymptotic distribution property in Theorem 2.2 and 2.3 says that $-2\ell\log\lambda^*$ is an asymptotic pivotal quantity, i.e., its distribution does not depend on θ_0. This provides a device to construct a confidence region and hypothesis test about θ_0.

A simulation to evaluate the coverage probability for the Intersection Method is presented in Table 2.1. 500 random samples of size 1000 are drawn from a N(0,1) and .7N(0,1) + .3N(1,1), respectively. The assumed model is $\theta_1 N(0,1) + \theta_2 N(-1,1) + (1 - \theta_1 - \theta_2)N(1,1)$, where $\theta_0 = (\theta_{01}, \theta_{02}) = (1,0)$ and $(.7,0)$, respectively. The coverage probabilities are excellent compared to nominal

levels.

3. Comparison of the Intersection Method with other competitors

One competitor of the Intersection Method is to use $\chi^2_{p,1-\alpha}$ to construct a confidence set centered at the maximum likelihood estimator, assuming a correct variance structure is used. This will be liberal since the asymptotic distribution of the generalized likelihood ratio based on the maximum likelihood estimator is always stochastically smaller than χ^2_p when the true parameter is on the boundary of the p-dimensional parameter space. This fact is implicitly stated in Chernoff (1954) and Self and Liang (1987), since the asymptotic distribution of $-2\,\ell\text{og}\lambda$ can be thought of as projecting a χ^2_p random variable onto Ω. Therefore, the confidence set constructed by the above approach would generally have coverage probability larger than $1-\alpha$. This set is always greater in volume than the confidence set obtained from the Intersection Method, since all points in the intersection confidence set have larger likelihood than all the other points in Ω not included in the intersection confidence set. Therefore, the intersection method is better in the sense of exact asymptotic coverage probability and containing the points which have the higher likelihood.

Another competitor is what we called the *Truncation Method*, which is used by some practitioners and some computing software. A number of popular nonlinear estimation routines handle the boundary restriction by computing an unrestricted step and truncating the results on the appropriate boundary whenever a boundary violation occurs, especially when the Newton-Raphson algorithm is used for searching the maximum likelihood estimator (Jennrich and Sampson, 1968). Searle (1988) argued that this method can provide the maximum likelihood estimator in some variance component estimation settings. We will show that the Truncation Method works when the dimension of θ is one but does not provide reliable inferences in general.

For the 1-dimensional case, we can assume without loss of generality that $\Omega = [0, +\infty)$ and $\theta_0 = 0$ is the unknown true parameter. The Truncation Method is defined as follows: Define $\hat{\theta}_t \equiv \hat{\theta}1(\hat{\theta} \geq 0)$ where $\hat{\theta}$ is the local maxima of $\ell^*(\theta;\mathbf{x})$. Given $\hat{\theta}_t$ and $I(\hat{\theta}_t)$, the estimated Fisher information, a 'natural' confidence interval would be constructed as

$$\text{C.I.} \equiv \hat{\theta}_t \pm Z_\alpha \hat{\sigma}_t \tag{2.7}$$

where

$$\hat{\sigma}_t = \left\{ -\frac{\partial^2 \ell^*(\theta;\, X)}{\partial \theta^2}\bigg|_{\theta=\hat{\theta}_t} \right\}^{-\frac{1}{2}} .$$

Theorem 3.1. Under assumptions (a)-(d), (2.7) has asymptotic coverage probability $1-\frac{\alpha}{2}$ when $\theta_0 = 0$ and asymptotic coverage probability $1-\alpha$ when $\theta_0 > 0$.

Theorem 3.1 indicates that when a one-dimensional parameter is on the boundary of the parameter space, the simple Truncation Method leads to a conservative confidence interval due to the fact that the truncation is always made in the correct direction (towards the true parameter). We will demonstrate that this is no longer true when the dimension is greater than one.

The Truncation Method is defined for the 2-dimensional case in a similar way: Let $\theta \in [0, +\infty) \times [0, +\infty)$. Define $\hat{\theta}_t \equiv \left[\hat{\theta}_1 1(\hat{\theta}_1 > 0),\ \hat{\theta}_2 1(\hat{\theta}_2 > 0)\right]$. $I(\theta_0)$ is estimated by $I(\hat{\theta}_t)$ and a nominal $1-\alpha$ confidence ellipsoid is constructed by

$$\mathcal{R}_\alpha \equiv \left\{\theta:\, Q(\hat{\theta}_t,\theta) \le \chi^2_{2,1-\alpha}\right\} \tag{2.8}$$

where

$$Q(\hat{\theta}_t,\theta) \equiv n(\hat{\theta}_t-\theta)^T I(\hat{\theta}_t)(\hat{\theta}_t-\theta) .$$

To examine the asymptotic coverage probability of (2.8), we consider two possible cases where one or two components of θ_0 are on the boundary of Ω. Without loss of generality, we may consider the following two cases:

$$\text{Case 1:}\quad \theta_{01} = 0,\ \theta_{02} > 0 ,$$

$$\text{Case 2:}\quad \theta_{01} = \theta_{02} = 0 .$$

Theorem 3.2. Under assumptions (a)-(d), the nominal $1-\alpha$ confidence ellipsoid constructed in (2.8) using the truncated estimator has asymptotic coverage probability:

$$\text{Case 1: If } \theta_{01} = 0 \text{ and } \theta_{02} > 0 ,$$

then

$$\lim_{n\to\infty} P\{\theta \in \mathcal{R}_\alpha\} = \frac{1-\alpha}{2} + \frac{1}{2} P\left\{\chi^2_1 \le (1-\rho^2)\chi^2_{2,1-\alpha}\right\} \tag{2.9}$$

$$\text{Case 2: If } \theta_{01} = \theta_{02} = 0 ,$$

then

$$\lim_{n\to\infty} P\{\theta \in \mathcal{R}_\alpha\} = \frac{\cos^{-1}(-\rho)}{2\pi}(2-\alpha) + \left(1-\frac{\cos^{-1}(-\rho)}{\pi}\right) P\left\{\chi^2_1 \le (1-\rho^2)\chi^2_{2,1-\alpha}\right\}, \tag{2.10}$$

where ρ = asymptotic correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$.

Case 3: If $\theta_{01} > 0$ and $\theta_{02} > 0$, then the asymptotic coverage probability of the truncation method is $1-\alpha$.

Proof: The asymptotic distribution of the likelihood ratio when the true parameter is on the boundary of Ω can be thought of as the distribution of the projection of a random variable with a chi-squared distribution onto Ω. Therefore, for any simple configuration of θ_0, it is possible to decompose the distribution into different regions corresponding to the components of the parameter (Chernoff, 1954 and Self and Liang, 1987). The aymptotic coverage probabilities can then be calculated from the decomposed distribution which is usually a mixture of chi-squared distributions. Some examples of the calculations can be found in the above two papers.

Figure gives a plot of (2.9) and (2.10) versus ρ using $1-\alpha = .9$. For Case 1 (2.9), the asymptotic coverage probability is close to or a little higher than 0.9 when the absolute value of correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$ is less than 0.8. However, it performs poorly when this correlation is high (> 0.8). For Case 2 (2.10), it shows that from moderate negative correlation (> -0.7) to high correlation between $\hat{\theta}_1$ and $\hat{\theta}_2$, the asymptotic coverage probability is higher than 0.9. It decreases quite rapidly when the correlation is less than -0.8.

The extension of the calculations to dimensions higher than two is possible but complicated and tedious. Analogues of the two-dimensional case will occur and the asymptotic $1-\alpha$ coverage probability will not hold.

4. Discussion

The Intersection Method is easy to implement while inferential procedures (testing or confidence regions) based on the asymptotic distribution of the maximum likelihood estimator are difficult to use since the limiting distributions are complex when incorporating boundary constraints. On the other hand, the simple Truncation Method or other variants, which are sometimes used by practitioners, are not reliable when the dimension of the parameter is larger than one and correlations between the estimators are high.

In finite samples, the Intersection Method might lead to a confidence set which does not contain the maximum likelihood estimator or even to an empty set when the confidence set does not intersect Ω. This can be avoided if the search for maxima out of the parameter space is restricted in the neighborhood of the boundary of the parameter space. The definition of the neighborhood can be defined in a way that the confidence region by the Intersection Method always includes the maximum likelihood estimator. This will always hold asymptotically by our assumptions on the likelihood and the property of the maximum likelihood estimator derived by Chernoff (1954) and Self and Liang (1987).

For point estimation, the maximum likelihood estimator should be used instead of the unrestricted maxima since it may not have a meaningful interpretation and most probably has a larger mean square error.

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. *Ann. Math. Statist.* 25: 573-578.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall. 511 pp.

Cramer, H. (1946). *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press.

Edlefsen, L. E. and Jones, S. D. (1988). *GAUSS Version 2.0.* Kent, WA: Aptech System, Inc.

Jennrich, R. I. and Sampson, P. F. (1968). Application of stepwise regression to non-linear estimation. *Technometrics* 10: 63-72.

Lehmann, E. L. (1983). *Theory of Point Estimation.* New York: John Wiley and Sons. 506 pp.

Moran, P. A. P. (1971). Maximum Likelihood Estimators in Non-Standard Conditions. *Proceedings of the Cambridge Philosophical Society* 70: 441-450.

Searle, S. R. (1987). *Variance Components: A Set of Notes.* (Unpublished lecture notes).

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.

Table 1. Coverage probabilities of confidence regions constructed by the Intersection Method for 500 simulations. Model is a mixture of normals with three components where one or two of the parameters are on the boundary. (Sample sizes are 1000 for each simulation.)

	Nominal Coverage Probability				
θ_0	.8	.85	.90	.95	.99
(1 ,0)	.784	.846	.890	.938	.984
(.7,0)	.786	.860	.906	.940	.986

Figure 1. Simulated cumulative distribution function of the likelihood ratio statistic and χ^2_1 and χ^2_2 distributions in a test of H_0: N(0,1) vs. H_1: π N(0,1) + (1−π) N(1,1) when H_0 is true. The likelihood ratio statistic is based on the unrestricted maxima. (Sample size 1000, 500 replications)

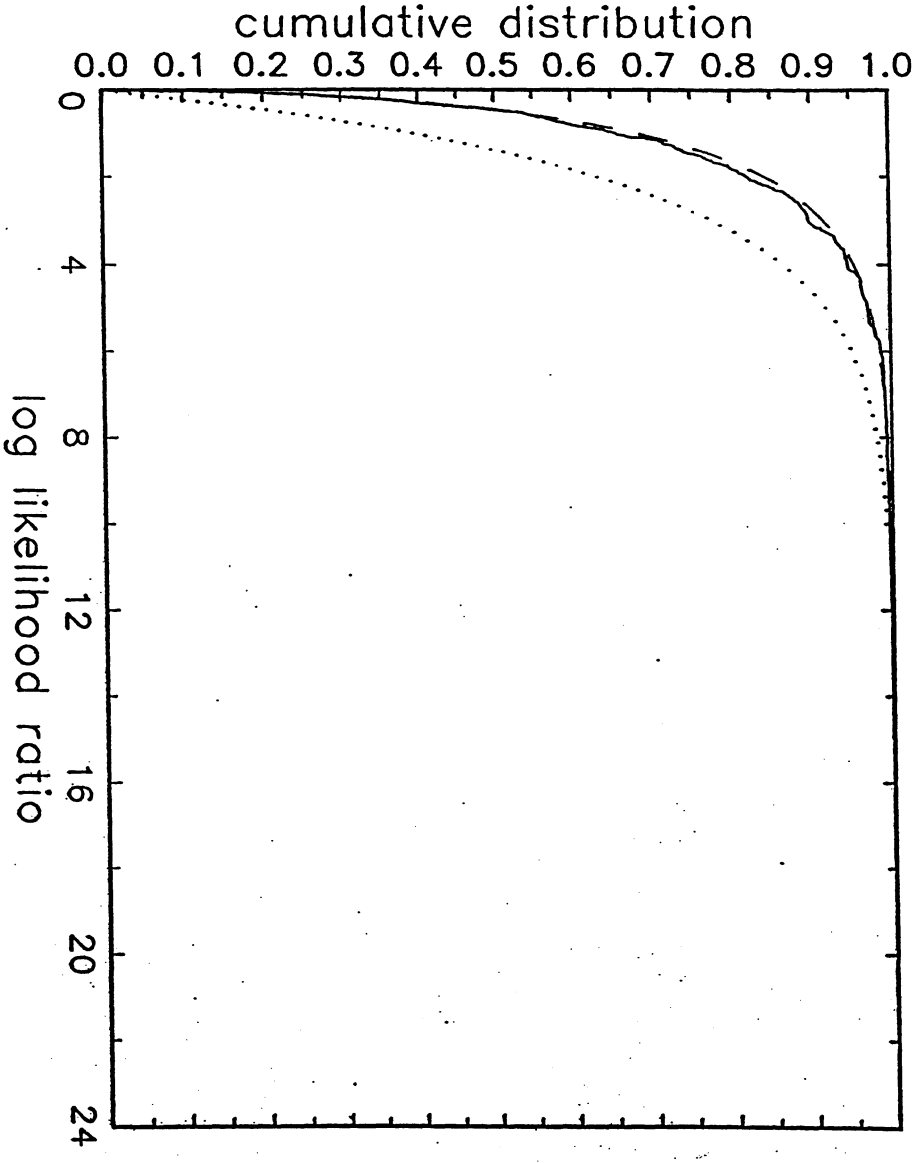

Figure 2. Asymptotic coverage probabilities of confidence regions based on the Truncation Method plotted versus ρ, the correlation between $\hat{\theta}_1$, and $\hat{\theta}_2$. Nominal coverage $1-\alpha = .9$.

Figure 1: Asymptotic coverage probabilities using the Truncation Method