ON CONSISTENCY AND ASYMPTOTIC EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATES

BY S. A. D. C. DOSS

Department of Mathematics and Statistics, University of Poona

1. INTRODUCTION

IN an earlier paper the author (Doss, 1962) extended Cramér's (1946) and Huzurbazar's (1948) results on consistency, uniqueness, asymptotic normality and asymptotic efficiency of the maximum likelihood estimates (m.l.e.'s) to multiparametric distributions under a set of conditions other than that of Chanda (1954). The set of conditions of that paper is weaker than that of Chanda (1954) as far as consistency is concerned. In the present paper the properties of m.l.e.'s will be established under another different set of conditions analogous to those of Kulldorff's (1957) Theorem 2. This set of conditions is necessarily weaker than that of Chanda (1954). It may be pointed that Kulldorff's (1957) conditions are themselves weaker than those of Cramér (1946); and Chanda's (1954) conditions are only straightforward generalization of Cramér's conditions to multiparametric distributions. The set of conditions presented in the present paper is not necessarily weaker than that given in the author's (Doss, 1962) earlier paper. However, they are only complementary to each other and together they extend the class of m.l.e.'s beyond the range covered by any one set of conditions alone.

Let $f(x;\ \theta_1, \theta_2, \ldots, \theta_k)$ be the probability density function of a distribution depending on k parameters. For brevity we shall write $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$. If $x_1, x_2, \ldots, x_n$ is a sample of n independent observations from $f(x;\ \theta)$ we shall call

$$L(\theta) = \sum_{i=1}^{n} \log f(x_i;\ \theta)$$

as the likelihood function of the sample. Let the parameter space be Ω a k-dimensional interval, and let θ° be the true unknown value of the parameter vector θ. θ° is assumed to be an inner point of Ω. We shall denote by $\delta(\theta', \theta'')$ the distance between the points θ' and θ''.

Throughout the paper we assume that the following set of conditions are satisfied. As is done in Kulldorff (1957) we also postulate the existence of at most one solution of the system of likelihood equations.

Condition 1.—For almost all x and for all $\theta \epsilon \Omega$

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_r} f(x;\theta)\, dx = \int_{-\infty}^{\infty} \frac{\partial^2}{\partial \theta_r \partial \theta_s} f(x;\theta)\, dx = 0$$

$$r, s = 1, 2, \ldots, k.$$

Condition 2.—There exists a set of functions $g_r(\theta)$ $(r = 1, 2, \ldots, k)$ which are positive and differentiable $(N-1)$ times $(N \geqslant 3)$ for all $\theta \epsilon \Omega$ and satisfy the following relations for all $\theta \epsilon \Omega$:

$$\text{(i)} \quad \left| \frac{\partial g_r(\theta)}{\partial \theta_s} \right| < \infty \qquad r, s = 1, 2, \ldots, k$$

$$\text{(ii)} \quad \left| \frac{\partial^m \left\{ g_r(\theta) \frac{\partial}{\partial \theta_r} \log f(x;\theta) \right\}}{\partial \theta_{r_1} \partial \theta_{r_2} \ldots \partial \theta_{r_m}} \right| < \infty$$

$$m = 1, 2, \ldots, N-2;\ r_1, r_2, \ldots, r_m = 1, 2, \ldots, k.$$

$$\text{(iii)} \quad \left| \frac{\partial^{N-1} \left\{ g_r(\theta) \frac{\partial}{\partial \theta_r} \log f(x;\theta) \right\}}{\partial \theta_{r_1} \partial \theta_{r_2} \ldots \partial \theta_{r_{N-1}}} \right| < H_{rr_1 \ldots r_{N-1}}(x)$$

while

$$\int_{-\infty}^{\infty} H_{rr_1 \ldots r_{N-1}}(x) \cdot f(x;\theta)\, dx < M < \infty.$$

Condition 3.—For all $\theta \epsilon \Omega$, the matrix $I(\theta) = \| I_{rs}(\theta) \|$ where

$$I_{rs}(\theta) = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta_r} \log f(x;\theta) \cdot \frac{\partial}{\partial \theta_s} \log f(x;\theta) \cdot f(x;\theta)\, dx$$

is non-singular and $|I(\theta)|$ is finite.

It may be noted that, for $N = 3$, these conditions are the same as those of Kulldorff's (1957) Theorem 2.

2. Consistency of the M.L.E.

In the following we shall prove both the existence of the solution of the system of likelihood equations and its consistency.

Theorem 1.—The solution of the system of likelihood equations θ converges to the true value of the parameter vector θ as $n \to \infty$.

Proof.—The system of likelihood equations for estimating the unknown parameter vector are:

$$\frac{\partial L(\theta)}{\partial \theta_r} = 0 \qquad r = 1, 2, \ldots, k.$$

Expanding

$$\frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r}$$

in a Taylor's series, up to N terms, in the neighbourhood of the true θ^0 of the unknown parameter vector θ we have

$$\frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r}$$

$$= \frac{g_r(\theta^0)}{n} \left(\frac{\partial L(\theta)}{\partial \theta_r} \right)_{\theta=\theta^0} + \sum_{s=1}^{k} \delta_s \left[\frac{\partial}{\partial \theta_s} \left\{ \frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r} \right\} \right]_{\theta=\theta^0}$$

$$+ \frac{1}{2} \sum_{s,t=1}^{k} \delta_s \delta_t \left[\frac{\partial^2}{\partial \theta_s \partial \theta_t} \left\{ \frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r} \right\} \right]_{\theta=\theta^0} + \ldots$$

$$+ \frac{1}{(N-1)!} \sum_{r_1, \ldots, r_{N-1}=1}^{k} \delta_{r_1} \ldots \delta_{r_{N-1}}$$

$$\times \left[\frac{\partial^{N-1}}{\partial \theta_{r_1} \partial \theta_{r_2} \ldots \partial \theta_{r_{N-1}}} \left\{ \frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r} \right\} \right]_{\theta=\theta'},$$

where $\delta_s = \theta_s - \theta_s^0$ and θ' is such that $\delta(\theta, \theta^0) > \delta(\theta', \theta^0)$, $\delta(\theta', \theta)$ for every x. The above relation may be written as

$$A_r(\theta)$$

$$= A_r(\theta^0) + \sum_{s=1}^{k} \delta_s A_{rs}(\theta^0) + \frac{1}{2} \sum_{s,t=1}^{k} \delta_s \delta_t A_{rst}(\theta^0) + \ldots$$

$$+ \frac{1}{(N-1)!} \sum_{r_1, \ldots, r_{N-1}=1}^{k} \delta_{r_1} \delta_{r_2} \ldots \delta_{r_{N-1}} A_{rr_1 \ldots r_{N-1}}(\theta') \tag{1}$$

where

$$A_r(\theta) = \frac{g_r(\theta)}{n} \frac{\partial L(\theta)}{\partial \theta_r},$$

and

$$A_{rr_1 \ldots r_m}(\theta) = \frac{\partial^m \left\{ \dfrac{g_r(\theta)}{n} \dfrac{\partial L(\theta)}{\partial \theta_r} \right\}}{\partial \theta_{r_1} \partial \theta_{r_2} \ldots \partial \theta_{r_m}}$$

$$m = 1, 2, \ldots, N-1.$$

Since $A_{rr_1 \ldots r_m}(\theta^\circ)$ is the arithmetic mean of identically distributed independent random variables, in virtue of Khintchine's theorem $A_{rr_1 \ldots r_m}(\theta^0)$ converges in probability to

$$E\left[\frac{\partial^m \left\{ g_r(\theta) \dfrac{\partial}{\partial \theta_r} \log f(x; \theta) \right\}}{\partial \theta_{r_1} \partial \theta_{r_2} \ldots \partial \theta_{r_m}}\right]_{\theta=\theta^0}$$

for all $r_1, r_2, \ldots, r_m = 1, 2, \ldots, N-2$ and $| A_{rr_1 \ldots r_{N-1}}(\theta') |$ converges in probability to a finite positive quantity less than M for all r_i's because of condition 2 (iii). In particular because of conditions 1 and 3, $A_r(\theta^0)$ converges in probability to

$$E\left[g_r(\theta) \frac{\partial}{\partial \theta_r} \log f(x; \theta)\right]_{\theta=\theta^0}$$

$$= g_r(\theta^0) \, E\left[\frac{\partial}{\partial \theta_r} \log f(x; \theta)\right]_{\theta=\theta^0} = 0,$$

and $A_{rs}(\theta)$ connverges to

$$E\left[\frac{\partial}{\partial \theta_s} \left\{ g_r(\theta) \frac{\partial}{\partial \theta_r} \log f(x; \theta) \right\}\right]_{\theta=\theta^0}$$

$$= \left\{\frac{\partial g_r(\theta)}{\partial \theta_s}\right\}_{\theta=\theta^0} E\left(\frac{\partial}{\partial \theta_r} \log f(x; \theta)\right)_{\theta=\theta^0}$$

$$+ g_r(\theta^0) \, E\left(\frac{\partial^2}{\partial \theta_r \partial \theta_s} \log f(x; \theta)\right)_{\theta=\theta^0}$$

$$= - g_r(\theta^0) \, I_{rs}(\theta^0).$$

Given two positive quantities ϵ, η and $B > M$, it is then possible to find the number $n_0 = n_0(\epsilon, \eta)$ such that for all $n > n_0(\epsilon, \eta)$

$$P\left[\,|A_r(\theta^0)| < \eta;\ \left| A_{rr_1\ldots r_m}(\theta^0) - E\left\{\frac{\partial^m\left(g_r(\theta)\dfrac{\partial}{\partial\theta_r}\log f(x;\ \theta)\right)_{\theta=\theta^0}}{\partial\theta_{r_1}\ldots\partial\theta_{r_m}}\right\}\right| < \eta;\ |A_{rr_1\ldots r_{N-1}}(\theta^0)| < B\right] > 1-\epsilon \tag{2}$$

$$m = 1, 2, \ldots, N-2;\ r_1, r_2, \ldots, r_m = 1, 2, \ldots, k.$$

Now, consider the system of likelihood equations

$$-\sum_{s=1}^{k}\delta_s A_{rs}(\theta^0) = A_r(\theta^0) + \frac{1}{2}\sum_{s,\,t=1}^{k}\delta_s\delta_t A_{rst}(\theta^0) + \ldots + \frac{1}{(N-1)!}\sum_{r_1,\ldots,r_{N-1}=1}^{k}\delta_{r_1}\delta_{r_2}\ldots\delta_{r_{N-1}}A_{rr_1\ldots r_{N-1}}(\theta') \tag{3}$$

$$r = 1, 2, \ldots, k.$$

Since, $\|I_{rs}(\theta^0)\|$ is non-singular and $g_r(\theta^0) > 0$ for all $r = 1, 2, \ldots, k$,

$$|g_r(\theta^0)\, I_{rs}(\theta^0)| = \prod_{i=1}^{k} g_i(\theta^0)\,|I_{rs}(\theta^0)| \neq 0$$

and so

$$\|g_r(\theta^0)\, I_{rs}(\theta^0)\|^{-1}$$

exists. Hence

$$\|A_{rs}(\theta^0)\|^{-1} = \|A^{rs}(\theta^0)\|$$

(say) also exists. Therefore we have from (3)

$$\delta_r = a_r(\theta^0) + \sum_{s,\,t=1}^{k}\delta_s\delta_t a_{rst}(\theta^0) + \ldots + \sum_{r_1,\ldots,r_{N-1}}^{k}\delta_{r_1}\delta_{r_2}\ldots\delta_{r_{N-1}}a_{rr_1\ldots r_{N-1}}(\theta') \tag{4}$$

where

$$a_r(\theta) = - \sum_{p=1}^{k} A_p(\theta) A^{pr}(\theta)$$

and

$$a_{rr_1 \ldots r_m}(\theta) = -\frac{1}{m!} \sum_{p=1}^{k} A_{pr_1 \ldots r_m}(\theta) \cdot A^{pr}(\theta).$$

Since $\| A^{rs}(\theta^0) \|$ is a matrix of finite elements, we can choose $n_0(t, \eta)$ for arbitrary ϵ, η such that for all $n > n_0(\epsilon, \eta)$

$$P\,[\,| a_r(\theta^0) | < \eta;\ | a_{rr_1 \ldots r_m}(\theta^0) | < \infty;$$
$$| a_{rr_1 \ldots r_{N-1}}(\theta') | < \infty] > 1 - \epsilon.$$

Hence, it is easily seen that the system of likelihood equations (3) admit a solution $\hat{\delta} = (\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_k)$ which are of the same order as η. It follows therefore that

$$P\,[\,| \hat{\delta} | < \eta] > 1 - \epsilon,$$

for all $n > n^0(\epsilon, \eta)$. Hence it immediately follows that the solution of the system of likelihood equations $\hat{\theta}$ exists and is a consistent estimate of the true parameter vector θ^0.

3. Absolute Maximum of the M.L.E.

In this section we shall establish the maximum property of maximum likelihood estimates. Before that we shall prove the following important and useful lemmas.

Lemma 1.—If θ^{T} is any consistent estimate (not necessarily the solution of the system of likelihood equations) of θ^0, then

$$\left[\frac{\partial}{\partial \theta_s}\left\{\frac{g_r(\theta)}{n}\ \frac{\partial L(\theta)}{\partial \theta_r}\right\}\right]_{\theta = \theta^{\mathrm{T}}}$$

converges in probability to $-g_r(\theta^0) I_{rs}(\theta^0)$ as $n \to \infty$.

Proof.—Expanding $A_{rs}(\theta^{\mathrm{T}})$ in a Taylor's series in the neighbourhood of θ^0 we have

$$A_{rs}(\theta^{\mathrm{T}}) = A_{rs}(\theta^0) + \sum_{t=1}^{k} \delta_t^{\mathrm{T}} A_{rst}(\theta^0)$$

$$+\frac{1}{2} \sum_{t,\, u=1}^{k} \delta_t^{\tau}\delta_u^{\tau} A_{rstu}(\theta^0) + \ldots$$

$$+ \frac{1}{(N-2)!} \sum_{r_1 \ldots r_{N-2}}^{k} \delta_{r_1}^{\tau}\delta_{r_2}^{\tau} \ldots \delta^{\tau}_{r_{N-2}} A_{r_1 r_2 \ldots r_{N-2}}(\theta'') \tag{5}$$

where

$$\delta_t^{\tau} = \theta_t^{\tau} - \theta_t^0$$

and

$$\delta(\theta^{\tau}, \theta^0) > \delta(\theta'', \theta^0),\ \delta(\theta'', \theta').$$

Now, consider the convergence of each random variable involved in the right-hand side of the above expansion. We know that $A_{rs}(\theta^0)$ converges in probability to $-g_r(\theta^0) I_{rs}(\theta^0)$, and δ_t^{τ}'s to zero. It can be easily seen from (2) that $A_{rst}(\theta^0)$, $A_{rstu}(\theta^0)$, etc., converge in probability to finite quantities. Hence, it follows from Sultsky's theorem that $A_{rs}(\theta^{\tau})$ converges in probability to $-g_r(\theta^0) I_{rs}(\theta^0)$. This completes the proof of the lemma.

Lemma 2.—The matrix

$$\left\| \frac{1}{n} \left(\frac{\partial^2 L(\theta)}{\delta\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}} \right\|$$

is negative-definite with probability tending to certainty.

Proof.—It can be easily seen that

$$\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}}$$

can be written as

$$\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta_r \partial\theta_s}\right)_{\theta=\hat{\theta}}$$

$$= \frac{1}{g_r(\hat{\theta})} \left[\left\{\frac{\partial}{\partial\theta_s}\left(\frac{g_r(\theta)}{n}\ \frac{\partial L(\theta)}{\partial\theta_r}\right)\right\}_{\theta=\hat{\theta}} - \left(\frac{\partial g_r(\theta)}{\partial\theta_s}\right)_{\theta=\hat{\theta}} \cdot \left(\frac{1}{n}\ \frac{\partial L(\theta)}{\partial\theta_r}\right)_{\theta=\hat{\theta}}\right]. \tag{6}$$

Since $\hat{\theta}$ is the consistent solution of the system of likelihood equations

$$\left(\frac{1}{n}\frac{\partial L(\theta)}{\partial \theta_r}\right)_{\theta=\hat{\theta}} = 0,$$

from lemma 1 it follows that

$$\left\{\frac{\partial}{\partial \theta_s}\left(\frac{g_r(\theta)}{n}\frac{\partial L(\theta)}{\partial \theta_r}\right)\right\}_{\theta=\hat{\theta}}$$

converges in probability to $-g_r(\theta^0) I_{rs}(\theta^0)$. Because of their continuity in Ω,

$$\left(\frac{\partial g_r(\theta)}{\partial \theta_s}\right)_{\theta=\hat{\theta}}$$

and $g_r(\hat{\theta})$ converge in probability, respectively to the finite quantities

$$\left(\frac{\partial g_r(\theta)}{\partial \theta_s}\right)_{\theta=\theta^0}$$

and $g_r(\theta^0)$ (*see* condition (2)). Hence, in virtue of Slutsky's theorem, it follows from (6) that

$$\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial \theta_r \partial \theta_s}\right)_{\theta=\hat{\theta}}$$

converges in probability to $-I_{rs}(\theta^0)$.

In order to prove lemma 2, we have to consider the quadratic from

$$\sum_{r,s=1}^{k}\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial \theta_r \partial \theta_s}\right)_{\theta=\hat{\theta}} u_r u_s$$

where $u = (u_1, u_2, \ldots, u_k)$ is any finite vector. Being a rational function of random variabes such as

$$\left(\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial \theta_r \partial \theta_s}\right)_{\theta=\hat{\theta}},$$

this form, in virtue of Slutsky's theorem, converges in probability to

$$-\sum_{r,s=1}^{k} I_{rs}(\theta^0) u_r u_s$$

which is negative-definite. Hence the lemma.

Now we have

Theorem 2.—The solution of the system of likelihood equations provides the absolute maximum of the likelihood function with probability tending to certainty as $n \to \infty$.

Proof.—From lemma 2, it immediately follows that $L(\theta)$ has a maximum at $\hat{\theta}$ with probability tending to certainty as $n \to \infty$. Since we have postulated that the system of likelihood equations have at most one solution, it follows that $L(\theta)$ has absolute maximum at $\hat{\theta}$.

4. Asymptotic Normality and Asymptotic Efficiency of the M.L.E.

Theorem 3.—The solution of the system of likelihood equations is distributed as a k-variate normal distribution with means zero and variance-covariance matrix

$$\frac{1}{n} \| I_{rs}(\theta^0) \|^{-1}.$$

And, it is asymptotically efficient estimate of the unknown parameter vector.

Proof.—Consider the following equation which is obtained by putting $\theta = \hat{\theta}$ in (4):

$$\hat{\delta}_r = a_r(\theta^0) + \sum_{s,t=1}^{k} \hat{\delta}_s \hat{\delta}_t a_{rst}(\theta^0) + \ldots$$

$$+ \sum_{r_1 \ldots r_{N-1}=1}^{k} \hat{\delta}_{r_1} \hat{\delta}_{r_2} \ldots \hat{\delta}_{r_{N-1}} a_{rr_1 \ldots r_{N-1}}(\theta'). \tag{7}$$

In virtue of multi-dimensional form of Liapounoff's central limit theorem, it follows that $A_1(\theta^0), A_2(\theta^0), \ldots, A_k(\theta^0)$ are asymptotically jointly distributed as a k-variate normal distribution with means zero and variance-covariance matrix

$$J = \frac{1}{n} \| g_r(\theta^0) g_s(\theta^0) I_{rs}(\theta^0) \|.$$

Since all the summands on the right-hand side of (7) converge in probability to zero, it then follows that $\hat{\delta}_1, \hat{\delta}_2, \ldots, \hat{\delta}_k$ are asymptotically jointly distributed as a k-variate normal distribution with zero means and variance-covariance matrix given by DJD' where

$$D = \| g_r(\theta^0) I_{rs}(\theta^0) \|^{-1}.$$

As shown earlier (Doss, 1962)

$$DJD' = \frac{1}{n} \| I_{rs}(\theta^0) \|^{-1}.$$

Hence the theorem.

In conclusion, the author wishes to express his gratitude to Dr. V. S. Huzurbazar for his constant encouragement and guidance; and to his colleague Mr. P. S. Swamy for the useful discussions he had with him during the preparation of the paper. He also wishes to thank the Government of India for awarding him a Research Training Scholarship.

References

1. Chanda, K. C. .. "A note on the consistency and maxima of the roots of likelihood equations," *Biometrika*, 1954, **41,** 56.
2. Cramér, H. .. *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 500–504.
3. Doss, S. A. D. C. ... "A note on consistency and asymptotic efficiency of maximum likelihood estimates in multi-parametric problems," *Cal. Stat. Assoc. Bull.*, 1962, **11,** 85.
4. Huzurbazar, V. S. .. "The likelihood equation, consistency and the maxima of the likelihood function," *Annals of Eugenies*, 1948, **14,** 185.
5. Kulldorff, G. .. "On the conditions for consistency and asymptotic efficiency of maximum likelihood estimates," *Skandinavisk Aktuarietidskrift*, 1957, **40,** 129.