ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Mathématiques

C. RADHAKRISHNA RAO

First and second order asymptotic efficiencies of estimators

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 8, série *Mathématiques*, nº 2 (1962), p. 33-40

<http://www.numdam.org/item?id=ASCFM_1962__8_2_33_0>

FIRST AND SECOND ORDER ASYMPTOTIC EFFICIENCIES OF ESTIMATORS

C. RADHAKRISHNA RAO
Indian Statistical Institute Calcutta

1 - INTRODUCTION -

The objects of the paper are two fold : (i) to examine the definition of asymptotic efficiency of an estimator as discussed in statistical literature and reformulate it in a way more appropriate to problems of statistical inference and (ii) to develop the concept of second order efficiency by which different estimators satisfying the asymptotic efficiency (to be called first order) could be distinguished. The method of maximum likelihood (m.l.) is known to be one out of an infinity of estimation procedures leading to asymptotically efficient (first order) estimators. It is proved, under certain conditions, that m.l. has the maximum second order efficiency, which distinguishes it from the others. The original ideas relating to these wider concepts of efficiency are contained in two fundamental papers by Fisher [7, 8]. A systematic study of estimation in large samples, based on these ideas, has been undertaken by the author in three different papers [19, 20, 21].

Before discussing the concept of efficiency it may be useful to specify the object of estimation. We do not look upon an estimate computed by a suitably chosen procedure as an end in itself, as it would be if the theory of decision functions as developed by Wald [26] is strictly adhered to. The limitations of such a theory are well known (Fisher [9]). The introduction of a loss function may be inappropriate in many situations and further there will be generally, multiple uses for an estimate for which it may be difficult to assign a consolidated loss function. For instance, if an anthropometrician is estimating the mean stature in a population on the basis of a sample he may need the estimate for a variety of purposes. He may compare it with an estimate of mean stature of another population, combine it with a previous estimate from the same population, assert with some confidence that the true stature lies between two values, preserve it for comparison or combination with future estimates, and so on. It may be argued that for each purpose a different estimate may be used depending on the consequences of the decision taken. But it may be more convenient to obtain an estimate which, without further reference to original data, would serve a variety of purposes. In any case, since estimation necessarily involves condensation of observed data, a good estimator is one which provides a satisfactory substitute for the entire data (Rao [19]).

More precisely, we shall investigate the consequences of using an estimator, instead of the whole sample, in constructing tests of significance of simple hypotheses or setting up of confidence intervals of the unknow parameter. It is shown that the loss incurred by adopting such restricted procedures is not serious in large samples if estimators are chosen to satisfy the criteria of first order efficiency stated in the paper.

It may also be thought, that for a function of the observations to be called an estimator, it must be in some sense close to the true value of the parameter; this aspect is examined by computing an asymptotio lower bound for the probability of an estimator lying in a fixed, but small, interval enclosing the true value as the sample size tends to infinity and laying down a criterion by which an estimator attaining this lower bound can be identified.

2 - FIRST ORDER EFFICIENCY -

Asymptotic efficiency of a consistent estimator T_n of a parameter ϑ is defined as the ratio stet $1/i$, where i is the information (as defined by Fisher [7]) per single observation, to the asymptotic variance of $\sqrt{n}\,(T_n - \vartheta)$. It is believed that (i) for a consistent estimator the asymptotic variance under consideration has the lower bound $1/i$ so that an estimator which has its asymptotic variance equal to $1/i$ is fully efficient, and (ii) an estimator with the smallest asymptotic variance is more useful than others in problems of statistical inference. We shall examine to what extent these two statements are valid.

Hodges constructed the following example (quoted in a paper by Lecam [12]). Let $\bar{X}_n$ be the mean of n independent observations on X from a normal distribution with mean ϑ and standard deviation unity. Consider the statistic

$$\begin{aligned} T_n &= \bar{X}_n \quad \text{if} \quad |\bar{X}_n| \geq n^{-1/4} \\ &= \alpha \bar{X}_n \quad \text{if} \quad |\bar{X}_n| < n^{-1/4} \end{aligned} \qquad (2.1)$$

It is easy to see that $\sqrt{n}\,(T_n - \vartheta)$ is also asymptotically normally distributed with variance 1 for $\vartheta \neq 0$ and α^2 for $\vartheta = 0$. Since α is arbitrary the asymptotic variance can be made as small as possible when $\vartheta = 0$. This example shows that statement (i) regarding the existence of a lower bound to asymptotic variance is not strictly true. Estimators such as (2.1) are called 'super efficient' since the ratio of $1/i$ to asymptotic variance exceeds unity at least for some values of the parameter.

It was pointed out by Kallianpur and Rao [11] that this situation does not arise if the estimator is Fisher consistent and Frechet differentiable as a functional of the empirical distribution function. Under these conditions $1/i$ is shown to be a lower bound to asymptotic variance. Earlier work by Neyman [14] and Barankin and Gurland [2] in some special cases also confirms this result. But the conditions imposed on the estimator are very restrictive.

In order to examine whether asymptotic variance is a reliable indicator of the usefulness of an estimator in statistical inference let us consider the following example, where X_m denotes the median of n independent observations from a normal population with variance unity. Define the statistic

$$\begin{aligned} T_n &= \alpha X_m \quad \text{if} \quad |\bar{X}| < n^{-1/4} \\ &= \bar{X}_n \quad \text{if} \quad |\bar{X}| \geq n^{-1/4} \end{aligned} \qquad (2.2)$$

The asymptotic distribution of $\sqrt{n}\,(T_n - \vartheta)$ is normal with variance $\alpha^2\pi/2$ when $\vartheta = 0$ and 1 when $\vartheta \neq 0$. Since α is arbitrary T_n is 'super efficient', but obviously is less useful than $\bar{X}_n$, which is sufficient for ϑ, in problems of inference. For instance a test based on T_n defined in (2.2), for the hypothesis $\vartheta = 0$ essentially depends on the median which is known to be less powerful than the mean.

In this connection we may also refer to an interesting but a different type of example due to Basu [3], where the ratio of the limiting variance of one statistic to that of another $\longrightarrow \infty$ but also the corresponding ratio of probabilities outside any given limits enclosing the true value $\longrightarrow 0$. Another example investigated by Sethuraman [23] shows that although one statistic has a smaller variance than another uniformly, a test based on the former is less powerful for a neighbourhood of values close to the value of the parameter under test. The criterion of minimum variance, by itself, is therefore somewhat misleading.

To define the new criterion of efficiency let us introduce some notations. We shall consider the case of identical and independently distributed observations and denote by $P(X_n, \vartheta)$ the probability density of the sample X_n in the n dimensional Euclidean space E^n and by $P(T_n, \vartheta)$ the corresponding density for the statistic T_n. The first derivatives of $P(X_n, \vartheta)$ and $P(T_n, \vartheta)$ assumed to exist, are denoted by $P'(X_n, \vartheta)$ and $P'(T_n, \vartheta)$. Let :

$$Z_n = n^{-\frac{1}{2}} \frac{P'(X_n, \vartheta)}{P(X_n, \vartheta)}$$

$$Y_n = n^{-\frac{1}{2}} \frac{P'(T_n, \vartheta)}{P(T_n, \vartheta)}$$

$$V(Z_n) = i(\vartheta) \quad \text{and} \quad V(Y_n) = i_T(\vartheta).$$

The functions $i(\vartheta)$ and $i_T(\vartheta)$ correspond to the information per observation contained in the sample and in the statistic respectively.

<u>Definition of first order efficiency</u> - T_n is said to have first order efficiency if

$$|Z_n - n^{\frac{1}{2}} \beta(\vartheta)\,(T_n - \vartheta)| \longrightarrow 0 \text{ in probability} \qquad (2.3)$$

where $\beta(\vartheta)$ is a function of ϑ only.

We shall examine the consequences of such a definition by studying the properties of T_n based on the condition (2.3).

Firstly the condition (2.3) implies that the asymptotic correlation between Z_n and $\sqrt{n}(T_n - \vartheta)$ is unity. If T_n' is an alternative consistent estimator of ϑ, then its efficiency can be defined as the square of its asymptotic correlation with Z_n. The pivotal quantity Z_n considered as a function of the observations and the parameter plays a fundamental role in statistical inference (i.e., in providing optimum procedures of testing of hypotheses, setting up confidence limits etc.) in large samples as shown by Wald [24, 25] and Rao and Poti [22]. By demanding that the asymptotic correlation should be unity we are hoping that T_n can be used as a good substitute for Z_n in sufficiently large samples.

On the basis of this criterion, Hodges' 'super efficient' estimator (2.1) is efficient in the new sense and is asymptotically equivalent to $\bar{X}_n$ i.e., has asymptotic correlation unity with $\bar{X}_n$ although their asymptotic variances are different. On the other hand the 'super efficient' estimator (2.2) is not fully efficient although it has a smaller asymptotic variance than $\bar{X}_n$, its efficiency in the new sense being $2/\pi < 1$ at $\vartheta = 0$. Thus the new definition enables us to distinguish between estimators such as those defined in (2.1) and (2.2) without restricting the class of estimators to well behaved functions of observations.

Secondly it has been established (Doob [5, 6], Rao [20]) that under some mild regularity conditions on the probability density $P(X_n, \vartheta)$, the condition (2.3) implies that $i_T \longrightarrow i$ as $n \longrightarrow \infty$, i.e., the limiting information per observation in the statistic tends to the information per observation in the entire sample. This is important since $n\, i_T$ provides in some sense a measure of distance (for discrimination) based on the distribution of T_n, between alternative values of the parameter close to each other (Rao [15, 19]), where as the distance based on the distribution of the entire sample is $n\, i$. It is known that $n i_T \leq n i$ and what has been shown is that $n i_T / n i \longrightarrow 1$, giving an assurance that for discriminating between alternative values of the parameter close to one another, the performance of the statistic is *as good as* that of the entire sample as the size of the sample increases.

Thirdly, if $r_n(\vartheta)$ denotes the power function of the test criterion $\sqrt{n}(T_n - \vartheta_0) \geq \lambda$ for testing the hypothesis $\vartheta = \vartheta_0$ against the alternatives $\vartheta > \vartheta_0$, where λ is chosen such that the limiting level of significance is $\alpha\phi$ then the condition (2.3) implies that :

$$\lim_{n \to \infty} n^{-\frac{1}{2}} r_n'(\vartheta_0) = \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp(-a^2/2) \tag{2.4}$$

where $r_n'(\vartheta)$ is the derivative of $r_n(\vartheta)$ and a is the $\alpha\phi$ point of the standard normal deviate. It is known from a lemma proved by the author (Lemma 2.1 in [21]) that for any test :

$$\overline{\lim_{n \to \infty}}\, n^{-\frac{1}{2}} r_n'(\vartheta_0) \leq \left(\frac{1}{2\pi}\right)^{\frac{1}{2}} \exp(-a^2/2). \tag{2.5}$$

A comparison of (2.4) with (2.5) shows that a test based on an efficient estimator has locally good power in large samples. For any statistic with asymptotic correlation $\rho < 1$ with Z_n, i.e., not efficient in the new sense, the limit of the left hand side is ρ times that of the right hand side in (2.4). From this it follows that a test based on an inefficient estimator such as the 'super efficient' estimator (2.2), is locally less powerful for at least a small neighbourhood of ϑ_0 for sufficiently large n, although this neighbourhood may depend on n.

Since the problem of confidence intervals and testing of hypothesis are inter-related, similar optimum properties are expected of confidence intervals based on first order efficient estimators.

3 - STRONGER FIRST ORDER EFFICIENCY -

We will now state a stronger form of first order efficiency and deduce the properties of estimators satisfying it.

Definition of stronger first order efficiency - A statistic T_n is said to have stronger first order efficiency as a consistent estimator of ϑ if :

$$\text{Prob.}\ (|n^{-\frac{1}{2}} Z_n - \beta(\vartheta)\ (T_n - \vartheta)| \geq \varepsilon |T_n - \vartheta|\ |\vartheta) < \rho_\varepsilon^n \tag{3.1}$$

or alternatively :

$$\text{Prob.}\ (|n^{-\frac{1}{2}} Z_n - \beta(\vartheta)\ (T_n - \vartheta)| \geq \varepsilon \mid n^{-\frac{1}{2}} Z_n \mid \mid \vartheta) < \rho_\epsilon^n \tag{3.2}$$

where $\rho_\epsilon < 1$ and is independent of ϑ for a small interval ϑ round the true value.

It may be seen that the condition (3.1) or (3.2) implies a stronger stochastic relationship between Z_n and $(T_n - \vartheta)$ than that implied by (2.3) of first order efficiency.

It has been proved in an earlier paper (Rao [21]) that if $\beta_n(\vartheta)$ is the second kind of error for any test of the hypothesis $\vartheta = \vartheta_0$ at a level of significance tending to a fixed value α, $0 < \alpha < 1$ as $n \to \infty$, then :

$$\underline{\lim}_{\theta \to \theta_0}\ \underline{\lim}_{n \to \infty} \frac{n^{-1} \log \beta_n(\vartheta)}{(\vartheta - \vartheta_0)^2} \geqslant -\frac{i(\vartheta_0)}{2} \tag{3.3}$$

Under the condition that Z_n admits a moment generating function and certain other mild conditions on Z_n as a function of ϑ it has also been shown that, for a test based on T_n satisfying the condition (3.1) or (3.2)

$$\lim_{\theta \to \theta_0}\ \lim_{n \to \infty} \frac{n^{-1} \log \beta_n(\vartheta)}{(\vartheta - \vartheta_0)^2} = -\frac{i(\vartheta_0)}{2} \tag{3.4}$$

The result (3.4) shows that a test based on T_n satisfying stronger first order efficiency has as good power or perhaps better than any other given test in small intervals of ϑ round the hypothetical value for all sufficiently large n. It may be noted that the corresponding result establishing local power in the case of estimators satisfying first order efficiency is somewhat weaker.

The results (3.3) and (3.4) can be restated in a form suitable for studying the probability of concentration of an estimator in fixed intervals round the true value as $n \to \infty$. It was thought that a statistic with a smaller asymptotic variance has necessarily higher concentration in the neighbourhood of the true value. This is true only when intervals of the type $(\vartheta_0 \pm \lambda\ n^{-\frac{1}{2}})$ which tend to zero as $n \to \infty$ are taken and the limiting value of the probability of concentration in such intervals is considered. Some other properties must be satisfied to ensure higher concentration in fixed intervals round the true value as the sample size increases. This line of work was initiated by Bahadur [1], whose results are derived here as a consequence of the stronger first order efficiency of estimators.

It may be seen that the result (3.3) can be restated in the form :

$$\underline{\lim}_{h \to o}\ \underline{\lim}_{n \to \infty} \frac{\log \Pr.(|T_n - \vartheta_0| > h)}{nh^2} \geq -\frac{i(\vartheta_0)}{2} \tag{3.5}$$

and for any consistent estimator of ϑ. For an estimator T_n satisfying the stronger first order efficiency :

$$\lim_{h \to o}\ \lim_{n \to \infty} \frac{\log \Pr.(|T_n - \vartheta_0| > h)}{nh^2} = -\frac{i(\vartheta_0)}{2} \tag{3.6}$$

The result (3.6) shows that a statistic T_n satisfying (3.1) or (3.2), has in some sense maximum concentration in small intervals round the true value as $n \to \infty$.

It may be enquired, under what conditions and for which methods of estimation, the weaker and stronger forms of efficiency hold. In the case of the multinomial distribution, which has been studied somewhat thoroughly, it is known that estimators exist for which condition (2.3) of first order efficiency is satisfied when the cell probabilities admit continuous first derivatives only (Rao, [18]) and the stronger conditions (3.1) and (3.2) are satisfied when continuous second derivatives exist (Rao, [16], [17]). If the parameter chosen is a continuous functional of the distribution function, such estimates may be derived by methods such as m.l., minimum chi-square, minimum modified chi-square (Neyman, [14]), minimum discrepancy (Haldane [10]), etc. It will be shown in section 4 that although all these methods lead to first order efficient estimators, they could be distinguished by another measure to be defined as the second order efficiency.

In the case of continuous distributions no comprehensive discussion is available to answer all the questions relating to first order efficiencies, except for a recent contribution due to Bahadur [1], who imposes rather severe restrictions on the probability density. Partial answers, however, exist in the papers by Cramer [4], Doob [5, 6], LeCam [12 , 13] and others.

4 - SECOND ORDER EFFICIENCY -

First order efficiency states that under a suitable norming factor $(T_n - \vartheta)$ is close to Z_n in large samples, in the sense that the difference $\longrightarrow 0$ in probability. There exist, indeed, a large number of estimation procedures which lead to estimators satisfying this condition. We may then try to distinguish among them by constructing a measure of the rate of convergence of the difference between Z_n and $\sqrt{n}\,(T_n - \vartheta)$. For this let us consider :

$$W_n = n^{\frac{1}{2}} [Z_n - n^{\frac{1}{2}} \beta(\vartheta) \; (T_n - \vartheta)] \tag{4.1}$$

which is $\sqrt{n}$ times the difference occurring in the condition (2.3) of first order efficiency and which may not converge to zero in probability. What is relevant is not the distribution of W_n by itself but its conditional distribution given T_n or some measure of variability of W_n given T_n. We may then define the limiting average conditional variance of W_n given T_n as second order efficiency. The importance of such a definition may be seen from the fact that the average conditional variance for any finite n is exactly

$$ni - ni_T \tag{4.2}$$

which is the difference between the actual amounts of information contained in the sample and in the statistic. It may be recalled that first order efficiency ensures that :

$$\lim_{n \to \infty} \; (i - i_T) = 0 \tag{4.3}$$

while the concept of second order efficiency is based on minimising the expression

$$\lim_{n \to \infty} \; (ni - ni_T). \tag{4.4}$$

It is extremely difficult to evaluate (4.4) except in special cases by using the actual knowledge of the joint distribution of T_n and W_n.

On the other hand it is somewhat simpler to evaluate, in a general way, the average conditional variance of W_n given $\sqrt{n}\,(T_n - \vartheta)$ from their joint asymptotic distribution, which we shall adopt as second order efficiency and denote it by E_2. In many cases it is possible that :

$$\lim_{n \to \infty} \; (ni - ni_T) = E_2 \tag{4.5}$$

when our definition will be satisfactory. It is worth investigating some general conditions under which (4.5) is true.

The quantity E_2 as defined can, however, be directly computed as the minimum asymptotic variance (minimised with respect to λ) of the statistic

$$n^{\frac{1}{2}} \; [Z_n - n^{\frac{1}{2}} \beta(\vartheta) \; (T_n - \vartheta) - \lambda \, n^{\frac{1}{2}} (T_n - \vartheta)^2]. \tag{4.6}$$

The exact computation of E_2 has been carried out in the case of the multinomial distribution in k classes with the following conditions on the probabilities, the parameter and the estimation procedure.

(i) The cell probabilities represented by $\pi_1(\vartheta), \dots, \pi_k(\vartheta)$ admit continuous derivatives up to the second order.

(ii) The parameter ϑ under consideration is a continuous functional of the distribution function.

(ii) The estimating equation :

$$f\left(\vartheta, \frac{n_1}{n}, \dots, \frac{n_k}{n}\right) = 0$$

where $n_1, \dots, n_k$ are observed frequencies in the k classes, is consistent, i. e., $f(\vartheta, \pi_1(\vartheta), \dots, \pi_k(\vartheta)) \equiv 0$ and has continuous derivatives up to the second order in ϑ as well as in n_i/n considered as variables.

Under the conditions (i), (ii) and (iii) it has been shown by the author [20] that :

$$E_2 \geqslant \frac{\mu_{02} - 2\mu_{21} + \mu_{40}}{i} - i - \frac{\mu_{11}^2 + \mu_{30}^2 - 2\mu_{11}\mu_{30}}{i} = \mu > 0$$

where :

$$\mu_{rs} = \Sigma\, \pi_j(\pi_j' / \pi_j)^r\, (\pi_j''/\pi_j)^s\,,\quad i = \mu_{20}$$

and further that in the case of an m.l. estimator the lower bound μ is actually attained. The following Table gives the values of E_2 for a number of methods of estimation, where in the value of Δ is

$$\frac{1}{2}\Sigma \left(\frac{\pi_r'}{\pi_r}\right)^2 - \frac{\mu_{40}}{i} + \frac{\mu_{30}^2}{2\,i^2} \geq 0$$

Table

Second order efficiencies of different methods of estimation applicable to a multinomial distribution.

method of estimation	function to be maximised or minimised	E_2
1. max. likelihood	$\Sigma\, n_i \log \pi_i$	μ
2. min. chi-square	$\Sigma\, (n_i - n\pi_i)^2/n\pi_i$	$\mu + \Delta$
3. min. mod. chi-square	$\Sigma\, (n_i - n\pi_i)^2/n_i$	$\mu + 4\Delta$
4. min. discrepancy	$\Sigma\, \pi_i^{k+1}/n_i^k$	$\mu + (k+1)^2\Delta$
5. min. Kullback-Liebler separator	$\Sigma\, \pi_i \log (\pi_i/n_i)$	$\mu + \Delta$
6. min. Hellinger distance	$\Sigma \sqrt{n_i \pi_i}$	$\mu + \frac{1}{4}\Delta$

Although all the six methods listed in the Table satisfy the criterion of first order efficiency both in weaker and stronger forms, they are clearly distinguishable by their second order efficiencies. Apart from m.l., which has been shown to have the smallest value of E_2 under the conditions assumed, minimum Hellinger distance appears to be better than the other methods of estimation considered. The method of minimum modified chi-square advocated by Neyman [14] seems to involve a higher loss of information than the usual minimum chi-square method.

REFERENCES

[1] R.R. BAHADUR - On asymptotic efficiency of tests and estimates. *Sankhya,* Vol. 22 (1960), pp. 229-252.

[2] E.W. BARANKIN and J. GURLAND - On asymptotically normal, efficient estimator I. *Univ. California Publ. Statist.,* Vol. 1 (1951), pp. 89-130.

[3] D. BASU - Choosing between two simple hypotheses and the criterion of consistency. *Proc. Nat. Int. Sc. (India),* Vol. 19 (1953), pp. 841-849.

[4] H. CRAMER - *Mathematical methods of Statistics.* Princeton, Princeton University Press (1946).

[5] J. DOOB - Probability and statistic, *Trans. Amer. Math. Soc.* Vol. 36 (1934), pp. 759-775.

[6] J. DOOB - Statistical estimation, *Trans. Amer. Math. Soc.,* Vol. 39 (1936), pp. 410-421.

[7] R.A. FISHER - On the mathematical foundations of theoretical statistics, *Philos. Trans. Roy. Sco. London,* Ser. A., Vol. 222 (1922), pp. 309-368.

[8] R.A. FISHER - Theory of statistical estimation, *Proc. Cambridge Philos. Soc.* Vol. 22 (1925), pp. 700-725.

[9] R.A. FISHER - *Statistical Methods and Scientific Inference*, Oliver and Boyd, London (1956).

[10] J.B.S. HALDANE - A class of efficient estimates of a parameter, *Bull. Int. Statist. Inst.*, Vol. 33 (1951), pp. 231.

[11] G. KALLIANPUR and C.R. RAO - On Fisher's lower bound to asymptotic variance of a consistent estimate, *Sankhya*, Vol. 15 (1955), pp. 331-342.

[12] L. LeCAM - On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *Univ. California Publ. Statist.* Vol. 1 (1953), pp. 227-330.

[13] L. LeCAM - On the asymptotic theory of estimation and testing of hypotheses, *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability*, 1, (1956), pp. 129-156.

[14] J. NEYMAN - Contribution to the theory of the X^2 test, *Proceedings of the Berkeley Symposium on Methematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, (1949), pp. 239-273.

[15] C.R. RAO - Information and Accuracy attainable in the estimation of statistical parameters, *Bulletin of the Calcutta Mathematical Society*, Vol. 37 (1945), pp. 81-91.

[16] C.R. RAO - Maximum likelihood estimation for the multinomial distribution, *Sankhya*, Vol. 18 (1957), pp. 139-148.

[17] C.R. RAO - Maximum likelihood estimation for the multinomial distribution with infinite number of cells, *Sankhya*, Vol. 20 (1958) pp. 211-218.

[18] C.R. RAO - A study of large sample test criteria through properties of efficient estimates, *Sankhya*, Vol. 23 (1960), pp. 25-40.

[19] C.R. RAO - Apparent anomalies and irregularities of the method of maximum likelihood, *Bulletin of the International Statistical Institute*, Vol. XXXVIII, Part IV (1961), pp. 439-453.

[20] C.R. RAO - Asymptotic efficiency and limiting information. *Proceedings of the Fourth Berkeley Symposium on Mathematical statistics and Probability ;* University of California Press (1961), pp. 531-545.

[21] C.R. RAO - Efficient estimates and optimum inference procedures in large sample. *To appear in J.R. Statist. Soc.* (1962).

[22] C.R. RAO and S.J. POTI - On locally most powerful tests when alternatives are one-sided, *Sankhya*, Vol. 7 (1946), pp. 439.

[23] J. SETHURAMAN - Conflicting criteria of goodness of statistics, *Sankhya*, Vol. 23 (1961), pp. 187-190.

[24] A. WALD - Asymptotically shortest confidence intervals, *Ann. Math. Statist.*, Vol. 13 (1942), pp. 127-137.

[25] A. WALD - Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426-482.

[26] A. WALD - *Statistical Decision Functions*, John Wiley and Sons, New York, 1950.

DISCUSSION

M. FERON - Is there any difficulty in generalising the results to the multiparameter case ?

M. RAO - In the case of more than one parameter the first and second order efficiencies cam be defined in a similar way. If there are k parameters we consider the vector Z of the derivatives :

$$\frac{1}{\sqrt{n}} \frac{\partial \operatorname{Log} P(X_n, \vartheta)}{\partial \vartheta_i}, \quad i = 1, \ldots, k$$

where Xn is the sample point and ϑ stands for the vector of parameters. If T_n represents the vector estimate, then first order efficiency of the estimate may be defined as the property that :

$$|Z_n - \sqrt{n}\ \beta(\vartheta)\ (T_n - \vartheta)| \longrightarrow 0$$

in probability where $\beta(\vartheta)$ is a square matrix of order k. The condition is of the same form as (2.3.) in the case of a single parameter except that the variables involved are vectors.

To define the second order efficiency let us consider the vector variable

$$W_n = n^{\frac{1}{2}}[Z_n - n^{\frac{1}{2}}\beta(\vartheta)\ (Tn - \vartheta)]$$

and define E_2 as the matrix of average conditional variances and covariances of Wn given T_n as computed from the joint asymptotic distribution of Wn and $n^{\frac{1}{2}}(T_n - \vartheta)$.

It may be shown, as in the case of a single parameter that there exists a matrix μ such that $E_2 - \mu$ is at least semi-positive definite. For maximum likelihood estimation $E_2 = \mu$.