CRITERIA OF ESTIMATION IN LARGE SAMPLES*

By C. RADHAKRISHNA RAO

Indian Statistical Institute

SUMMARY. The existing criteria of consistency and efficiency of estimation have been examined in the light of recent criticisms and controversies concerning them. A new criterion called *uniform first order efficiency* which is a better indicator of the performance of an estimator in statistical inference has been introduced. It is, however, pointed out that the anomaly in the earlier criterion of efficiency can be removed by considering consistent estimators which converge to a normal distribution uniformly in compacts of the parameter space. First order efficiency by itself cannot discriminate among a large number of estimation procedures. Therefore, an additional criterion called the *second order efficiency* has been introduced, which considerably restricts the class of useful estimation procedures and by which several well established estimation procedures could be eliminated in favour of the method of maximum likelihood.

1. INTRODUCTION

Estimation, as conceived by the late Sir Ronald Fisher, is one of the methodological processes by which data are analysed or reduced for purposes of drawing inferences on the unknown population from which data are observed. For instance a sample survey of consumer expenditure may provide a mass of data which by themselves are difficult to interpret. We therefore need summary figures or *estimates* which provide a fair idea of the characteristics of the population sampled and enable us to answer a variety of questions. Has the per-capita expenditure on rice increased over time and is it different in different regions? Does a given estimate reasonably agree with what is believed to be the per-capita expenditure, or with another estimate obtained by a parallel agency? No clear indication of answers to such questions would be available without computing from the data an estimate which represents the per-capita expenditure and other quantities which indicate the possible extent of error in the estimate and guide us in making judicious statements about the population. Further questions may suggest themselves after some initial questions are answered with the estimates already obtained.

There has been a tendency to consider the problem of estimation as a part of decision theory, requiring a prestated purpose for the estimate and specification of loss resulting from any given magnitude of error in the estimate. It is not, however, my view that the latter approach should be completely abandoned. There may be situations where such an approach is necessary and appropriate as in the case of acceptance procedures in industrial statistics. But in a majority of situations the framework of decision theory may not be applicable and it may be necessary to consider the problem of estimation from a wider point of view as 'extraction of information' for drawing inferences and for recording it, as a substitute for the entire data, for possible future uses.

Since estimation, however it may be viewed, involves reduction of data, it may entail some loss of information for we are interpreting the data through the

*Lecture delivered on the occasion of the presentation of Shanti Swarup Bhatnagar award for 1959.

This paper has been included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70th birthday.

estimates. The criteria for choice of estimators should then relate to minimisation of loss of information. Unfortunately, no objective measurement of information is possible and hence the difficulty in the formulation of suitable criteria. However, asymptotic theories of estimation based on the criteria of *consistency* and *efficiency* (to be referred to as v-efficiency) have been constructed and certain methods have been shown to yield estimators satisfying these criteria. It was thought that the criteria of consistency and v-efficiency ensure minimum loss of information due to estimation as the sample size increases.

These theories are not satisfactory due to three main reasons. Firstly, all the results relate to limiting properties as the sample size tends to infinity and no indication is available of their applicability to samples of sizes ordinarily met with in actual practice. Secondly, there seem to exist infinitely many procedures leading to estimators satisfying the stated criteria and no further criteria have been suggested to distinguish among them. Thirdly, the criterion of v-efficiency does not provide a satisfactory index of the performance of an estimator from the view point of statistical inference.

I have attempted to resolve these difficulties in some ways (Rao, 1960b, 1961, 1962). Firstly, the criterion of v-efficiency has been reformulated to ensure some optimum asymptotic properties of an estimator used in the place of the sample for purposes of inference. This is called *first order efficiency*. Secondly, another criterion known as *second order efficiency* has been introduced to distinguish among different procedures leading to first order efficient estimators. On the basis of the latter criterion several well-known procedures, such as the minimum chi-square, modified minimum chi-square etc., which are considered as competitors to maximum likelihood on the basis of v-efficiency, could be eliminated. The second order efficiency also provides a partial answer to the question of sample size. Correction terms of order $O(n^{-1})$ to the estimate and of order $O(n^{-2})$ to its precision have been determined for several estimation procedures.

The present paper is intended for a further discussion of first and second order efficiencies and to introduce a new concept of *uniform efficiency* which seems to be important when asymptotic theories are considered. Some new light is thrown on the use of asymptotic variance of an estimator as an index of efficiency. Further the second order efficiency is linked with terms of order (n^{-2}) in the asymptotic expansion of the variance of an estimator. Problems requiring further investigation are indicated.

In undertaking these studies I have been guided by the basic ideas contained in two fundamental papers on estimation by Fisher (1922, 1925). I wish to record my debt of gratitude to the late Sir Ronald Fisher for the encouragement I received from him when I was working under his guidance at Cambridge and during his recent visits to the Indian Statistical Institute. I also wish to thank Professor P. C. Mahalanobis, the Director of the Indian Statistical Institute for his stimulating discussions on the logic of statistical inference and the purpose of statistics to which I have been constantly exposed.

2. Consistency

The criterion of consistency is in the nature of identifying the parameter for which a statistic is said to be an estimator. This is important from the practical point of view of interpreting the estimates. There are various definitions of consistency of which the one frequently referred to in literature is probability consistency (PC).

Definition 2A : *Probability consistency* (PC). A sequence of statistics T_n is said to be consistent for a parameter θ if $T_n \to \theta$ in probability.

But one criticism of such a definition is that it places no restriction on the statistic for any given n. An alternative definition of consistency due to Fisher, called Fisher consistency (FC) seems to be more satisfactory in this respect, but somewhat restrictive in application.

Definition 2B : *Fisher consistency* (FC). A statistic $T_n = f(S_n)$, where S_n is the empirical distribution function based on n observations and f is a weakly continuous functional defined on the space of distribution functions is said to be Fisher consistent if $f(F_\theta) \equiv \theta$, where F_θ is the true distribution function from which observations are drawn.

It is easy to see that FC $\Longrightarrow$ PC and that FC refers to a restriction on the estimate for any finite n and is not just a limiting property of a sequence of statistics. But it is applicable only in situations where independent observations are drawn from a population characterised by a distribution function.

3. Efficiency

Efficiency of an estimator, which we rename as v-efficiency because it is linked with asymptotic variance, is usually defined as follows :

Definition 3A : *v-efficiency.* Consider the class $\{T_n\}$ of consistent asymptotically normal (CAN) estimators of θ, i.e., for each T_n, $n^{\frac{1}{2}}(T_n-\theta)\xrightarrow{L} N[0, v(\theta)]$. Any member of the sub-class for which $v(\theta) = 1/i(\theta)$ is said to be an efficient estimator of θ.

It was believed that for a CAN estimator, the asymptotic variance $v(\theta)$ satisfies the inequality

$$v(\theta) \geqslant \frac{1}{i(\theta)} \qquad \ldots \quad (3.1)$$

and that an estimator with the smallest $v(\theta)$ has maximum concentration round the true value in sufficiently large samples. Unfortunately, both these results are not strictly true without any restrictions on the estimating function or the mode of convergence to normality. About ten years ago Hodges (see LeCam, 1953) constructed an example to show that the result (3.1) is not true in general. Let

$$\left.\begin{aligned} T_n &= \bar{x} \quad (|\bar{x}| \geqslant n^{-1/4}) \\ &= \alpha\, \bar{x}\, (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \qquad \ldots \quad (3.2)$$

where $\bar{x}$ is the average of n observations from $N(\theta, 1)$ and α is arbitrary. It may be verified that T_n is also CAN with

$$\left.\begin{aligned} v(\theta) &= 1, \qquad \text{for} \quad \theta \neq 0 \\ &= \alpha^2, \qquad \text{for} \quad \theta = 0 \end{aligned}\right\}$$

so that the variance at $\theta = 0$ can be made arbitrarily small. Such an estimate has been termed 'super efficient.' This example throws in doubt the exact significance of v-efficiency.

Even if there is no lower bound to asymptotic variance, the question remains as to whether we should prefer the estimator T_n as defined in (3.2) to $\bar{x}$ because of smaller asymptotic variance at least at one point and equivalence elsewhere. It can be easily seen that for any given n, T_n has better concentration than $\bar{x}$, in the sense of higher probabilities for intervals enclosing the true value, only for the special values of $\theta = 0$ and a small neighbourhood of zero, and thereafter for a continuous set of θ, T_n has less concentration than $\bar{x}$. This may also be inferred by comparing the mean square errors (m.s.e.) of T_n and $\bar{x}$. For any given n the m.s.e. of T_n is smaller than that of $\bar{x}$ for θ close to zero and thereafter it stays larger, although the difference tends to zero as θ increases. It may, however, be observed that the m.s.e. in either case tends to the corresponding asymptotic value but the anomaly arises due to convergence being not uniform in the case of T_n. We shall have occasion to stress the importance of uniform convergence in a later section of this paper. An attempt to improve the concentration in the neighbourhood of a particular value of the parameter seems to have injured the performance of the estimator at other values. A general statement to this effect is proved by LeCam (1953) using bounded risk functions. Superiority as judged by asymptotic variance function need not therefore indicate greater concentration for all values of the unknown parameter even in sufficiently large samples.

Consider another super efficient estimator U_n,

$$\left.\begin{aligned} U_n &= \bar{x} \qquad (|\bar{x}| \geqslant n^{-1/4}) \\ &= \alpha x_m \quad (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \quad \ldots \quad (3.3)$$

where x_m is the sample median and α is arbitrarily small. The statistics (3.2) and (3.3) have the same asymptotic variance and are therefore indistinguishable on the basis of v-efficiency. There must, however, be some difference in the performance of these two statistics, the estimator (3.3) being essentially equivalent to the sample median when $\theta = 0$.

Since there is no lower bound to the asymptotic variance of a CAN estimator, it may be thought that an improvement over $\bar{x}$ is possible by constructing a statistic T'_n with a uniformly lower asymptotic variance and thereby increasing the concentration at every value of the parameter, as at $\theta = 0$ in examples (3.2) and (3.3). LeCam (1953) has demonstrated that such an improvement is not possible for any continuous interval of the parameter and the set of points with a lower asymptotic variance has to be of Lebesgue measure zero.

Can we avoid all these troubles by considering only efficient estimators in the sense of Definition 3A and not trying to improve upon the asymptotic variance $1/i(\theta)$? The following example provides an answer to this question.

Let

$$\left.\begin{aligned} W_n &= \bar{x} \qquad\qquad (|\bar{x}| \geqslant n^{-1/4}) \\ &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} x_m \ (|\bar{x}| < n^{-1/4}) \end{aligned}\right\} \quad \ldots \quad (3.4)$$

where $\bar{x}$ is the sample mean and x_m is the sample median. W_n is also CAN with the same asymptotic variance $v(\theta) = 1$ for all θ as that of $\bar{x}$. The estimator W_n is thus indistinguishable from $\bar{x}$ so far as consistency and v-efficiency are concerned. Yet for any given large n, W_n has less concentration than that of $\bar{x}$ for all values of θ.

It is no doubt true that an estimator having a higher concentration than another for every value of θ is more useful in drawing inferences on θ from an observed estimate. That such a situation is realised for an estimator compared to another for sufficiently large n cannot be judged by comparing the asymptotic variances only as shown by examples (3.2), (3.3) and (3.4). It is, however, difficult to choose between two estimators when one does not have uniformly better concentration than another without bringing in other considerations. For instance, we may have an estimator whose distribution for a particular value of θ is highly concentrated but it will be a poor discriminator between this value of θ and other values close to it if the concentration at the other values is low. To compare the estimators $\bar{x}$, T_n, U_n and W_n, we may examine one aspect of their usefulness in statistical inference e.g., the power functions of tests based on these statistics to test the hypothesis that θ has an assigned value. It may be inferred from the optimum properties possessed by $\bar{x}$, that in large samples $\bar{x}$ and T_n tend to have the same local power (Rao, 1962) whereas U_n and W_n being equivalent to the sample median when $\theta = 0$, will have a smaller local power. Since v-efficiency does not enable us to distinguish between estimators such as $\bar{x}$ or T_n and U_n or W_n we shall consider an alternative definition of efficiency (to be called first order) which appears to be more satisfactory.

Definition 3B : *First order efficiency*. A statistic T_n is said to be efficient if

$$n^{\frac{1}{2}}|(T_n-\theta)-\beta(\theta)Z_n| \xrightarrow{L} 0 \qquad \ldots \quad (3.5)$$

where $\beta(\theta)$ is a function of θ only, and $Z_n = n^{-1}[d \log P(X_n, \theta)/d\theta]$, $P(X_n, \theta)$ being the density of the observations. The condition (3.5) implies that the asymptotic correlation between T_n and Z_n is unity.

I have shown elsewhere (Rao, 1960b) that according to definition 3B, T_n is just as efficient as $\bar{x}$, although T_n is super efficient in the sense of v-efficiency and U_n and W_n are not efficient in the new sense at $\theta = 0$ although U_n and W_n are super efficient and efficient respectively in the old sense. If the efficiency of an estimator is measured by the square of its asymptotic correlation with Z_n, then U_n and W_n have the same efficiency $2/\pi < 1$, although U_n and W_n have different asymptotic variances. It is also shown (Theorem 2 in Rao, 1962) that an estimator satisfying, or efficient in the sense of Definition 3B provides a locally more powerful test of a simple hypothesis concerning θ than any other test in sufficiently large samples. Another important consequence of Definition 3B of efficiency is that the ratio of $I(T_n)$ the Fisher's information contained in the estimator T_n to I, the total information in the sample tends to unity as $n \to \infty$ (Doob, 1934; Rao, 1961).

The Definition 3B of efficiency implies that the limiting distribution of $n^{\frac{1}{2}}(T_n-\theta)$ is normal for any given θ and in large samples, any simple hypothesis on θ can be tested by using the normal approximation. But in problems of statistical inference, it is often necessary to express our preference for different values of θ, on the basis of the estimate as in the case of interval estimation, and not just examine whether a particular value is true or not. There is thus for a given n, a need to consider the whole set of distributions of the estimator for all values of θ at least in a small interval (in large samples) where different values of θ have to be distinguished. If the distributions are to be approximated by appropriate normal distributions, it seems to be a logical necessity that the convergence to normality of the chosen estimator should be uniform in compacts of θ. Under fairly general conditions the convergence to normality of $n^{\frac{1}{2}}Z_n(\theta)$ is found to be uniform in which case the desired property is assured by the following definition of uniform first order efficiency.

Definition 3C : *Uniform first order efficiency.* An estimator is said to have uniform first order efficiency if

$$n^{\frac{1}{2}}|T_n-\theta-Z_n(\theta)/i(\theta)| \xrightarrow{UL} 0 \quad \dots \quad (3.6)$$

in compacts of θ, where the symbol UL stands for uniform convergence in law and $i(\theta)$ is Fisher's information per observation.

It would have been more natural to define uniform first order efficiency as

$$n^{\frac{1}{2}}|T_n-\theta-\beta(\theta)\, Z_n(\theta)| \xrightarrow{UL} 0 \quad \dots \quad (3.7)$$

without specifying the value of $\beta(\theta)$ as in (3.6). It appears that if the condition (3.7) is satisfied for various values of $\beta(\theta)$, then it is desirable to choose an estimator for which $\beta(\theta)$ is a minimum which is shown to be $[i(\theta)]^{-1}$ in section 4 of this paper.

4. Some lemmas

Notations and assumptions. We consider only sequences of independent and identically distributed variables with probability density $p(x, \theta)$, where θ is a parameter with values in an open interval Θ. In the case of discrete variables, $p(x, \theta)$ represents the probability of x. The probability density of n observations is denoted by $P(X_n, \theta)$. The first derivative $p'(x, \theta) = dp(x, \theta)/d\theta$ exists. Let $a(x, \theta) = p'(x, \theta)/p(x, \theta)$ and

$$i(\theta) = E_\theta[a(x, \theta)]^2$$

Fisher's information per observation be continuous in θ. The following assumptions are made in the various lemmas of this section.

Assumption I : (i) $\mu(\theta_0, \theta) = E_\theta[a(x, \theta_0)] = (\theta-\theta_0)i(\theta_0)+o(\theta-\theta_0)$

(ii) $[\sigma(\theta_0, \theta)]^2 = V_\theta[a(x, \theta_0)] = i(\theta_0)+o(1)$

(iii) $c(\theta_0, \theta) = \mathrm{cov}_\theta\,[a(x, \theta), a(x, \theta_0)] = i(\theta_0)+o(1)$

Assumption II : $E_\theta\left|\dfrac{p'(x, \theta)}{p(x, \theta)}\right|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ in compacts of θ.

Assumption III :

Let $b(x, \theta, \theta_0) = \log[p(x, \theta)/p(x, \theta_0)]$

$$\text{(i)} \quad \xi(\theta, \theta_0) = E_{\theta_0}[b(x, \theta, \theta_0)] = -\frac{(\theta-\theta_0)^2}{2} i(\theta_0) + o(\theta-\theta_0)^2$$

$$\text{(ii)} \quad [\eta(\theta, \theta_0)]^2 = V_{\theta_0}[b(x, \theta, \theta_0)] = (\theta-\theta_0)^2 i(\theta_0) + o(\theta-\theta_0)^2$$

$$\text{(iii)} \quad \zeta(\theta, \theta_0) = \text{cov}_{\theta_0}[b(x, \theta, \theta_0), a(x, \theta_0)] = (\theta-\theta_0) i(\theta_0) + o(\theta-\theta_0)$$

Assumption IV : (i) $\int_{E_n} \frac{d^2}{d\theta^2}[P(X_n, \theta)]dv = \frac{d^2}{d\theta^2} \int_{E_n} P(X_n, \theta)dv$

for every Lebesgue measurable set E_n,

(ii) $E_\theta \left| \frac{p''(x, \theta)}{p(x, \theta)} \right|$ is bounded in compacts of θ.

The Assumptions I, II, and III are not severe. Conditions may be imposed directly on the probability density to ensure them. For instance restrictions such as those imposed by Danials (1961) on the probability density will imply the conditions (i)-(iii) of Assumption I.

Lemma 1 : *Let $\beta_n(\theta)$ be the power function of any test of the hypothesis $\theta = \theta_0$ based on a sample of n independent observations, at probability level α. Then under Assumptions* I-III

$$\overline{\lim_{n\to\infty}} \beta(\theta_0+\delta n^{-\frac{1}{2}}) \leqslant \Phi(a-\delta i^{\frac{1}{2}}) \qquad \dots \quad (4.1)$$

where Φ is the distribution function of $N(0, 1)$ and a is the upper α point of $N(0, 1)$.

The limit of $\beta(\theta_0+\delta n^{-\frac{1}{2}})$ when it exists is known as Pitman power of the test. Lemma 1 gives an upper bound to Pitman power under some conditions on the probability density of the observations. Two limit theorems of a different type concerning the local power of a test have been given in an earlier paper of the author (Rao, 1962).

Let

$$Z_n(\theta) = \frac{1}{n} \frac{P'(X_n, \theta)}{P(X_n, \theta)} = \frac{1}{n} \Sigma a(x_i, \theta)$$

$$Y_n = \frac{1}{n} \log \frac{P(X_n, \theta)}{P(X_n, \theta_0)} = \frac{1}{n} \Sigma b(x_i, \theta, \theta_0)$$

$$u_n(\theta) = n^{\frac{1}{2}}[Y_n - \xi(\theta, \theta_0)]/\eta(\theta, \theta_0)$$

$$v_n(\theta) = n^{\frac{1}{2}}[Y_n + \xi(\theta_0, \theta)]/\eta(\theta_0, \theta)$$

$$w_n(\theta) = n^{\frac{1}{2}} Z_n(\theta)/[i(\theta)]^{\frac{1}{2}}.$$

Under the Assumptions I, II, and III, it is easy to show that

$$\text{(i)} \quad V_{\theta_0}[u_n(\theta) - w_n(\theta_0)] \to 0 \quad \text{as} \quad \theta \to \theta_0 \qquad \dots \quad (4.2)$$

$$\text{(ii)} \quad V_\theta[v_n(\theta) \frac{\eta(\theta_0, \theta)}{\eta(\theta, \theta_0)} - w_n(\theta)] \to 0 \quad \text{as} \quad \theta \to \theta_0 \qquad \dots \quad (4.3)$$

$$\text{(iii)} \quad w_n(\theta) \xrightarrow{UL} N(0, 1) \text{ in compacts of } \theta. \qquad \dots \quad (4.4)$$

The best test of the hypothesis $H_0 : \theta = \theta_0$ against the alternative $\theta_n = \theta_0+\delta n^{-\frac{1}{2}}$ is

$$u_n(\theta_n) \geqslant c_n \quad \dots \quad (4.5)$$

where c_n is chosen such that the size of the test $\to \alpha$ as $n \to \infty$.

$$\lim_{n\to\infty} P_{\theta_0}[u_n(\theta_n) \geqslant c_n] = \lim_{n\to\infty} P_{\theta_0}[u_n(\theta_n)-w_n(\theta_0)+w_n(\theta_0) \geqslant c_n]$$

$$= \lim_{n\to\infty} P_{\theta_0}[w_n(\theta_0) \geqslant c_n] \text{ by (4.2).}$$

Since the limiting distribution of $w_n(\theta_0)$ is $N(0, 1)$, $c_n \to a$ the upper α point of $N(0, 1)$.

The power of the test (4.5) is

$$\beta^*(\theta_n) = P_{\theta_n}(u_n(\theta_n) \geqslant c_n)$$

$$= P_{\theta_n}(u_n(\theta_n)-w_n(\theta_n)+w_n(\theta_n) \geqslant c_n$$

$$= P_{\theta_n}\left\{ v_n(\theta_n) \frac{\eta(\theta_0, \theta_n)}{\eta(\theta_n, \theta_0)} - w_n(\theta_n)+w_n(\theta_n) > c_n + \frac{n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]}{\eta(\theta_n, \theta_0)} \right\}$$

writing $u_n(\theta_n)$ in terms of $v_n(\theta_n)$ using their definitions.

$$\lim_{n\to\infty} \beta^*_n(\theta_n) = \lim_{n\to\infty} P_{\theta_n} \left\{ w_n(\theta_n) > c_n + \frac{n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]}{\eta(\theta_n, \theta_0)} \right\}$$

$$= \Phi(a-\delta i^{\frac{1}{2}}) \qquad \text{using (4.4) of uniform convergence,}$$

where $-\delta i^{\frac{1}{2}} = \lim_{n\to\infty} n^{\frac{1}{2}}[\xi(\theta_0, \theta_n)+\xi(\theta_n, \theta_0)]/\eta(\theta_n, \theta_0)$. The result of Lemma 1 follows by observing that $\beta^*_n(\theta) \geqslant \beta_n(\theta)$ for each θ, where $\beta_n(\theta)$ is the power of any other test.

Lemma 2 : *Let* $n^{\frac{1}{2}}(T_n-\theta) \xrightarrow{UL} N(0, [\gamma(\theta)]^2)$ *in compacts of* θ, *where* $\gamma(\theta)$ *is bounded. Then*

(i) $\gamma(\theta)$ *is continuous if the probability density* $p(x, \theta)$ *is continuous in* θ.

(ii) $[\gamma(\theta)]^2 \leqslant 1/i(\theta)$ *under Assumptions* I–III.

We use an argument similar to that of LeCam (1960) to prove (i) of lemma 2 :

If $p(x, \theta)$ is continuous in θ the distribution function $F_{\theta,n}$ of T_n is continuous in θ and consequently the characteristic function $c_n(t, \theta)$ of $U_n = n^{\frac{1}{2}}(T_n-\theta)$ is continuous in θ. Since U_n converges uniformly, $c_n(t, \theta)$ converges uniformly to $c(t, \theta)$ the characteristic function of the asymptotic distribution $N(0, [\gamma(\theta)]^2)$. But $c(t, \theta)$ is continuous. Hence $\gamma(\theta)$ is continuous in the interval of the uniform convergence of U_n.

Let us consider the test

$$\frac{n^{\frac{1}{2}}(T_n-\theta_0)}{\gamma(\theta_0)} \geqslant \lambda_n$$

of the hypothesis $\theta = \theta_0$, at a probability level α. The power of the test at θ is

$$\beta_n(\theta) = P_\theta\{n^{\frac{1}{2}}(T_n-\theta_0) > \lambda_n \gamma(\theta_0)\}$$

$$= P_\theta \left\{ \frac{n^{\frac{1}{2}}(T_n-\theta)}{\gamma(\theta)} > \lambda_n \frac{\gamma(\theta_0)}{\gamma(\theta)} - \frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\gamma(\theta)} \right\}.$$

Substituting $\theta = \theta_0+\delta n^{-\frac{1}{2}}$ and observing that the convergence to normality of $n^{\frac{1}{2}}(T_n-\theta)$ is uniform in θ, we find

$$\lim_{n\to\infty} \beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = \Phi[a-\delta/\gamma(\theta_0)] \qquad \dots \quad (4.6)$$

where the argument of Φ in (4.6) is the limit of

$$\frac{\lambda_n\gamma(\theta_0)}{\gamma(\theta)} - \frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\gamma(\theta)}$$

with $\theta = \theta_0+\delta n^{-\frac{1}{2}}$, as $n\to\infty$.

It is shown in Lemma 1 under Assumptions I–III

$$\overline{\lim_{n\to\infty}}\, \beta_n(\theta_0+\delta n^{-\frac{1}{2}}) \leqslant \Phi\,(a-\delta i^{\frac{1}{2}}).$$

Hence from (4.6)

$$\Phi(a-\delta/\gamma(\theta_0)) \leqslant \Phi\,(a-\delta i^{\frac{1}{2}})$$

or
$$a-\delta/\gamma(\theta_0) \geqslant a-\delta i^{\frac{1}{2}}$$

i.e.,
$$\gamma^2(\theta_0) \geqslant 1/i(\theta_0) \quad \text{(for any given } \theta_0).$$

We thus see that the asymptotic variance of CUAN (consistent uniformly asymptotically normal) estimator has Fisher's lower bound $1/i(\theta)$ when the probability density satisfies some regularity conditions. It appears then that in the examples of Hodges and LeCam, super efficiency in the sense of having asymptotic variance less than $1/i(\theta)$ has been achieved at the sacrifice of uniform convergence.

Lemma 3 : *Let*

$$n^{\frac{1}{2}}\left\{\frac{Z_n\theta)}{[i(\theta)]^{\frac{1}{2}}} - \frac{T_n-\theta}{\gamma(\theta)}\right\} \xrightarrow{UL} 0. \qquad \dots \quad (4.7)$$

Then

(i) $n^{\frac{1}{2}}(T_n-\theta)\xrightarrow{UL} N(0,\, [\gamma(\theta)]^2)$ *in compacts of* θ, *where* $\gamma(\theta)$ *is continuous, under Assumption* II *and continuity of* $p(x, \theta)$, *and*

(ii) $\gamma(\theta) = [i(\theta)]^{-\frac{1}{2}}$ *under Assumptions* II *and* IV.

Under Assumption II,

$$\frac{n^{\frac{1}{2}}Z_n(\theta)}{[i(\theta)]^{\frac{1}{2}}} \xrightarrow{UL} N(0, 1) \qquad \dots \quad (4.8)$$

and hence
$$\frac{n^{\frac{1}{2}}(T_n-\theta)}{\gamma(\theta)} \xrightarrow{UL} N(0, 1) \qquad \dots \quad (4.9)$$

since by the condition (4.7) of Lemma 3, the difference of (4.8) and (4.9) $\xrightarrow{UL} 0$, Hence the result (i) of Lemma 3 follows.

Consider the test

$$n^{\frac{1}{2}}(T_n-\theta_0) \geqslant c_n\gamma(\theta_0) \qquad .. \quad (4.10)$$

of the hypothesis $\theta = \theta_0$ at a probability level α, where $c_n \to a$, the upper α probability point of $N(0, 1)$. The power of the test (4.10) at $\theta_n = \theta_0+\delta n^{-\frac{1}{2}}$ is

$$\beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = P_{\theta_n}(n^{\frac{1}{2}}(T_n-\theta_0) \geqslant c_n\gamma(\theta_0)$$

$$= P_{\theta_n}\left\{\frac{n^{\frac{1}{2}}(T_n-\theta)}{\gamma(\theta)} \geqslant \frac{c_n\gamma(\theta_0)}{\gamma(\theta)} - \frac{n^{\frac{1}{2}}(\theta-\theta_0)}{\gamma(\theta)}\right\}$$

$$\lim_{n\to\infty} \beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = \Phi[a-\delta/\gamma(\theta_0)] \qquad \ldots \quad (4.11)$$

using the uniform convergence proved in (i) of Lemma 3. It appears from (4.11) that a test of the hypothesis $\theta = \theta_0$ based on T_n does not attain the full Pitman power $\Phi(a-\delta i^{\frac{1}{2}})$ unless $\gamma^2 = i^{-1}$. It is therefore interesting to know whether the condition (4.7) of Lemma 3 itself implies that $\gamma^2 = i^{-1}$. I have been able to establish this result only under the additional Assumption IV but it is worth examining whether such a strong assumption is necessary.

Under condition (i) of Assumption IV we have the expansion of the power function

$$\beta_n(\theta_0+\delta n^{-\frac{1}{2}}) = \beta_n(\theta_0)+\frac{\delta}{n^{\frac{1}{2}}}\beta_n'(\theta_0)+\frac{\delta^2}{2n}\beta_n''(\theta') \qquad \ldots \quad (4.12)$$

and under condition (ii) of Assumption IV, $\beta_n''(\theta')/n$ is bounded in an interval of θ' enclosing θ_0. From (4.12) we find

$$\lim_{\delta\to 0}\lim_{n\to\infty} \frac{\beta_n(\theta_0+\delta n^{-\frac{1}{2}})-\beta(\theta_0)}{\delta} = \lim_{n\to\infty} n^{-\frac{1}{2}}\beta_n'(\theta_0).$$

Hence
$$\lim_{\delta\to 0} \frac{\Phi(a-\delta/\gamma)-\Phi(a)}{\delta} = \lim_{n\to\infty} n^{-\frac{1}{2}}\beta_n'(\theta_0) \qquad \ldots \quad (4.13)$$

The limit of the R.H.S. of (4.13) is

$$(i/2\pi)^{\frac{1}{2}}e^{-a^2/2} \qquad \ldots \quad (4.14)$$

using the result of Theorem 1 in an earlier paper (Rao, 1962). The value of the L.H.S. of (4.13) is

$$(2\pi\gamma^2)^{-\frac{1}{2}}e^{-a^2/2}. \qquad (4.15)$$

Comparing (4.14) and (4.15) we find $\gamma^2 = i$ which establishes (ii) of Lemma 3.

Lemma 4 : *Let* $\{n^{\frac{1}{2}}(T_n-\theta), n^{\frac{1}{2}}Z_n(\theta)\}\to$ *in law to a bivariate normal distribution uniformly in compacts of* θ, *with the asymptotic covariance matrix*

$$\begin{pmatrix} \beta^2(\theta)/i(\theta) & \rho(\theta)\beta(\theta) \\ \rho(\theta)\,\beta(\theta) & i(\theta) \end{pmatrix}$$

Then under Assumptions I–III

$$\beta(\theta) = 1 \Longrightarrow \rho(\theta) = 1 \Longrightarrow n^{\frac{1}{2}}\,|\,Z_n(\theta)-i(T_n-\theta)\,| \overset{UL}{\to} 0.$$

The Lemma 4 implies that v-efficiency of UCAN is equivalent to uniform first order efficiency.

Consider the test

$$n^{\frac{1}{2}}[T_n-\theta_0+\lambda Z_n(\theta_0)] > c_n\sigma$$

where $\sigma^2 = 1/i(\theta)+\lambda^2 i(\theta_0)+2\lambda\rho(\theta_0)$ the asymptotic variance of the test statistic. Using an argument similar to that of Lemma 3, the Pitman power of the test is

$$\Phi(a-\delta(1+\lambda i)/\sigma).$$

By the result of Lemma 1,

$$\frac{\delta(1+\lambda i)}{\sigma} \leqslant \delta i^{\frac{1}{2}}, \text{ for any arbitrary } \lambda$$

or

$$(1+\lambda i)^2 \leqslant 1+2\lambda i\rho+\lambda^2 i^2$$

which implies $\rho = 1$ at $\theta = \theta_0$ (any chosen value). The asymptotic variance of $n^{\frac{1}{2}}[Z_n(\theta)-i(T_n-\theta)]$ is then zero, and since the convergence is assumed to be uniform the desired result follows.

The results of Lemmas 1—4 under the conditions assumed on the probability density of the observations can be summarised as follows.

(i) If T_n is UCAN, the asymptotic variance of T_n has Fisher's lower bound $1/ni$. This implies that the concept of v-efficiency is not void when the class of estimators is restricted to UCAN.

It may be noted that the existence of such a lower bound to the asymptotic variance was established by Kallianpur and Rao (1955) under some conditions on the estimator such as Fisher consistency (FC) and Frechét differentiability. Recently (Kallianpur, 1963) relaxed the restriction of Frechét differentiability to a weaker form due to Volterra. Some observations on lower bound to asymptotic variance of a CAN estimator have also been made by Bahadur (1960) from a different point of view.

(ii) Uniform first order efficiency of T_n implies that it is CUAN and v-efficient.

(iii) The converse of (ii) has been established under the additional assumption that the joint asymptotic distribution of T_n and Z_n is bivariate normal and the convergence is uniform in compacts of θ.

It may be interesting to examine other conditions under which the existence of a CUAN estimator T_n with v-efficiency implies uniform first order efficiency. Restrictions on the estimator such as those imposed by Kallianpur and Rao (1955) and Kallianpur (1963) may be sufficient.

The investigations of Section 4 show that v-efficiency is a valid and useful concept if only we restrict our consideration to estimators which are consistent and uniformly asymptotically normal in compact intervals of the unknown parameter.

5. Second order efficiency

The second order efficiency is defined in earlier papers by Rao (1961, 1962) as the minimum asymptotic variance of

$$n[Z_n-\beta(T_n-\theta)-\gamma(T_n-\theta)^2] \qquad \dots \quad (5.1)$$

when minimised with respect to γ. Under some conditions this minimum value is equivalent to the limiting value of the difference in the actual amounts of information contained in the sample and in the statistic. It was also shown (Rao, 1961) that for the m.l. estimate the asymptotic variance of (5.1) is the least, thus establishing its highest second order efficiency.

It may be seen that the concepts of first and second order efficiencies are not explicitly linked with any loss function. It is also not important which function of θ is under estimation. We could, for instance, define first order efficiency as

$$n^{\frac{1}{2}} | Z_n - \beta[f(T_n) - f(\theta)] | \to 0$$

in probability for any function f admitting a continuous first derivative. Similarly the second order efficiency could be defined as the minimum asymptotic variance of

$$n(Z_n - \beta[f(T_n) - f(\theta)] - \gamma[f(T_n) - f(\theta)]^2) \qquad \ldots \quad (5.2)$$

where f admits a continuous second derivative. The expression for the minimum asymptotic variance in either case (5.1) or (5.2) would be exactly the same. Similarly if T_n is altered as

$$T_n + \frac{g(T_n)}{n}$$

where g is a smooth function, the first and second order efficiencies remain the same although from the point of view of quadratic loss function there would be difference in terms of order $(1/n^2)$. So the first and second order efficiencies as defined refer to some intrinsic properties of an estimator (statistic) used as a substitute for the whole sample for purposes of inference on the unknown parameter.

In a discussion on my paper (Rao, 1962), Lindley thought that the superiority of the m.l. estimate is probably established through some specific loss function implicit in the definition of second order efficiency. It is, therefore, proposed to compare different estimators in a more direct way by assuming a quadratic loss function. Before doing this, the procedure has to be cleared of some unpleasantness arising out of some samples of relatively small frequency leading to large deviations in the estimator and making the expected loss unduly large. We shall, therefore, omit a portion of the sample space and compare the performance of estimators over the rest of the sample space. Usually the total probability of the portion so omitted rapidly diminishes to zero as the sample size increases and the value of the estimator over this portion could be defined arbitrarily except that it should be bounded.

We shall consider the case of the finite multinomial distribution as in the earlier paper (Rao, 1961). Let us represent the theoretical frequencies in the k cells by

$$\pi_1(\theta), \ldots, \pi_k(\theta)$$

where θ is an unknown parameter, the observed proportions by

$$p_1, \ldots, p_k$$

and the estimating equation by

$$f(\theta, p_1, \ldots, p_k) = 0 \qquad \ldots \quad (5.3)$$

where

$$f(\theta, \pi_1(\theta), \ldots, \pi_k(\theta)) \equiv \theta$$

so that the estimator satisfies Fisher consistency. We shall assume that f as a function of θ, $p_1, \ldots, p_k$ admits third order partial derivatives which are bounded in a closed region P of the cube C

$$0 \leqslant p_i \leqslant 1, \qquad i = 1, \ldots, k$$

and for values of θ satisfying (5.3) with $(p_1, \ldots, p_k) \epsilon P$. The true point $\pi_1(\theta), \ldots, \pi_k(\theta)$ is assumed to be an interior point of P. Let θ^* be a solution of the equation (5.3) such that $\theta^* \to \theta$ as $p_i \to \pi_i(\theta)$. Then expanding $f(\theta^*, p_1, \ldots, p_k)$ by Taylor's theorem at θ, $\pi_1(\theta), \ldots, \pi_k(\theta)$, we have

$$\frac{\delta f}{\delta \theta}(\theta^* - \theta) + \Sigma \frac{\delta f}{\delta \pi_r}(p_r - \pi_r)$$

$$= -\frac{1}{2} \Sigma \Sigma \frac{\delta^2 f}{\delta \pi_r \delta \pi_s}(p_r - \pi_r)(p_s - \pi_s)$$

$$-\frac{1}{2}(\theta^* - \theta)^2 \frac{\delta^2 f}{\delta \theta^2} - (\theta^* - \theta) \Sigma \frac{\delta^2 f}{\delta \theta \delta \pi_r}(p_r - \pi_r) + \varepsilon, \qquad \ldots \quad (5.4)$$

Due to the boundedness of the third order partial derivatives, if we define θ^* arbitrarily in $C - P$, except that it should be bounded, it follows that

$$E(\theta^* - \theta) = O(n^{-1}), \qquad E(\varepsilon^2) = O(n^{-3}).$$

If the equation (5.3) is such that first order efficiency is satisfied then

$$\frac{\delta f}{\delta \pi_r} \div \frac{\delta f}{\delta \theta} = -\frac{1}{i} \frac{\pi_r'}{\pi_r}$$

as shown in (Rao, 1961) in which case, dividing (5.4) by $\delta f / \delta \theta$, the left hand side expression can be written

$$\theta^* - \theta - Z_n(\theta)/i$$

where $Z_n = \Sigma[\pi_r'(p_r - \pi_r)/\pi_r]$. If the right hand side of (5.4) without ε, divided by $\partial f / \partial \theta$ and $(\theta^* - \theta)$ replaced by Z_n / i is represented by S_n, we have the approximate relation

$$\theta^* - \theta - Z_n / i \sim S_n. \qquad \ldots \quad (5.5)$$

Hence

$$E(\theta^* - \theta) \sim E(S_n) = b(\theta)/n$$

where $b(\theta)/n$ is the bias in the estimator up to terms of $O(1/n)$. Such a bias has no effect if the mean square error is evaluated up to terms of $O(1/n)$. Otherwise correction for bias seems to be necessary. The correction can easily be done by considering the estimator

$$\hat{\theta} = \theta^* - \frac{b(\theta^*)}{n}$$

in which the bias is $o(1/n)$. We shall evaluate $E(\hat{\theta} - \theta)^2$ upto terms of $O(1/n^2)$.

Consider the approximate relationship

$$E(\theta^*) \sim \theta + \frac{b(\theta)}{n}$$

which on differentiation with respect to θ yields

$$nE(\theta^* Z_n) \sim 1 + \frac{b'(\theta)}{n}. \qquad \ldots \quad (5.6)$$

Further

$$V(\hat{\theta}) \sim V(\theta^* - Z_n b'(\theta)/ni)$$
$$\sim V(\theta^*) - 2b'(\theta)/n^2 i$$

using (5.6) and

$$V(\theta^* - Z_n/i) = V(\theta^*) + V(Z_n/i) - 2\,\mathrm{cov}\,(\theta^*, Z_n/i)$$
$$= V(\theta^*) + \frac{1}{ni} - \frac{2}{ni}\left(1 + \frac{b'}{n}\right)$$
$$= V(\theta^*) - \frac{2b'}{n^2 i} - \frac{1}{ni}$$
$$= V(\hat{\theta}) - \frac{1}{ni}. \qquad \ldots \quad (5.7)$$

From (5.5)

$$V(\theta^* - Z_n/i) \sim V(S_n) = \frac{\psi(\theta)}{n^2} \text{ (say)}$$

Using (5.7) we have

$$V(\hat{\theta}) = \frac{1}{ni} + \frac{\psi(\theta)}{n^2} + o\left(\frac{1}{n^2}\right). \qquad \ldots \quad (5.8)$$

We shall compute $\psi(\theta)$ for some methods of estimation and compare the values. The variance of θ^*, without correction for bias, is

$$V(\theta^*) = \frac{1}{ni} + \frac{\psi(\theta)}{n^2} + \frac{2b'(\theta)}{n^2 i} + o\left(\frac{1}{n^2}\right) \qquad \ldots \quad (5.9)$$

(i) *Maximum likelihood.* For the method of maximum likelihood (m.l.)

$$S_n = \frac{Z_n(W_n - gZ_n)}{i^2} - \frac{\mu_{11}\, Z_n^2}{2i^3}$$

where

$$W_n = \Sigma(d^2 \log \pi_r/d\theta^2)(p_r - \pi_r), \; g = (\mu_{11} - \mu_{30})/i$$
$$\mu_{rs} = \Sigma \pi_i (\pi_i'/\pi_i)^r (\pi_i''/\pi_i)^s.$$

The bias in the estimator and the value of $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = E(S_n) = -\frac{\mu_{11}}{2ni^2},$$

$$\psi(\theta) = n^2 V(S_n) = \frac{V[Z_n(W_n - gZ_n)]}{i^4} + \frac{\mu_{11}^2}{4i^6}\, V(Z_n^2)$$

$$= \frac{\mu_{02} - 2\mu_{21} + \mu_{40}}{i^3} - \frac{1}{i} - \frac{(\mu_{11} - \mu_{30})^2}{i^4} + \frac{\mu_{11}^2}{2i^4}$$

$$= \psi(\text{m.l.}). \qquad \ldots \quad (5.10)$$

The variance of the m.l. estimator without correction for bias is

$$\frac{1}{ni}+\frac{\psi(\text{m.l.})}{n^2}-\frac{1}{i}\,\frac{d}{d\theta}\left(\frac{\mu_{11}}{ni^2}\right)+o\left(\frac{1}{n^2}\right)$$

which agrees with the expression given by Haldane and Smith (1956). It may be seen that ψ(m.l.) is connected with E_2(m.l.), the index of second order efficiency defined in the earlier paper (Rao, 1961) by the relation

$$i^2\psi(\text{m.l.}) = E_2(\text{m.l.})+\frac{\mu_{11}^2}{2i^2}\ .$$

It may be seen that the m.l. estimator corrected for bias is similar to the estimator given by Lindley (1961). For other properties of m.l. estimators reference may be made to papers by Cramér (1946), Daniels (1961), Doob (1934, 1936), LeCam (1953, 1956), Rao (1957, 1958, 1960a), Wald (1949) and others. Uniform consistency and convergence to normality of m.l. estimators are considered by Kraft (1955) and Parzen (1954).

(ii) *Minimum chi-square.* A theoretical investigation of the asymptotic properties of minimum chi-square estimates is contained in papers by Neyman (1949) and Rao (1955). The estimating equation is

$$\Sigma\, \pi_r' \frac{p_r^2}{\pi_r^2} = 0$$

and the value of S_n is

$$\left(Q+\frac{\mu_{30}}{2i^3}\,Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}\,Z_n^2$$

where $$Q = \frac{1}{2i}\,\Sigma\,\frac{\pi_r'^2}{\pi_r}(p_r-\pi_r)^2-\frac{1}{i^2}\,Z_n\,\Sigma\left(\frac{\pi_r'}{\pi_r}\right)^2(p_r-\pi_r).$$

By using the expressions already derived in Rao (1961), the bias in the minimum chi-square estimate and the value of $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{\frac{1}{2i}\,\Sigma\,\frac{\pi_r'}{\pi_r}-\frac{\mu_{30}+\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = n^2V(S_n) = \delta+\psi(\text{m.l.})$$

where $$\delta = \frac{1}{2i^2}\,\Sigma\left(\frac{\pi_r'}{\pi_r}\right)^2-\frac{\mu_{40}}{i^3}+\frac{\mu_{30}^2}{2i^4} \qquad \dots \quad (5.11)$$

which is non-negative and zero only in special cases.

(iii) *Minimum modified chi-square (Neyman,* 1949). The estimating equation is

$$\Sigma\,\frac{\pi_r\pi_r'}{p_r^2} = 0$$

leading to the value of S_n

$$-2\left(Q+\frac{\mu_{30}}{2i^3}\,Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}\,Z_n^2.$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{-\frac{1}{i}\Sigma\frac{\pi_r}{\pi_r}+\frac{2\mu_{30}-\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = 4\delta+\psi(\text{m.l.}).$$

(iv) *Haldane's minimum discrepancy* (*Haldane*, 1953). The estimating equation, after a slight modification which does not effect the treatment of the present paper, is

$$\Sigma\frac{\pi_r^k\,\pi_r'}{p_r^k} = 0$$

giving the value of S_n

$$-(k+1)\left(Q+\frac{\mu_{30}}{2i^3}\;Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}\;Z_n^2,$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{-\frac{(k+1)}{2i}\Sigma\frac{\pi_r'}{\pi_r} \text{[illegible]} -\mu_{11} \text{[illegible]}\right\}$$

$$\psi(\theta) = (k+1)^2\delta+\psi(\text{m.l.})$$

(v) *Minimum Hellinger distance.* The estimating equation is

$$\Sigma\frac{\pi_r'\,p_r^{\frac{1}{2}}}{\pi_r^{\frac{1}{2}}} = 0$$

giving the value of S_n

$$-\frac{1}{2}\left(Q+\frac{\mu_{30}}{2i^3}\;Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}\;Z_n^2.$$

The bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{-\frac{1}{2i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{\mu_{30}-2\mu_{11}}{4i^2}\right\}$$

$$\psi(\theta) = \frac{\delta}{4}+\psi(\text{m.l.}).$$

(vi) *Minimum Kullback-Liebler separator.* The estimating equation is

$$\Sigma\,\pi_r'\log\frac{\pi_r}{m} = 0$$

giving the value of S_n

$$-\left(Q+\frac{\mu_{30}}{2i^3}\,Z_n^2\right)+\frac{Z_n(W_n-gZ_n)}{i^2}-\frac{\mu_{11}}{2i^3}\,Z_n^2.$$

The values of bias and $\psi(\theta)$ are

$$\frac{b(\theta)}{n} = \frac{1}{n}\left\{-\frac{1}{2i}\Sigma\frac{\pi_r'}{\pi_r}+\frac{\mu_{30}-\mu_{11}}{2i^2}\right\}$$

$$\psi(\theta) = \delta+\psi(\text{m.l.})$$

It is seen that among the six methods compared, the mean square error in the estimator corrected for bias is the least in the case of the m.l., when terms up to the order $(1/n^2)$ are considered. It may be shown more generally (following the mechanism developed in the earlier paper, Rao 1961) that under the assumptions made on the estimating equation $f(\theta, p) = 0$, the m.l. estimator has the least value for $\psi(\theta)$.

The bias and variance for estimators corrected for bias, obtained by the different methods considered in this section are given below, where δ and ψ(m.l.) are as defined in (5.10) and (5.11).

method of estimation	bias (coefficient of n^{-1})	variance of estimator corrected for bias: coefficient of n^{-1}	coefficient of n^{-2}
maximum likelihood	$-\frac{\mu_{11}}{2i^2}$	$\frac{1}{i}$	ψ(m.l.)
minimum chi-square	$\frac{1}{2i}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)-\frac{\mu_{30}+\mu_{11}}{2i^2}$	$\frac{1}{i}$	$\delta+\psi$(m.l.)
modified minimum chi-square	$-\frac{1}{i}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)+\frac{2\mu_{30}-\mu_{11}}{2i^2}$	$\frac{1}{i}$	$4\delta+\psi$(m.l.)
Haldane's minimum discrepancy	$\frac{-(k+1)}{2i}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)+\frac{(k+1)\mu_{30}-\mu_{11}}{2i^2}$	$\frac{1}{i}$	$(k+1)^2\delta+\psi$(m.l.)
minimum Hellinger distance	$-\frac{1}{2i}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)+\frac{\mu_{30}-2\mu_{11}}{4i^2}$	$\frac{1}{i}$	$\frac{\delta}{4}+\psi$(m.l.)
minimum K.-L. separate	$-\frac{1}{2i}\Sigma\left(\frac{\pi_r'}{\pi_r}\right)+\frac{\mu_{30}-\mu_{11}}{2i^2}$	$\frac{1}{i}$	$\delta+\psi$(m.l.)

The expressions for bias and variance will be similar in the case of estimation of parameters in a continuous distribution. The conditions to be assumed on the estimating equation and the probability density will be very severe if an expansion of the asymptotic variance up to terms of order $(1/n^2)$ is desired. A recent paper by Linnik and Mitrafanova (1963) on the computation of the variance of the m.l. estimator in a continuous case shows the nature of the complexities involved.

REFERENCES

BAHADUR, R. R. (1960): On asymptotic efficiency of tests and estimates. *Sankhyā*, **22**, 229–252.

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

DANIELS, H. E. (1961): The asymptotic efficiency of a maximum likelihood estimator. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 151–164.

DOOB, J. L. (1934): Probability and Statistics. *Trans. Amer. Math. Soc.*, **36**, 759–772.

——— (1936): Statistical estimation. *Trans. Amer. Math. Soc.*, **39**, 410–421.

FISHER, R. A. (1922): On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy Soc.*, A, **222**, 309–365.

——— (1925): Theory of statistical estimation. *Proc. Camb. Phil. Soc.*, **22**, 700–725.

HALDANE, J. B. S. and SMITH, SHEILA MAYNARD (1956): The sampling distribution of a maximum likelihood estimate. *Biom.*, **43**, 96–103.

HALDANE, J. B. S. (1953): A class of efficient estimates of a parameter. *Bull. Int. Stat. Inst.*, **33**, 231.

KALLIANPUR, G. (1963): Von Mises functionals and maximum likelihood estimation. *Contributions to Statistics*, presented to Professor P. C. Mahalanobis on his 70th birthday.

KALLIANPUR, G. and RAO, C. R. (1955): On Fisher's lowerbound to asymptotic variance of a consistent estimate. *Sankhyā*, **15**, 331–342.

KRAFT, CHARLES H. (1955): Some conditions for consistency and uniform consistency of statistical procedures. *University of California Publications in Statistics*, **2**, 125–42.

LECAM, L. (1953): On some asymptotic properties of maximum likelihood estimates and related Baye's estimates. *University of California Publications in Statistics*, **1**, 227–330.

——— (1956): On the asymptotic theory of estimation and testing hypotheses. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, **1**, 129–156.

——— (1960): Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, **3**, 37–98.

LINDLEY, D. V. (1961): The use of prior probability distributions in statistical inference and decisions. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 453-468.

LINNIK, YU. V. and MITRAFANOVA, N. M. (1963): Some asymptotic expansions for the distribution of the maximum likelihood estimate. *Contributions to Statistics*, presented to Professor P. C. Mahalanobis on his 70th birthday.

PARZEN, E. (1954): On uniform convergence of families of sequences of random variables. *University of California Publications in Statistics*, **2**, 23–54.

RAO, C. R. (1955): Theory of the method of estimation by minimum chi-square. *Bull. Int. Statist. Inst.*, **35**, 25–32.

——— (1957): Maximum likelihood estimation for multinomial distribution. *Sankhyā*, **18**, 139–148.

——— (1958): Maximum likelihood estimation for the multinomial distribution with an infinite number of cells. *Sankhyā*, **20**, 211–218.

——— (1960a): A study of large sample test criteria through properties of efficient estimates. *Sankhyā*, **23**, 25–40.

——— (1960b): Apparent anomalies and irregularities in maximum likelihood estimation. *32nd Session of the Int. Stat. Inst.*, Tokyo. Reprinted with discussion in *Sankhyā*, **24**, 73–102.

——— (1961): Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 531–546.

——— (1962): Efficient estimates and optimum inference procedures in large samples, (with discussion). *J. Roy. Stat. Soc.*, **24**, 1, 46–72.

Wald, A. (1949): A note on the consistency of maximum likelihood estimation. *Ann. Math. Stat.* **20**, 595–601.