ASYMPTOTIC EFFICIENCY AND LIMITING INFORMATION

C. RADHAKRISHNA RAO
INDIAN STATISTICAL INSTITUTE, CALCUTTA

1. Introduction

In a recent paper [15], the author gave new formulations of the concepts of asymptotic efficiency and consistency, which seem to throw some light on the principle of maximum likelihood (m.l.) in estimation. An attempt was also made in that paper to link up the concept of asymptotic efficiency with the limiting information per observation contained in a statistic, as the sample size tends to infinity, information on an unknown parameter θ being defined in the sense of Fisher [5], [6].

The object of the present paper is to pursue the investigation of the earlier paper and establish some further propositions which might be of use in understanding the m.l. method of estimation.

The first proposition is concerned with the conditions under which i_T, the information per observation in a statistic T_n tends to i, the information in a single observation, as the sample size $n \to \infty$. It may be noted that i_T cannot exceed i for any n. A sufficient condition for convergence of i_T to i is

$$\left| n^{-1/2}\left(d\,\frac{\log L}{d\theta}\right) - \alpha - \beta n^{1/2}(T_n - \theta)\right| \to 0 \tag{1.1}$$

in probability, where L denotes the likelihood of θ given the sample and α, β are constants possibly depending on θ. This proposition was first proved by Doob [3], but his proof appears to be somewhat complicated. Under the same conditions assumed by Doob, the problem has been formulated in a more general way and a simple proof has been provided. It is observed that asymptotic efficiency of an estimator T_n may be defined as the property (1.1), or a less restrictive condition such as the asymptotic correlation between $n^{-1/2}(d\log L/d\theta)$ and $n^{1/2}(T_n - \theta)$ being unity, which imply that $i_T \to i$. This new definition of asymptotic efficiency is applicable to a wider class of statistics, while for the application of the usual definition in terms of the asymptotic variance of the estimator some regularity conditions on the estimator have to be imposed.

It is known that under some regularity conditions the m.l. estimate or an estimate obtained as a particular (consistent) root of the equation $(d\log L/d\theta) = 0$, has the property (1.1). But the m.l. estimate is only one member of a large class of estimates satisfying this property. The minimum chi-square, modified minimum chi-square (Neyman [11]), minimum discrepancy (Haldane [8]), and a few other methods are known to provide estimates efficient in this sense, so that judged by the criterion of asymptotic efficiency alone the m.l. estimate has nothing special to distinguish it from the others.

As emphasized in the writings of Fisher [5], [6] and in my recent paper (Rao [15]) the efficiency of a statistic has to be judged by the degree to which the estimate provides an approximation to $d \log L/d\theta$. The property (1.1) states that under a suitable norming factor $(T_n - \theta)$ is close to $n^{-1/2}(d \log L/d\theta)$ in large samples. We may now investigate the rate at which the convergence in (1.1) takes place. For this we consider $n^{1/2}$ times the random variable in (1.1)

(1.2)

$$n^{1/2}\left[n^{-1/2}\frac{d \log L}{d\theta} - \alpha - \beta n^{1/2}(T_n - \theta)\right] = \frac{d \log L}{d\theta} - n^{1/2}\alpha - n\beta(T_n - \theta),$$

which may not tend to zero and may have a limiting distribution. The variance of this limiting distribution is considered as a measure of convergence. Since our object is to determine the extent to which $d \log L/d\theta$ can be approximated by a function of T_n we may as well consider the random variable

$$\frac{d \log L}{d\theta} - n^{1/2}\alpha - n\beta(T_n - \theta) - \lambda n(T_n - \theta)^2, \tag{1.3}$$

which provides a second degree polynomial approximation in $(T_n - \theta)$ to $d \log L/d\theta$ and define E_2, the minimum asymptotic variance of (1.3) when minimized with respect to λ, as a measure of *second order efficiency* of an estimator. The smaller the value of E_2, the greater is the second order efficiency.

The advantage of such a definition is that when the minimum asymptotic variance of (1.3) is the limit of the average conditional variance of $d \log L/d\theta$ given T_n (a proposition which is not proved in the present paper), the measure E_2 we consider is the limit of $(ni - ni_T)$ as $n \to \infty$. This is a very satisfactory situation as we can then study the behavior of the difference between the *actual* amounts of information contained in the statistic and in the sample.

When the stated convergence of the average conditional variance takes place the random variable (1.3) is equivalent, in large samples, to

$$\frac{d \log L}{d\theta} - \frac{d \log \phi}{d\theta}, \tag{1.4}$$

where ϕ is the density of the estimate T_n. Of course, the best fit to $d \log L/d\theta$ is $d \log \phi/d\theta$ in the sense of minimum variance. The variance of (1.4) is exactly $(ni - ni_T)$ and its limiting value is of interest. The random variable (1.3) is considered in preference to (1.4) as the asymptotic properties of the former are easy to determine.

The second proposition established here is that under some regularity conditions, in the special case of multinomial populations, the m.l. estimate has the maximum second order efficiency. The actual expressions for second order efficiency have also been computed for a number of alternative methods of estimation applicable to multinomial populations. The extension of this investigation to populations with continuous distributions needs a more elaborate analysis involving the consideration of functionals of the empirical distribution function and their differentiability, some related ideas of which are contained in a paper by Kallianpur and Rao [9]. The result that the m.l. estimate has the maximum second order efficiency was stated by Fisher [6]. The present paper only supplies a rigorous demonstration of this result under some assumptions. Further, Neyman [12] observed that there exist a large class of estimation procedures leading to best asymptotically normal (BAN) estimates, of which the m.l. estimate is only one member "but the question remains open concerning how good these estimates are when the number of observations is only moderate." The concept of second order efficiency provides an answer to this problem as it enables comparison of different BAN estimates.

2. Definitions, notations, and assumptions

Let $p(x, \theta)$ denote the probability density of observation x, depending on the parameter θ, and $P_n = p(x_1, \theta) \cdots p(x_n, \theta)$ the density for n independent observations $x_1, \cdots, x_n$. Let T_n be a Lebesgue measurable function of the observations and $\phi(T_n, \theta) = \phi_n$ its probability density.

ASSUMPTION 1. *There is a domain D, such that for each value of θ, $P_n > 0$ on D except for a set of points of Lebesgue measure zero and $P_n = 0$ on the complement of D except for a set of points of Lebesgue measure zero.*

It is easy to see that if this property (D) is true for P_n, it is also satisfied for ϕ_n (Doob [4]).

ASSUMPTION 2. *The derivative $dP_n/d\theta$ exists and i, defined by the following equation, is finite,*

$$ni = E\left(\frac{1}{P_n}\frac{dP_n}{d\theta}\right)^2 = nE\left(\frac{1}{p}\frac{dp}{d\theta}\right)^2. \tag{2.1}$$

It may be noted that as a consequence of assumption 2 we have the following. If $Z_n = n^{-1/2}(d \log P_n/d\theta)$, then $E(Z_n) = 0$, $V(Z_n) = i$, and the asymptotic distribution of Z_n is normal with mean zero and variance i.

ASSUMPTION 3. *For E_n any Lebesgue measurable set in Euclidean space R_n*

$$\frac{d}{d\theta}\int_{E_n}\cdots\int P_n\, dv = \int_{E_n}\cdots\int \frac{dP_n}{d\theta}\, dv, \tag{2.2}$$

$$\frac{d}{d\theta}\int_{E_1} \phi_n\, dT_n = \int_{E_1} \frac{d\phi_n}{d\theta}\, dT_n, \tag{2.3}$$

where $dv = dx_1 \cdots dx_n$ and the derivative $(d\phi_n/d\theta)$ is also assumed to exist.

The information contained in T_n is, by definition, $V(d \log \phi_n/d\theta) = ni_T$. We are concerned with the limiting value of i_T. If we define $Y_n = n^{-1/2}(d \log \phi_n/d\theta)$, then $E(Y_n) = 0$ and $V(Y_n) = i_T$.

ASSUMPTION 4. *The joint asymptotic distribution of Z_n and $U_n = n^{1/2}(T_n - \theta)$ exists and has finite second order moments.*

3. The limiting value of i_T

We briefly recall the essential notations

$$Z_n = n^{-1/2} \frac{d \log P_n}{d\theta} = n^{-1/2} \sum_1^n \frac{d \log p(x_i, \theta)}{d\theta}, \tag{3.1}$$

$$U_n = n^{1/2}(T_n - \theta), \tag{3.2}$$

$$Y_n = n^{-1/2} \frac{d \log \phi_n(T_n, \theta)}{d\theta}, \tag{3.3}$$

where T_n is a Lebesgue measurable function of $x_1, \cdots, x_n$.

LEMMA 1. *Under the assumptions* 1 *to* 4 *of section* 2 *we have*

(i) $E(Z_n|T_n) = Y_n$,

(ii) $E(Y_n)^2 = i_T \leqq i$,

(iii) $E(Z_nY_n) = i_T$.

To prove (i) we use the definitions of conditional expectation and probability density,

$$\int_{A'} Z_nP_n \, dv = \int_A E(Z_n|T_n)\phi_n \, dT_n, \tag{3.4}$$

$$\int_{A'} P_n \, dv = \int_A \phi_n \, dT_n, \tag{3.5}$$

where A is any Lebesgue measurable set on the real axis and A' is the corresponding set in E_n. By assumption 3,

$$\begin{aligned} n^{1/2} \int_{A'} Z_nP_n \, dv &= \frac{d}{d\theta} \int_{A'} P_n \, dv = \frac{d}{d\theta} \int_A \phi_n \, dT_n \\ &= \int_A \frac{d\phi_n}{d\theta} \, dT_n = n^{1/2} \int_A Y_n\phi_n \, dT_n. \end{aligned} \tag{3.6}$$

Comparing the last terms in (3.4) and (3.6) we have

$$E(Z_n|T_n) = Y_n, \tag{3.7}$$

excluding the set of values of T_n for which the density $\phi_n = 0$. To prove (ii) we consider the inequality

$$E(Z_n^2|T_n) \geqq [E(Z_n|T_n)]^2 \tag{3.8}$$

with probability 1, which follows from the convexity of the function Z_n^2. Taking further expectation with respect to the density of T_n we have

$$E(Z_n^2) \geqq E[E(Z_n|T_n)]^2 = E(Y_n^2) = i_T. \tag{3.9}$$

Therefore $i \geqq i_T$. The result (iii) is a consequence of (i). These results are not new and they are contained in papers by Fisher [6] and Doob [4].

LEMMA 2. *If $U_n^c = n^{1/2}(T_n - \theta)$ when $|n^{1/2}(T_n - \theta)| \leqq c$ and zero otherwise, where c is a continuity point of the asymptotic distribution of U_n, we have*

(i) $E(Z_n U_n^c) = EU_n^c[E(Z_n|T_n)] = E(U_n^c Y_n) = \gamma_n$.

(ii) *As $n \to \infty$, $\lim \gamma_n = \gamma^c$ for fixed c, where γ^c is the corrected product moment of the asymptotic distribution of U_n^c and Z_n. It may be noted that $\gamma^c \to \gamma$ as $c \to \infty$, where γ is the corresponding product moment of the asymptotic distribution of U_n and Z_n.*

(iii) *If $v_n = V(U_n^c)$, then $v_n \to v^c$ as $n \to \infty$, and $v^c \to v$ as $c \to \infty$, where v^c and v are second central moments of the asymptotic distributions of U_n^c and U_n respectively. All these results follow from the finiteness of all moments of U_n^c and the boundedness of $E(Z_n^2)$.*

THEOREM 1. $\liminf_{n\to\infty} i_T \geqq \rho^2 i$, *where ρ is the asymptotic correlation of U_n and Z_n.*

From [i], by applying Schwarz's inequality,

$$V(U_n^c)V(Y_n) \geqq [E(U_n^c Y_n)]^2 = \gamma_n^2 \tag{3.10}$$

or

$$i_T \geqq \frac{\gamma_n^2}{v_n}. \tag{3.11}$$

Hence

$$\liminf_{n\to\infty} i_T \geqq \frac{(\gamma^c)^2}{v_c}. \tag{3.12}$$

Since c is arbitrary we have, by letting $c \to \infty$,

$$\liminf_{n\to\infty} i_T \geqq \frac{\gamma^2}{v} = \frac{\rho^2 v\, i}{v} = \rho^2 i. \tag{3.13}$$

THEOREM 2. *If T_n is such that*

$$|n^{1/2}(T_n - \theta) - \alpha - \beta Z_n| \to 0 \tag{3.14}$$

in probability, then as $n \to \infty$, $\lim i_T = i$.

Condition (3.14) implies that the asymptotic correlation between Z_n and $n^{1/2}(T_n - \theta)$ is unity. From theorem 1, putting $\rho = 1$, we have

$$\liminf_{n\to\infty} i_T \geqq i. \tag{3.15}$$

But from (ii), $i_T \leqq i$ and therefore

$$\limsup_{n\to\infty} i_T \leqq i. \tag{3.16}$$

Hence $\lim i_T$ exists and is equal to i.

COROLLARY. *If $n^{1/2}(T_n - \theta)$ has correlation unity with Z_n, then $Y_n = n^{-1/2}(d \log \phi_n/d\theta)$ has asymptotic normal distribution with mean zero and variance i, which is the same as the asymptotic distribution of Z_n. The statistic $U_n = n^{1/2}(T_n - \theta)$ has also an asymptotic normal distribution.*

By theorem 2, limit $i_T = i$ when the asymptotic correlation between U_n and Z_n is unity. Hence $V(Y_n - Z_n) \to 0$ which implies that $|Y_n - Z_n| \to 0$ in probability. This shows that Y_n has the same asymptotic distribution as Z_n. Since U_n has asymptotic correlation unity with Z_n, the asymptotic distribution of U_n is also normal.

It is known (Doob [3], [4], Cramér [2], Rao [13], [14]) that the maximum likelihood (m.l.) estimate or more generally the consistent root of the m.l. equation, θ_n^*, satisfies the condition

$$\left| in^{1/2}(\theta_n^* - \theta) - n^{-1/2} \frac{d \log L}{d\theta} \right| \to 0 \tag{3.17}$$

in probability under some regularity conditions on the probability density. For such an estimate the limiting information per observation attains the maximal value i.

4. Asymptotic efficiency

The motivation for a new formulation of the concept of efficiency is fully explained in an earlier paper (Rao [15]). It is observed that Z_n as a function of the observations and the parameter plays an important role in statistical inference. For instance, it is known that the best test of the hypothesis $H_0{:}\theta = \theta_0$ (an assigned value), which has maximum local power in the direction $\theta > \theta_0$ for any n is of the form, "Reject H_0 if $Z_n(\theta_0) \geqq \lambda$" (see Rao [15]). Or in other words, $Z_n(\theta_0)$ affords the best discrimination between values of the parameter θ_0 and an alternative close to θ_0 *for any given n*. Wald [16], [17] has shown that Z_n can be used as a "pivotal quantity" in the sense of Fisher [7] to obtain a confidence interval for θ which is asymptotically the best.

If we replace the sample by a statistic T_n, then for purposes of statistical inference based on T_n the pivotal quantity is $d \log \phi/d\theta = n^{1/2}Y_n$. The pivotal quantity based on the whole sample is $d \log P_n/d\theta = n^{1/2}Z_n$ and there is no loss of information if $Y_n = Z_n$ for all θ and all samples, a situation which arises when T_n is a sufficient statistic. If no such (Lebesgue measurable) function T_n exists, we may find one, for which Y_n and Z_n are as close as possible in some sense, so that inferences based on Y_n and Z_n may not be widely discrepant.

As a first step we may demand that

$$|Y_n - Z_n| \to 0 \tag{4.1}$$

in probability so that in sufficiently large samples the equivalence of Y_n and Z_n is assured. We may then state asymptotic efficiency in one of the following ways, not all of which are equivalent.

DEFINITION. *A statistic T_n is said to be asymptotically efficient if*

(i) $|Y_n - Z_n| \to 0$ *in probability, or*

(ii) $i_T \to i$ *as* $n \to \infty$, *or*

(iii) *the asymptotic correlation between* $n^{1/2}(T_n - \theta)$ *and* Z_n *is unity, or*

(iv) $|Z_n - \alpha - \beta n^{1/2}(T_n - \theta)| \to 0$ *in probability, where* α *and* β *are constants possibly depending on* θ.

Condition (iv) is perhaps the most stringent; (iii) follows (iv); (ii) follows from (iii) by theorem 2; (i) follows from (ii) by Chebecheff's lemma. In the earlier paper of the author (Rao [15]), only conditions (iii) and (iv) were mentioned.

Although condition (i) provides a logical definition of asymptotic efficiency, it is difficult to use it in practice and its verification depends on the computation of the distribution of the estimate. Condition (iv) is, perhaps, the easiest to verify. Further, (iv) implies (i) and therefore we shall use (iv) for further investigations.

The new definition of efficiency implies minimum asymptotic variance for the statistic T_n only under some regularity conditions (Neyman [12], Barankin and Gurland [1], Kallianpur and Rao [9]). But when these conditions are not satisfied the definition of efficiency as minimum asymptotic variance is not applicable as there is no nonzero lower bound to the minimum asymptotic variance of a consistent estimate, as shown by Hodges (Le Cam [10]). Given a consistent estimate, Hodges gives a method by which another consistent estimate can be constructed with an asymptotic variance not greater than that of the former and in fact less for some values of the parameter. With the new definition of asymptotic efficiency no such anomaly arises.

Further, the choice of a statistic on the criterion of a smaller asymptotic variance can be misleading. For instance, using the method of Hodges and Le Cam (Le Cam [10], p. 287) one can construct the following example. Let $\bar{x}$ and x_m be the mean and median in samples of size n from a normal population with an unknown mean θ and variance unity. Define the statistic

$$T_n = \begin{cases} \bar{x} & \text{if } |\bar{x}| \geqq n^{-1/4}, \\ \alpha x_m & \text{if } |\bar{x}| < n^{-1/4}. \end{cases} \tag{4.2}$$

It is easy to see that the asymptotic variance of $n^{1/2}(T_n - \theta) = 1$ if $\theta \neq 0$ and $\alpha^2\pi/2$ if $\theta = 0$; the latter can be made to approach zero by choosing α arbitrarily small. The statistic T_n has thus smaller asymptotic variance for at least one value of θ and is as good as the alternative estimate $\bar{x}$ for the other values of θ. Should we then prefer T_n to $\bar{x}$? It is clear that we are using essentially the median when θ has the value zero and the mean otherwise, and consequently the use of T_n for statistical inference entails some loss of information. This example can be generalized by the method of Le Cam [10] to obtain a statistic which is equivalent to the median for a denumerable closed subset S of values of θ and to the mean elsewhere, but with a smaller asymptotic variance on S. Clearly that would be making the situation worse. Under the new definition T_n as defined in (4.2) is clearly inefficient.

5. Second order efficiency

5.1 *The concept.* There may be several methods of estimation all leading to asymptotically efficient estimates in the sense

$$\left| n^{-1/2}\frac{d\log L}{d\theta} - \alpha - \beta n^{1/2}(T_n - \theta)\right| \to 0 \tag{5.1.1}$$

in probability. For instance, under some regularity conditions, m.l., minimum chi-square, and related methods provide estimates satisfying condition (5.1.1), so that no distinction among these methods is possible on the basis of asymptotic efficiency alone. For this purpose we may examine the rate of convergence in (5.1.1) for each method of estimation. One measure of the rate of convergence is the asymptotic variance of the random variable

$$\begin{aligned} n^{1/2}\left[n^{-1/2}\frac{d\log L}{d\theta} - \alpha - n^{1/2}\beta(T_n - \theta)\right] \\ = \frac{d\log L}{d\theta} - \alpha n^{1/2} - \beta n(T_n - \theta), \end{aligned} \tag{5.1.2}$$

which is $\sqrt{n}$ times the random variable in (5.1.1). Instead of (5.1.2) we may consider the random variable

$$\frac{d\log L}{d\theta} - \alpha n^{1/2} - \beta n(T_n - \theta) - \lambda n(T_n - \theta)^2, \tag{5.1.3}$$

which is the difference between $d\log L/d\theta$ and a second degree polynomial approximation based on the estimate T_n. The constant λ may be chosen so as to minimize the asymptotic variance of (5.1.3) and this minimum value, denoted by E_2 (smaller values being preferable), may be used to indicate the second order efficiency of the estimate. It may be noted that the best fit to $d\log L/d\theta$ in terms of the estimate T_n is $d\log\phi_n/d\theta$, in the sense of least variance of the difference, and indeed the ideal measure of second order efficiency would be

$$\lim_{n\to\infty} V\left(\frac{d\log L}{d\theta} - \frac{d\log\phi_n}{d\theta}\right) = \lim_{n\to\infty}(ni - ni_T), \tag{5.1.4}$$

which is the limiting difference in the actual amounts of information contained in the sample and in the statistic, while the first order efficiency is concerned only with the limit of $(i - i_T)$. The second degree polynomial in T_n is considered instead of $d\log\phi_n/d\theta$ as it is more convenient to handle. Further, if the average conditional variance of (5.1.2) given T_n tends to the corresponding value for the joint asymptotic distribution of (5.1.2) and $n^{1/2}(T_n - \theta)$, a result which is not proved in the present paper but is likely to be true under regularity conditions assumed, then the minimum asymptotic variance of (5.1.3) would be same as the limit of $(ni - ni_T)$, that is,

$$\lim_{n\to\infty}(ni - ni_T) = E_2. \tag{5.1.5}$$

We may say that in large samples the amount of information lost in replacing the sample by the statistic T_n is equivalent to that of (E_2/i) independent observations, as stated by Fisher [6]. In the next two sections the second order efficiencies are computed for a number of estimation procedures and their magnitudes are compared.

5.2 *Maximum attainable second order efficiency.* In this section we restrict our investigation to multinomial populations with a finite number k of cells. The true proportions are represented by $\pi_1(\theta), \cdots, \pi_k(\theta)$, where θ is an unknown parameter, and the observed proportions by $p_1, \cdots, p_k$. We exhibit an estimate as a suitably chosen root of an equation

$$f(\theta, p) = f(\theta, p_1, \cdots, p_k) = 0, \tag{5.2.1}$$

as in the case of a number of known methods of estimation. The following assumptions are made.

ASSUMPTION 5. *$\pi_i(\theta), i = 1, \cdots, k$ admit derivatives up to the second order, which are continuous in the neighbourhood of the true value of θ.*

ASSUMPTION 6. *The estimating equation is consistent, that is,*

$$f[\theta, \pi(\theta)] = f[\theta, \pi_1(\theta), \cdots, \pi_k(\theta)] \equiv 0, \tag{5.2.2}$$

and $f(\theta, p_1, \cdots, p_k)$ has continuous derivatives up to the second order in θ as well as in $p_1, \cdots, p_k$ considered as variables. This implies that a Fisher consistent estimate (Kallianpur and Rao [9]) *can be obtained by solving equation* (5.2.1).

Let us denote by f', f_r, and f_{rs} the derivatives

$$\frac{\partial f}{\partial \theta}, \quad \frac{\partial f}{\partial p_r}, \quad \text{and} \quad \frac{\partial^2 f}{\partial p_r \partial p_s} \tag{5.2.3}$$

at the value θ and $p_r = \pi_r(\theta)$. The first and second derivatives of $\pi_r(\theta)$ with respect to θ are denoted by π_r' and π_r''. The same notations f', f_r, and f_{rs} are used for values of the derivatives at the true value θ_0 of θ in some of the expansions such as (5.2.5) and so on.

LEMMA 3. *Under assumptions* 5 *and* 6 *there exists a root θ^* of the equation $f(\theta, p) = 0$ such that $\theta^* \to \theta_0$, the true value, with probability* 1, *and for asymptotic efficiency* (*first order*) *to hold uniformly for all θ, it is necessary and sufficient that*

$$\frac{f_r}{f'} = -\frac{1}{i}\frac{\pi_r'}{\pi_r}, \qquad r = 1, \cdots, k. \tag{5.2.4}$$

Expanding $f(\theta, p)$ at $\theta_0, \pi_1(\theta_0), \cdots, \pi_k(\theta_0)$, we find

$$f(\theta, p) = f(\theta_0, \pi) + (\theta - \theta_0)(f' + \epsilon) + \sum (f_r + \epsilon_r)(p_r - \pi_r). \tag{5.2.5}$$

Since $p_r \to \pi_r$ with probability 1, $f' + \epsilon$ does not change sign and $(f_r + \epsilon_r)$ is bounded when p_r is close to π_r and θ to θ_0, it follows that $f(\theta, p)$ changes sign as θ passes through θ_0 and hence there exists θ^* such that $f(\theta^*, p) = 0$ with probability 1. It also follows that we can choose a root such that $\theta^* \to \theta_0$ with probability 1.

For such a root we find, from equation (5.2.5),

$$|n^{1/2}(\theta^* - \theta_0)f' + n^{1/2} \sum f_r(p_r - \pi_r)| \to 0 \tag{5.2.6}$$

with probability 1. If θ^* satisfies the criterion of asymptotic efficiency, then

$$\left|n^{-1/2}\frac{d\log L}{d\theta} - \alpha - \beta n^{1/2}(\theta^* - \theta)\right| \to 0 \tag{5.2.7}$$

in probability, or

$$\left|n^{1/2}\sum\left(\frac{\pi_r'}{\pi_r} + \frac{\beta f_r}{f'}\right)(p_r - \pi_r) - \alpha\right| \to 0 \tag{5.2.8}$$

in probability, the necessary and sufficient condition for which is

$$\alpha = 0 \quad\text{and}\quad \frac{\pi_r'}{\pi_r} + \frac{\beta f_r}{f'} = 0, \qquad r = 1, \cdots, k. \tag{5.2.9}$$

Since $f(\theta, \pi) \equiv 0$, we find on differentiation and substitution of the true value of θ,

$$f' = -\sum f_r \pi_r', \tag{5.2.10}$$

which in conjunction with (5.2.9) gives the value of β as

$$\beta = \sum \frac{\pi_r'^2}{\pi_r} = i \text{ (information)}. \tag{5.2.11}$$

Hence the result of lemma 3.

The next step is to compute the difference

$$\frac{d\log L}{d\theta} - ni(\theta^* - \theta) - n\lambda(\theta^* - \theta)^2 \tag{5.2.12}$$

at the true value θ_0. For this we consider the expansion

$$\begin{aligned} f(\theta^*, p) = 0 = {} & (\theta^* - \theta_0)f' + \sum f_r(p_r - \pi_r) \\ & + \frac{1}{2}\sum\sum f_{rs}(p_r - \pi_r)(p_s - \pi_s) + (\theta^* - \theta_0)\sum f_r'(p_r - \pi_r) + c_1(\theta^* - \theta_0)^2 + \epsilon, \end{aligned} \tag{5.2.13}$$

where ϵ is such that $n\epsilon \to 0$ in probability and c_1 is a constant (c_2, c_3, c_4, used later, are also constants). Substituting the value of $(\theta^* - \theta_0)f'$ from equation (5.2.13) and using (5.2.9), expression (5.2.12) reduces to

$$\begin{aligned} \frac{ni}{f'}\Bigg[\frac{1}{2}\sum\sum f_{rs}(p_r - \pi_r)(p_s - \pi_s) + (\theta^* - \theta_0)\sum f_r'(p_r - \pi_r) \\ + c_2(\theta^* - \theta_0)^2\Bigg] + \epsilon'. \end{aligned} \tag{5.2.14}$$

The asymptotic distribution of (5.2.14) is not altered if we neglect ϵ' which $\to 0$ in probability and if

$$n^{-1/2}\frac{d\log L}{d\theta} = n^{1/2}\sum\frac{\pi_r'}{\pi_r}(p_r - \pi_r) = Z_n \tag{5.2.15}$$

is substituted for $i\sqrt{n}\,(\theta^* - \theta)$. Hence the expression we have to consider is

$$\frac{i}{f'}\left[\frac{n}{2}\sum\sum f_{rs}(p_r - \pi_r)(p_s - \pi_s) + \frac{Z_n}{i}\sqrt{n}\sum f_r'(p_r - \pi_r) + c_3 Z_n^2\right]. \tag{5.2.16}$$

If relation (5.2.4) is true for all θ we have on differentiating with respect to θ,

$$f_r' = \frac{f'}{i}\frac{d}{d\theta}\left(\frac{\pi_r'}{\pi_r}\right) - \frac{\pi_r'}{\pi_r}\frac{d}{d\theta}\left(\frac{f'}{i}\right) - \sum f_{rs}\pi_s'. \tag{5.2.17}$$

Substituting for f_r' in (5.2.16), we obtain

$$\frac{ni}{2f'}\sum\sum b_{rs}(p_r - \pi_r)(p_s - \pi_s) + Z_n \sum \sqrt{n}\, a_r(p_r - \pi_r) + c_4 Z_n^2, \tag{5.2.18}$$

where

$$a_r = \frac{1}{i}\frac{d}{d\theta}\left(\frac{\pi_r'}{\pi_r}\right) - g\frac{\pi_r'}{\pi_r} \tag{5.2.19}$$

and g is chosen such that $\operatorname{Cov}[Z_n, \sum \sqrt{n}\, a_r(p_r - \pi_r)] = 0$, and

$$\begin{aligned} b_{rr} &= \left[f_{rr} - \frac{1}{i}\frac{\pi_r'}{\pi_r}\sum f_{rs}\pi_s' - \frac{1}{i}\frac{\pi_r'}{\pi_r}\sum f_{rs}\pi_s'\right], \\ b_{rs} &= \left[f_{rs} - \frac{1}{i}\frac{\pi_s'}{\pi_s}\sum f_{rs}\pi_s' - \frac{1}{i}\frac{\pi_r'}{\pi_r}\sum f_{sr}\pi_r'\right]. \end{aligned} \tag{5.2.20}$$

If we denote the first term of (5.2.18) by Q, it can be shown by actual computation that the second term in (5.2.18) has zero asymptotic covariance with Q as well as with Z_n^2. To do this, we need the fourth order raw moments of the variables $X_r = n^{1/2}(p_r - \pi_r)$ up to the constant term, that is, neglecting terms which $\to 0$ as $n \to \infty$. The fourth order moments are expressible in terms of the second order moments, that is the variances (v) and covariances (c) of the variables X_i, which are well known.

$$\begin{aligned} E(X_1^4) &= 3[V(X_1)]^2,\ E(X_1^3X_2) = 3V(X_1)\operatorname{Cov}(X_1, X_2), \\ E(X_1^2X_2^2) &= V(X_1)V(X_2) + 2[\operatorname{Cov}(X_1, X_2)]^2, \\ E(X_1^2X_2X_3) &= V(X_1)\operatorname{Cov}(X_2, X_3) + 2\operatorname{Cov}(X_1, X_2)\operatorname{Cov}(X_1, X_3), \\ E(X_1X_2X_3X_4) &= \operatorname{Cov}(X_1, X_2)\operatorname{Cov}(X_3, X_4) \\ &\quad + \operatorname{Cov}(X_1, X_3)\operatorname{Cov}(X_2, X_4) + \operatorname{Cov}(X_1, X_4)\operatorname{Cov}(X_2, X_3). \end{aligned} \tag{5.2.21}$$

Consequently the asymptotic variance of (5.2.18) is

$$V(Q) + V(c_4Z_n^2) + 2\operatorname{Cov}(Q, c_4Z_n^2) + V[Z_n \sum n^{1/2}a_r(p_r - \pi_r)], \tag{5.2.22}$$

where all second order moments refer to the asymptotic distributions. The minimum value of (5.2.18) with respect to c_4 is

$$E_2 = V(Q) - \frac{[\operatorname{Cov}(Q, Z_n^2)]^2}{V(Z_n^2)} + V[Z_n \sum n^{1/2}a_r(p_r - \pi_r)]. \tag{5.2.23}$$

The last term in (5.2.23) is independent of the equation of estimation, while the result of the first two terms is not less than zero. Therefore, the lower bound to E_2 (corresponding to maximum second order efficiency) is

$$V[Z_n \sum n^{1/2}a_r(p_r - \pi_r)] \tag{5.2.24}$$

when

$$V(Q) - \frac{[\mathrm{Cov}\,(Q, Z_n^2)]^2}{V(Z_n^2)} = 0. \tag{5.2.25}$$

A sufficient condition for equation (5.2.25) to hold is that $Q = 0$, which is true when the estimating equation is linear in the frequencies and consequently all the second order derivatives f_{rs} are zero. So we have

THEOREM 3. *Under assumptions* (5) *and* (6) *on the multinomial probabilities* $\pi_1(\theta), \cdots, \pi_k(\theta)$ *and the estimating equation* $f(\theta, p) = 0$, *the m.l. estimate has the maximum possible second order efficiency.*

There may exist other methods of estimation for which the property stated in theorem 3 is true. The necessary and sufficient condition for this is that $Q = cZ_n^2$ where c is a constant. But this does not hold for a number of estimating equations which have been suggested as alternatives to the m.l. equation, although they satisfy the condition of first order efficiency. We shall consider some of these methods.

5.3. *The computation of* E_2 *for various methods of estimation.*

(i) *Maximum likelihood.* The estimating equation for the m.l. method is

$$\sum p_r \frac{\pi_r'}{\pi_r} = 0. \tag{5.3.1}$$

Expression (5.2.18), in this case, reduces to

$$-Z_n W_n + \lambda Z_n^2, \tag{5.3.2}$$

where $iW_n = \sum n^{1/2}(d^2 \log \pi_r/d\theta^2)(p_r - \pi_r)$. The minimum variance of (5.3.2) is

$$V(Z_n W_n) - \frac{[\mathrm{Cov}\,(Z_n^2, Z_n W_n)]^2}{V(Z_n^2)}. \tag{5.3.3}$$

Using the expressions for the moments given in (5.2.21), the value of (5.3.3), which is E_2 for m.l., is found to be

$$E_2(\text{m.l.}) = \frac{\mu_{02} - 2\mu_{21} + \mu_{40}}{i} - i - \frac{\mu_{11}^2 + \mu_{30}^2 - 2\mu_{11}\mu_{30}}{i^2}, \tag{5.3.4}$$

where

$$\mu_{rs} = \sum \pi_j \left(\frac{\pi_j'}{\pi_j}\right)^r \left(\frac{\pi_j''}{\pi_j}\right)^s. \tag{5.3.5}$$

Expression (5.3.4) is the same as that obtained by Fisher ([6], p. 719), using a different approach, but the corresponding expression for the minimum chi-square given below in equation (5.3.9) does not agree with that of Fisher as given in the original paper (Fisher [6], p. 721) or in a revised form in the collected papers.

(ii) *Minimum chi-square.* The estimating equation is

$$\sum \pi_r' \frac{p_r^2}{\pi_r^2} = 0 \tag{5.3.6}$$

and the corresponding expression (5.2.18) is

(5.3.7) $$Q_n - Z_n W_n + \lambda Z_n^2,$$

where

(5.3.8) $$Q_n = -n \sum \frac{\pi_r'}{2\pi_r^2}(p_r - \pi_r)^2 + \frac{Z_n}{i} \sum n^{1/2} \left(\frac{\pi_r'}{\pi_r}\right)^2 (p_r - \pi_r).$$

The value of E_2 is

(5.3.9) $$E_2(\chi^2) = \Delta + E_2 \text{ (m.l.)},$$

where

(5.3.10) $$\Delta = \frac{1}{2} \sum \left(\frac{\pi_r'}{\pi_r}\right)^2 - \frac{\mu_{40}}{i} + \frac{1}{2}\frac{\mu_{30}^2}{i^2}$$

is not less than zero and is zero only in very special cases.

(iii) *Minimum modified chi-square* (mod. χ^2). The modified chi-square defined by Neyman [11] is

(5.3.11) $$\sum \frac{(p_r - \pi_r)^2}{p_r},$$

which, on differentiation, yields the estimating equation

(5.3.12) $$\sum \frac{\pi_r \pi_r'}{p_r} = 0.$$

The expression (5.2.18) in this case is

(5.3.13) $$2Q_n - Z_n W_n + \lambda Z_n^2$$

and, therefore,

(5.3.14) $$E_2(\text{mod. } \chi^2) = 4\Delta + E_2 \text{ (m.l.)}$$

so that the second order efficiency of the minimum modified chi-square is less than that of minimum chi-square.

(iv) *Haldane's minimum discrepancy* (D_k) *method.* Consider the estimating equation

(5.3.15) $$\sum \frac{\pi_r^k \pi_r'}{p_r^k} = 0,$$

which is essentially the same as that proposed by Haldane [8], except that Haldane used in the denominator a suitable function of p_r which does not vanish when $p_r = 0$. This modification does not alter the asymptotic results we are concerned with. Equation (5.3.15) reduces to the m.l. equation when $k = -1$ and to the minimum modified chi-square equation when $k = 1$. Expression (5.2.18), for an arbitrary k, is

(5.3.16) $$-(k+1)Q_n - Z_n W_n - \lambda Z_n^2.$$

The second order efficiency is

(5.3.17) $$E_2(D_k) = (k+1)^2 \Delta + E_2 \text{ (m.l.)}.$$

It appears that the second order efficiency is maximum when $k = -1$, which corresponds to the m.l. method. The larger the numerical value of $(k + 1)$, the greater is the loss in efficiency.

(v) *Minimum Hellinger distance* (HD). The Hellinger distance between the hypothetical probabilities and the observed proportions is

$$(p_1\pi_1)^{1/2} + \cdots + (p_k\pi_k)^{1/2}. \tag{5.3.18}$$

The estimating equation is

$$\sum \frac{\pi_r' p_r^{1/2}}{\pi_r^{1/2}} = 0, \tag{5.3.19}$$

which is a special case of (5.3.15) with $k = (-1/2)$. Hence

$$E_2 \text{ (H.D.)} = \frac{1}{4}\Delta + E_2 \text{ (m.l.)}. \tag{5.3.20}$$

(vi) *Minimizing Kullback-Liebler* (KL) *separator* $\sum \pi_r \log (\pi_r/p_r)$. The estimating equation is

$$\sum \pi_r' \log \frac{\pi}{p} = 0 \tag{5.3.21}$$

and expression (5.2.18) is

$$-Q_n - Z_n W_n + \lambda Z_n^2. \tag{5.3.22}$$

The value of E_2 is

$$E_2\left[\sum \pi \log \frac{\pi}{p}\right] = \Delta + E_2 \text{ (m.l.)} = E_2(\chi^2). \tag{5.3.23}$$

Table I provides a comparison of the different methods in terms of E_2, the index of second order efficiency, and the limiting loss of information in terms of the number of observations in large samples. The quantities E_2 (m.l.) and Δ are defined in (5.3.4) and (5.3.10) respectively.

TABLE I

Smaller values of E_2 indicate higher second order efficiency.

Method of Estimation	E_2, the Index of Second Order Efficiency	Limiting Loss in Number of Observations, over the m.l.
maximum likelihood	E_2 (m.l.)	0
minimum χ^2	$\Delta + E_2$ (m.l.)	Δ/i
minimum modified χ^2	$4\,\Delta + E_2$ (m.l.)	$4\,\Delta/i$
minimum discrepancy D_k	$(k + 1)^2\,\Delta + E_2$ (m.l.)	$(k + 1)^2\,\Delta/i$
minimum Hellinger distance	$1/4\,\Delta + E_2$ (m.l.)	$\Delta/4i$
minimum KL separator	$\Delta + E_2$ (m.l.)	Δ/i

REFERENCES

[1] F. W. Barankin and J. Gurland, "On asymptotically normal, efficient estimator. I," *Univ. California Publ. Statist.*, Vol. 1 (1951), pp. 89–130.

[2] H. Cramér, *Mathematical Methods of Statistics*, Princeton, Princeton University Press, 1946.

[3] J. Doob, "Probability and statistics," *Trans. Amer. Math. Soc.*, Vol. 36 (1934), pp. 759–775.

[4] ———, "Statistical estimation," *Trans. Amer. Math. Soc.*, Vol. 39 (1936), pp. 410–421.

[5] R. A. Fisher, "On the mathematical foundations of theoretical statistics," *Philos. Trans. Roy. Soc. London, Ser. A.*, Vol. 222 (1922), pp. 309–368.

[6] ———, "Theory of statistical estimation," *Proc. Cambridge Philos. Soc.*, Vol. 22 (1925), pp. 700–725.

[7] ———, *Statistical Methods and Scientific Inference*, London, Oliver and Boyd, 1956.

[8] J. B. S. Haldane, "A class of efficient estimates of a parameter," *Bull. Int. Statist. Inst.*, Vol. 33 (1951), p. 231.

[9] G. Kallianpur and C. R. Rao, "On Fisher's lower bound to asymptotic variance of a consistent estimate," *Sankhyā*, Vol. 15 (1955), pp. 331–342.

[10] L. Le Cam, "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *Univ. California Publ. Statist.*, Vol. 1 (1953), pp. 227–330.

[11] J. Neyman, "Contribution to the theory of certain test criteria," *18th Session de L'Institut International de Statistique*, Vol. 1 (1929).

[12] ———, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1949, pp. 239–273.

[13] C. R. Rao, "Maximum likelihood estimation for the multinomial distribution," *Sankhyā*, Vol. 18 (1957), pp. 139–148.

[14] ———, "Maximum likelihood estimation for the multinomial distribution with infinite number of cells," Sankhyā, Vol. 20 (1958), pp. 211–218.

[15] ———, "Apparent anomalies and irregularities of the method of maximum likelihood," to appear in *Proceedings of the 32nd Session of the International Statistical Conference, Tokyo*, 1960.

[16] A. Wald, "Asymptotically shortest confidence intervals," *Ann. Math. Statist.*, Vol. 13 (1942), pp. 127–137.

[17] ———, "Tests of statistical hypotheses concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, Vol. 54 (1943), pp. 426–482.