Sankhyā : The Indian Journal of Statistics
1974, Volume 36, Series A, Pt. 4, pp. 325-358.

SECOND ORDER EFFICIENCY OF MAXIMUM LIKELIHOOD ESTIMATORS

By J. K. GHOSH
Indian Statistical Institute
and
K. SUBRAMANYAM[1]
University of Dar es Salaam

SUMMARY. The results of Fisher and Rao are extended to exponential families with more than one parameter; the proof is new even when specialised to one parameter multinomials of Fisher and Rao. The results are applied to a bioassay problem of Berkson.

The paper also develops Bhattacharya type inequalities, a (formal) Bayesian proof of the results on second order efficiency and the notion of asymptotic sufficiency up to $O(1/n)$.

1. Introduction

The word second order efficiency was introduced by Rao (1961) but, as noted there, the concept as well as the first main result in this area, occurs in Fisher (1925). In that famous paper, Fisher proposes

$$E_2' = \lim (nI - I_{T_n})$$

as a measure of second order efficiency to discriminate between different asymptotically efficient estimators in a problem of estimation involving a multinomial population; here I is Fisher's information contained in a single observation and I_{T_n} the information contained in T_n. Fisher stated, without any sort of proof, that the maximum likelihood estimator minimizes E_2', i.e., maximizes second order efficiency. He also calculated the values of this or rather an intuitively plausible approximation to it (in fact E_2 of Rao (1961)) for the maximum likelihood and minimum chi-square estimators. There is a slight mistake in the calculations; see Rao (1961) for the corrections. For some clarification of Fisher's calculations see

[1]This work was started when both authors were at the Indian Statistical Institute. It was finished while the first author was at the University of Illinois and the second author at the University of Dar es Salaam. The research of the first author was supported in part by NSF grant GP 28154 Amendment Section 2. The paper is dedicated to the memory of the late Professor P. C. Mahalanobis.

Kendall (1946) and Nandi (1956). Some discussion of the relation among Fisher's criterion E_2', the quantity he actually calculated and the quantity E_2 of Rao (1961) is contained in remark 6 after Theorem 1 and Section 5. Rao (1961) has provided the motivation for measures like E_2' and E_2.

Somewhat surprisingly second order efficiency remained neglected till it was picked up by Rao (1961) who makes major progress by proving Fisher's result. However the result actually proved differs in two ways from what Fisher stated. First Rao introduces a more easily computed and a more useful measure E_2 and secondly he restricts attention to what he calls Fisher consistent estimators with continuous second order derivatives. This result will be referred to as the Fisher-Rao theorem. We shall call Fisher consistent estimators with continuous second order derivatives, or rather a slightly wider class, locally stable (II). The definitions of E_2 and local stability (II, III) are given in Section 2.

The Fisher-Rao theorem has one unpleasant feature—its decision theoretic implications are far from clear. In fact this has been the main criticism against its use to justify the use of maximum likelihood estimators. Rao (1963) has, therefore, sought a direct comparison of the truncated mean squares. Let $W(a, \theta) = \min\{(a-\theta)^2, d\}$ be the squared error loss truncated at $d > 0$. (Actually Rao's loss is slightly different; see the remark after proposition 2.) Suppose T_n is an asymptotically efficient estimator with

$$E_\theta\{W(T_n, \theta)\} = \frac{1}{nI}+\frac{\psi}{n^2}+o\left(\frac{1}{n^2}\right)$$

then ψ may be taken as a third measure of second order efficiency. Again restricting to Fisher consistent estimators (with third order continuous derivatives instead of second order) and applying a bias correction to the estimators considered, he shows ψ is minimized by the (corrected) maximum likelihood estimator. The effect of the bias correction is to make the estimators unbiased up to terms of $O(1/n)$; we shall call this Rao's theorem.

There are quite a number of papers on the expansion of the variance of the maximum likelihood estimator and some other estimators in a multinomial set-up. A recent paper is by Robertson (1972).

Both these theorems pertain to the case of independent random samples from a multinomial population with proportions depending on an unknown parameter θ. Data which appears to contradict this sort of result in a particular bio-assay problem, has been presented by Berkson (1955). Berkson's data seems to indicate that for moderate sample size his minimum logit-chi-square estimator performs better than the maximum likelihood estimator as regards bias and mean square error. In this connection see also Berkson and Hodges (1961). A summary of the results of Berkson is available in Ferguson (1967). Since the population which Berkson considers is not multinomial but belongs to the Korpman-Darmois exponential family it seemed to us worth extending the results of Rao (1961, 1963) to exponential families to see what is really happening in Berkson's problem.

The extension to exponential families is carried out in Section 2. The main idea is simple. It is shown that all locally stable efficient estimators T_n which are unbiased up to $O(1/n)$ have same covariance up to $o(1/n^2)$ with $Z_n, Z_n^2, Z_n W_n$ (which are defined on pp. 331, 335). Moreover $\hat{\theta}_n$ even after bias correction is easily shown to be a linear function of Z_n, Z_n^2 and $Z_n W_n$ up to $o(1/n)$. So up to $o(1/n)$ we can write T_n as a sum of two orthogonal components the first of which is the bias corrected maximum likelihood estimator. Rao's theorem is an immediate consequence. The Fisher-Rao theorem follows similarly. The expansions given in Theorem 1 try to make clear the relation between the two types of results from the present point of view. Moreover it is shown that if T_n is an efficient l.s. (III) estimator then one can find $h(\hat{\theta}_n) = \hat{\theta}_n + g(\hat{\theta}_n)/n$ such that $E_\theta\{W(T_n, \theta)\} \geqslant E_\theta\{W(h(\hat{\theta}_n), \theta)\} + o(1/n^2) \forall \theta$. Extension to the multi-parameter case is briefly indicated. An asymptotic Bhattacharya bound is developed and necessary and sufficient conditions are given for the maximum likelihood estimator to attain it. The calculations in this section, though similar to Rao's are, we believe, somewhat simpler and more illuminating even when specialized to the multinomial case. Unlike Rao (1961, 1963), all the details necessary for rigour and precision have been spelled out to make the results easily accessible to all readers. We have not hesitated to present more than one derivation of a result whenever it helps clarification.

In Section 3 these results are applied to Berkson's problem. It is shown that if a correction is made to the maximum likelihood estimator so that its bias is the same as that of Berkson's minimum logit-chi-square estimator up to terms $O(1/n)$, then the maximum likelihood estimator has a lower variance up to the terms of $O(1/n^2)$.

In the next section we approach the problem from a Bayesian point of view. Using the results of Lindley (1961) a heuristic argument is given to show that these theorems hold quite generally and not merely in the restricted set-up considered in Section 2. We hope to present later a rigorous justification of this result using the expansions of Johnson (1970). It is pointed out that Lindley's comments in the discussion following Rao (1962) are not justified.

The last part of the paper is devoted to miscellaneous remarks about second order asymptotic sufficiency, expansion for the asymptotic distribution of a stable estimator and related matters.

After this paper was prepared Pfanzagl (1973) has published a very interesting paper on closely related results. The techniques are quite different. The assumptions are not quite comparable. Pfanzagl considers only the absolutely continuous case but for this case his assumptions are much weaker than ours. The results are also not comparable. We use different criteria and our results are true up to terms of smaller order than Pfanzagl's.[1] However both our results and Pfanzagl's are consequences of asymptotic sufficiency of $\hat{\theta}_n$ and another statistic up to $o(n^{-1})$; we shall return to this problem elsewhere. The definition of asymptotic sufficiency up to $o(n^{-1})$ is given in the last section. Incidentally we are unable to understand Pfanzagl's criticism (1973, .p 1006) of Rao's results as unmotivated. Adequate motivation is provided by (ii) and (iii) of our Theorem 1.

The techniques and results of Efron (1974), of which the authors came to know at the final stage of revision, are much more relevant for our purpose. For most of the paper Efron also considers multiparameter exponential families. Efron also introduces the Bhattacharya bounds. It is interesting to note that Efron has provided a counter example to show $E_2' \neq E_2$ in general. (See in this connection our remarks in Section 5). Efron's main contribution in this paper is the very useful and elegant notion of the "curvature" of the problem which in the exponential case happens to be a geometrical curvature of the parameter curve $\{\beta(\theta), \theta \epsilon \Theta\} \epsilon V'$, invariant under 1—1 smooth parametric transformations. (For the definition of these symbols see Section 2.). Efron shows its relevance for many concrete statistical problems including that of second order efficiency.

2. Second order efficiency for exponential families

$\{x_i\}$ is a sequence of i.i.d. random variables taking values in some measurable space (S, A). Let R^k be the k-dimensional Euclidean space. For each $\beta = (\beta_1, ..., \beta_k)$ lying in some fixed open set $V' \subset R^k$, let x_i's have probability density $f'(x, \beta)$ with respect to some non-degenerate σ-finite measure μ on (S, A). We assume this is an exponential family, i.e.,

$$f'(x, \beta) = c'(\beta) \exp\left\{\sum_{j=1}^{k} \beta_j p_j(x)\right\}$$

where the p_j's are real-valued measurable functions. Let $p_0(x) = 1 \forall x$. We

[1]For example Pfanzagl's Theorem 6 would fail to discriminate in the class of estimators considered here.

assume $p_0, p_1, \ldots, p_k$ are linearly independent in the sense $\sum_{j=1}^{k} c_j p_j = 0$ a.e. $\mu \Longrightarrow c_0 = c_1 = \ldots = c_k = 0$.

Let
$$h_i(\beta) = E(p_i|\beta) = \int p_i(x) f'(x, \beta) d\mu. \tag{2.1}$$

Then
$$\frac{\partial h_i}{\partial \beta_j} = -\frac{\partial^2 \log c'(\beta)}{\partial \beta_i \partial \beta_j} = \text{cov}\,(p_i, p_j|\beta). \quad \ldots \tag{2.2}$$

Thus
$$\left[\frac{\partial h_i}{\partial \beta_j}\right] \quad i, j, = 1, \ldots, k$$

is the $k \times k$ dispersion matrix of $p_1, \ldots, p_k$ and it is positive definite since p_0, $p_1 p_k \ldots$, are linearly independent. So for each $\beta^0 \epsilon V' \exists$ an open neighbourhood V of β^0 such that it is contained in V' and restricted to V the map

$$\beta \xrightarrow{h} h(\beta)$$

is one-one and onto an open set W in R^k. We fix such a V and such a W and denote the inverse of this map as

$$\pi \xrightarrow{\beta} \beta(h)$$

from W onto V. We now introduce an alternative parameterization $\{f(x, \pi); \pi \epsilon W\}$ for the family $\{f'(x, \beta); \beta \epsilon V\}$ where for $\pi \epsilon W, f(x, \pi) = f'(x, \beta(\pi))$. Writing $c(\pi) = c'(\beta(\pi))$, we get

$$f(x, \pi) = c(\pi) \exp\left\{\sum_{j=1}^{k} \beta_j(\pi) p_j(x)\right\}.$$

The statistical problem that we consider is one where $\pi_1, \ldots, \pi_k$ are known functions $\pi_1(\theta), \ldots, \pi_k(\theta)$ of a single unknown real parameter θ lying in the parametric space Θ and θ has to be estimated on the basis of observations $x_1, \ldots, x_n$.

Our assumption on the functions $\pi_1(\theta), \ldots, \pi_k(\theta)$ are stated below.

Assumption 1 : *Θ is an open set. For $\theta \epsilon \Theta$, $\pi(\theta) \epsilon W$ and $\pi_i(\theta)$ is thrice continuously differentiable on Θ, $i = 1, \ldots, k$. The rank of $(\pi_1'(\theta), \ldots, \pi_k'(\theta)) = 1 \forall \theta \epsilon (\Theta)$.*

Note that partial derivatives of all orders of $\beta_1, \ldots, \beta_k$ with respect to $\pi_1, \ldots, \pi_k$ exist at all points of W. Hence by Assumption 1 $\beta_1, \ldots, \beta_k$ are thrice continuously differentiable functions of θ. Since (i) rank of $(\pi_1'(\theta), \ldots, \pi_k'(\theta))$ is one, (ii) $\beta(\pi(\theta))$ is an interior point of V and (iii) μ is non-degenerate it is easy to see that the Fisher information $I(\theta) = E\left\{\frac{d\log L}{d\theta}\right\}^2 = E_\theta\left\{\Sigma \frac{\partial \log L}{\partial \pi_i}. \pi_i'(\theta)\right\}^2$ is finite and positive. $I'(\theta)$ also exists. Here

$$L = \{c(\pi(\theta))\}^n \exp\left\{n \sum_{j=1}^{k} \beta_j(\pi(\theta))p_j^n\right\}.$$

We shall henceforth write $c(\theta)$ for $c(\pi(\theta))$ and $\beta(\theta)$ for $\beta(\pi(\theta))$.

Let $p_i^n = \left(\sum_1^n p_i(x_j)\right)/n$. Since $p^n = (p_1^n, \ldots, p_k^n)$ is sufficient, we consider only estimators of the form $T_n = T_n(p^n)$ which depend on $x_1, \ldots, x_n$ only through p^n. By an estimator T_n we shall actually mean a sequence of estimators $\{T_n\}$.

Consider the following conditions

(i) For each $\theta \epsilon \Theta$ there is an open neighbourhood V_θ of $\pi(\theta)$ with compact closure $\overline{V}_\theta$ s.t. $\overline{V}_\theta \subset W$ and the domain of definition of T_n includes $\overline{V}_\theta$. Moreover, $T_n(p) = T(p) \forall n$ if $p \epsilon V_\theta$ and $\frac{\partial T}{\partial p_i}, \frac{\partial^2 T}{\partial p_i \partial p_j}$ $i, j = 1, \ldots, k$, exist and are continuous on $\overline{V}_\theta$.

(ii) $T_n(\pi(\theta)) = \theta \forall \theta \epsilon \Theta$.

If T_n satisfies these conditions we shall say T_n is locally stable of order two which will be abbreviated as l.s. (II). If T_n satisfies these conditions and has third order continuous derivatives, we shall say T_n is l.s. (III). If T_n has only continuous first order derivatives but otherwise satisfies (i) and (ii) it is l.s.(I).

Condition (i) is a stationarity and smoothness requirement which is likely to stabilize the large sample properties of T_n. For example convergence to an asymptotic distribution may be expected to be, in general, more rapid with condition (i) than without it. If T_n satisfies (i) then T_n is consistent iff it satisfies (ii). Note that the neighbourhood V_θ may be different for different estimators.

We consider the likelihood equation

$$\frac{d\log L}{d\theta} = 0$$

i.e.,

$$\frac{d\log c(\theta)}{d\theta} + \sum_{i=1}^{k} \beta_i'(\theta)p_i^n = 0 \quad \ldots \quad (2.3)$$

If $p_i^* = \pi_i(\theta_0)$ then $\theta = \theta_0$ is a solution. Since, moreover,

$$\left.\frac{d^2 \log c(\theta)}{d\theta^2}\right|_{\theta_0} + \sum_{i=1}^{k} \beta_i' \pi_i(\theta) = I(\theta_0) \neq 0,$$

it follows by the implicit function theorem that in a suitable neighbourhood V_θ of $\pi(\theta_0)$ the likelihood equation has a solution $\hat{\theta}_n = \hat{\theta}(p^n)$ which is a thrice continuously differentiable function of p^n under Assumption 1 and thrice continuously differentiable under Assumption 1'. The maximum likelihood estimator $\hat{\theta}_n$ is l.s. (III) if Assumption 1 or 1' holds. We have defined $\hat{\theta}_n$ only locally but if Θ has a compact closure and π on Θ has a continuous extension on the closure then it is not hard to combine the local definitions to get a global definition on a suitable neighbourhood of the curve $\{\pi(\theta); \theta \epsilon \Theta\}$. $\hat{\theta}_n$ can be defined in any way one likes outside the neighbourhood.

Following Rao (1961) T_n is said to be efficient up to first order or asymptotically efficient or simply efficient if for some α and $\beta > 0$, which may depend on θ,

$$|n^{\frac{1}{2}}Z_n - \alpha - \beta n^{\frac{1}{2}}(T_n - \theta)| \to 0 \qquad \ldots \quad (2.4)$$

in probability under θ, where $Z_n = \frac{1}{n}\frac{d \log L}{d\theta}$. Hajec (1970) has proved under quite general conditions that T_n has a certain locally asymptotically minimax property iff T_n is efficient up to first order and $\beta = I$ where I is Fisher's information.

Suppose T_n is l.s. (I). Then

$$\begin{aligned}\sqrt{n}(T_n - \theta) &= \sqrt{n}(T_n(p^n) - T_n(\pi(\theta)) \\ &= \sqrt{n}\Sigma(p_j^n - \pi_j(\theta))T^j + o_p(1) \qquad \ldots \quad (2.5)\end{aligned}$$

where $T^j = \left.\frac{\partial T(p)}{\partial p_j}\right|_{\pi(0)}$ and $o_p(1)$ is a term which tends to zero in probability.

It follows from (2.5) that (2.4) holds for a l.s.(I) estimator T_n iff

$$\alpha = 0 \text{ and } \beta_i' = \beta T^i. \qquad \ldots \quad (2.6)$$

To evaluate β we proceed as in Rao (1961).

Since $T_n(\pi(\theta)) = T(\pi(\theta)) = \theta$, we get on differentiating with respect to θ,

$$\Sigma T^j \pi_j'(\theta) = 1 \qquad \ldots \quad (2.7)$$

where $$\pi_j' = \frac{d\pi_j}{d\theta}.$$

From (2.6) and (2.7),

$$\beta = \beta \Sigma T^j \pi_j' = \Sigma \beta_j' \pi_j'. \qquad \dots \quad (2.8)$$

But
$$\frac{d^2 \log c}{d\theta^2} + \Sigma \beta_j' \pi_j + \Sigma \beta_j' \pi_j' = \frac{d}{d\theta}\left[\frac{d \log c}{d\theta} + \Sigma \beta_j' \pi_j\right]$$

$$= \frac{d}{d\theta} E_\theta\left(\frac{d \log f_\theta(x_1)}{d}\right) = 0.$$

So
$$\Sigma \beta_j' \pi_j' = -\frac{d^2 \log c}{d\theta^2} - \Sigma \beta_j' \pi_j = -E_\theta\left(\frac{d^2 \log f_\theta(x_1)}{d\theta^2}\right) = I \qquad \dots \quad (2.9)$$

where $I \equiv I(\theta)$ is Fisher's information.

From (2.8) and (2.9), we get $\beta = I$. Hence it follows from (2.9) that a necessary and sufficient condition for a l.s. (I) estimator to be efficient up to first order is

$$T^j = \beta_j' \times \frac{1}{I} \;\forall \theta \epsilon \Theta. \qquad \dots \quad (2.10)$$

Before defining second order efficiency let us state a simple lemma.

Let us fix $\theta \epsilon \Theta$. Let $U = \{p; |p_i - \pi_i(\theta)| < \delta,\ i = 1, \dots, k\}$ where $\delta > 0$ is chosen so that $U \subset W$. Let I_U and I_{U^c} denote the indicator functions of U and its complement U^c. We shall also use I_U and I_{U^c} to denote $I_U(p^n)$ and $I_{U^c}(p^n)$.

Lemma 1 :
$$P_\theta\{p^n \epsilon U^c\} < A\rho^n \qquad \dots \quad (2.11)$$

$$E_\theta[\{|p_i^n - \pi_i(\theta)|^r\}\{|p_j^n - \pi_j(\theta)|^s\} I_{U^c}] < B\rho^{n/2} \qquad \dots \quad (2.12)$$

for some $0 < \rho < 1$, $A > 0$, $B > 0$, *provided* $r, s \geqslant 0$.

Before proving the lemma we note that ρ depends on θ and B depends on i, j, r and s in addition to θ.

Proof : Let
$$\rho_{i_1} = \inf_{t \geqslant 0} E_\theta[\exp\{t(p_i' - \pi_i(\theta) - \delta)\}]$$

$$\rho_{i_2} = \inf_{t \leqslant 0} E_\theta[\exp\{t(p_i' - \pi_i(\theta) + \delta)\}]$$

$$\rho_i = \max(\rho_{i_1}, \rho_{i_2}),\ \rho = \max_{1 \leqslant i \leqslant k} \rho_i.$$

Of course $0 < \rho < 1$. Clearly,

$$P_\theta\{p^n \epsilon U^c\} \leqslant \sum_1^k P_\theta\{|p_i^n - \pi_i(\theta)| \geqslant \delta\}$$

$$\leqslant 2 \sum_1^k \rho_i^n(\theta) \text{ by Chernoff's (1952, p. 495) inequality}$$

$$\leqslant 2k\rho^n.$$

So (2.11) holds. Also,

$$E_\theta\{|p_i^n - \pi_i(\theta)|^r |p_j^n - \pi_j(\theta)|^s . I_{U^c}\}$$

$$\leqslant [E_\theta\{|p_i^n - \pi_i(\theta)|^{4r}\}]^{1/4} [E_\theta\{|p_j^n - \pi_j(\theta)|^{4s}\}]^{1/4} . [E_\theta\{I_{U^c}\}]^{\frac{1}{2}}$$

by two applications of the Cauchy-Schwarz inequality. The first two terms on the right hand side are bounded in n (in fact, go to zero). For

$$E_\theta\{|p_i^n - \pi_i(\theta)|^{4r}\} \leqslant E_\theta\{|p_i^1 - \pi_i(\theta)|^{4r}\} \qquad \text{if } r \geqslant 1/4$$

$$\leqslant 1 + E_\theta\{|p_i^1 - \pi_i(\theta)|\} \quad \text{if } 0 \leqslant r \leqslant 1/4.$$

So (2.12) now follows from (2.11).

We shall now describe Rao's first measure of second order efficiency for a l.s. (II) estimator T_n, which is efficient up to first order. Fix θ_0. We shall think of θ_0 as the true value of the parameter. Let V_{θ_0} be the open neighbourhood of $\pi(\theta_0)$ which we may associate with T_n by definition of local stability (II) and

$$U = \{p;\ |p_i - \pi_i(\theta_0| < \delta\} \subset V_{\theta_0} \qquad i = 1, \ldots, k.$$

For any random variable Z let $E^U(Z)$ denote $E_{\theta_0}(Z.\ I_U)$ where $I_U = I_U(p^n)$. Let

$$Z_n = n^{-1} \frac{d \log L}{d\theta}\bigg|_{\theta_0} \qquad \ldots \quad (2.13)$$

Recall that $I = I(\theta_0) = n.E_{\theta_0}(Z_n^2)$ and $E_{\theta_0}(Z_n) = 0$.

The proof of the following auxiliary proposition is given in the appendix. We use $\{T_n\}$ to indicate the sequence of estimators T_n.

Proposition 1 : Let T_n be l.s. (II) and efficient. Then

(i) $a_\lambda(\theta_0) = \lim_{n\to\infty} E^U n\{Z_n-(T_n-\theta_0)I-\lambda(T_n-\theta_0)^2\}$ exists

(ii) $E_2(\{T_n\}, \theta_0, \lambda, U) = \lim_{n\to\infty} [n^2 . E^U\{Z_n-(T_n-\theta_n)I$

$$-\lambda(T_n-\theta_0)^2-a_\lambda(\theta_0)/n\}^2]$$

exists.

(iii) $E_2(\{T_n\}, \theta_0, \lambda, U_1) = E_2(T_n, \theta_0, \lambda, U_2)$

where U_i *is a neighbourhood of* $\pi(\theta_0)$ *contained in* V_{θ_0}, $i = 1, 2$.

In view of (iii) we shall write $E_2(\{T_n\}, \theta_0, \lambda)$ for $E_2(\{T_n\}, \theta_0, \lambda, U)$. Let

$$E_2(\{T_n\}, \theta_0) = \inf_\lambda E_2(\{T_n\}, \theta_0, \lambda).$$

We can think of E_2 as a measure of how well a quadratic in T_n approximates Z_n. If Z_n were a function of $T_n \forall \theta_0$, T_n would be a sufficient statistic. So E_2 measures, in a sense, how "nearly" sufficient T_n is. The reason for taking a quadratic in T_n is mainly one of expediency. In Rao (1961) E_{θ_0} is used instead of E^U but the calculations can be justified only with E^U. See in this connection Rao (1963) where essentially the present approach is followed. The intuitive justification for using E^U is that we do not wish our measure to be unduly affected by the tail of the distribution of the estimator. If we are comparing two l.s. (II) efficient estimators $T_n^{(1)}$, $T_n^{(2)}$ we may take $U \subset V_{\theta_0}^{(1)} \cap V_{\theta_0}^{(2)}$ for the calculation of E_2 for both estimators, to remove the apparent arbitrariness of U and hence of the method of comparing $T_n^{(1)}$ and $T_n^{(2)}$.

To define Rao's other measure of second order efficiency we need the following proposition the proof of which is given in the appendix.

Proposition 2 : Let T_n be l.s. (II) and efficient.

Then

(i) $b(\theta_0) = \lim n\{E^U(T_n)-\theta_0\}$ exist.

If moreover T_n is l.s. (III) then the following results hold.

(ii) $b(\theta)$ is a continuously differentiable function on Θ.

(iii) If $T_n^* = T_n-b(T_n)/n$ then

$$E^U(T_n^*) = \theta_0+o(1/n).$$

(iv) If $T_n' = T_n + m(T_n)/n$ where m is continuously differentiable in a neighbourhood of θ_0, then

$$E^U(T_n'-\theta_0)^2 = \frac{1}{nI} + \frac{\psi(\{T_n'\}, \theta_0)}{n^2} + o\left(\frac{1}{n^2}\right)$$

where $\psi(\{T_n'\}, \theta_0)$ does not depend on U.

(v) $E_{\theta_0}\{W(T_n', \theta_0)\} - E^U(T_n'-\theta_0)^2 = o\left(\frac{1}{n^2}\right)$

where T_n' is defined as in (iv) $W(a, \theta) = \min\{(a-\theta)^2, d\}$ is the squared error loss truncated at $d > 0$.

Rao (1963) takes $E^U(T_n'-\theta_0)^2$ as the risk function of T_n' but Proposition 2 shows that it does not matter up to $o(n^{-2})$ whether we take $E^U(T_n'-\theta_0)^2$ or $E_{\theta_0}\{W(T_n', \theta_0)\}$ as our risk function. Following Rao we take $\psi(\{T_n'\}, \theta_0)$ as our second measure of second order efficiency of $\{T_n'\}$.

We shall now introduce a few more notations and then state our main result. Let

$$W_n = \frac{1}{n} \left.\frac{d^2 \log L}{d\theta^2}\right|_{\theta_0} + I(\theta_0) \qquad \text{... (2.14)}$$

$$= \frac{d^2 \log c(\theta)}{d\theta^2} + \Sigma \beta_i''(\theta) p_i^n + I(\theta_0). \qquad \text{... (2.15)}$$

Clearly $E_{\theta_0}(W_n) = 0$. Let

$$\mu_{rs} = E_{\theta_0}(Z_1^r W_1^s) \qquad \text{... (2.16)}$$

where Z_n is defined earlier by (2.13). Note that $I(\theta_0) = \mu_{02}$. As stated before we shall often write I for $I(\theta_0)$. Let

$$J = E_{\theta_0}\left\{\left.\frac{d^3 \log f_\theta(X_1)}{d\theta^3}\right|_{\theta_0}\right\} = \frac{d^3 \log c(\theta)}{d\theta^3} + \Sigma \beta_i'''(\theta_0)\pi_i(\theta_0) \qquad \text{... (2.17)}$$

The random variable

$$S_n = \{Z_n W_n - \mu_{11}/n\} I^2 + \{Z_n^2 - I/n\} J/2I^3 \qquad \text{... (2.18)}$$

will play an important role in what follows.

If Y_n is a sequence of random variables such that $E^U\{Y_n^2\} = o(a_n^2)$ or $O(a_n^2)$ we shall write $Y_n o_E(a_n)$ or $O_E(a_n)$ accordingly. A random variable X

will be called E^U-orthogonal to another random variable Y if covariance, under θ_0, of XI^U and Y is zero; X and Y will be said to be E^U-orthogonal up to $o\left(\frac{1}{n^2}\right)$ if the covariance of XI^U and Y is $o\left(\frac{1}{n^2}\right)$.

Recall from Proposition 1 that if T_n is efficient and l.s. (II) then $E_{\theta_0}(T_n) = \theta_0+b(\theta_0)/n+o\left(\frac{1}{n}\right)$. Under Assumption 1, $\hat{\theta}_n$ is efficient and l.s. (III) and so we may write

$$E_{\theta_0}(\hat{\theta}_n) = \theta_0+b_0(\theta_0)/n+o\left(\frac{1}{n}\right). \qquad \ldots \quad (2.19)$$

We can now state our main result.

Theorem 1 : *Suppose Assumption 1 holds.*

(i) *Then*

$$\hat{\theta}_n-\theta_0-Z_n/I = b_0(\theta_0)/n+S_n+\hat{R}_n \qquad \ldots \quad (2.20)$$

where $\hat{R}_n$ *is* $o_E(n^{-1})$ *and* E^U*-orthogonal to* Z_n^2 *and* Z_nW_n *up to* $o(n^{-2})$,

$$\hat{\theta}_n^*-\theta_0-Z_n/I = \frac{-Z_n}{2nI^4}\{2\mu_{21}I^5+J\mu_{30}\}+S_n+R_n^* \qquad \ldots \quad (2.21)$$

where $\hat{R}_n^*$ *is* $o_E(n^{-1})$ *and* E^U*-orthogonal to* Z_n, Z_n^2, Z_nW_n *up to* $o(n^{-2})$.

(ii) *Let* T_n *be efficient and l.s.* (II). *Then*

$$T_n-\theta_0-Z_n/I = b(\theta_0)/n+S_n+R_n \qquad \ldots \quad (2.22)$$

where R_n *is* $O_E(n^{-1})$ *and* E^U*-orthogonal to* Z_n^2 *and* Z_nW_n *up to* $o(n^{-2})$ *and* $E^U(R_n) = o(n^{-1})$. *Also*

$$E_2(\{T_n\},\ \theta_0) \geqslant E_2(\{\hat{\theta}_n\},\theta_0)\,\forall\theta_0 \in\Theta. \qquad \ldots \quad (2.23)$$

Let T_n *be efficient and l.s.* (III)

$$T_n^*-\theta_0 = (\hat{\theta}_n^*-\theta_0)+R_n^* \qquad \ldots \quad (2.24)$$

where R_n^* *is* $O_E\left(\frac{1}{n}\right)$, E^U*-orthogonal to* Z_n, Z_n^2, $Z_n W_n$ *and hence to* $(\hat{\theta}_n^*-\theta_0)$ *up to* $o\left(\frac{1}{n^2}\right)$ *and*

$$\psi(\{T_n^*\},\ \theta_0) \geqslant \psi(\{\hat{\theta}_n^*\},\theta_0)\,\forall\theta_0\in\Theta \qquad \ldots \quad (2.25)$$

Moreover

$$E_2(\{T_n\},\theta_0) = I^2\psi(\{T_n^*\},\theta_0)-\frac{2}{I^2}\{J/2+\mu_{11}\}^2. \qquad \ldots \quad (2.26)$$

(iii) *Let* T_n *be efficient and l.s.*(III) *and* $m(\theta)$ *a continuously differentiable function on* Θ. *Let* $T_n' = T_n + m(T_n)/n$. *Then* $\exists$ *a continuously differentiable function* g *on* Θ *such that* $\theta_n' = \theta_n + g(\theta_n)/n$ *is better than* T_n' *up to* $o\left(\frac{1}{n^2}\right)$ *in the sense that*

$$\lim n^2[E_{\theta_0}\{W(T_n', \theta_0)\} - E_{\theta_0}\{W(\theta_n', \theta_0)\}] \geqslant 0$$

where $W(a, \theta) = \min\{(a-\theta)^2, d\}$ *is the squared error loss, truncated at* $d > 0$.

We shall need a few lemmas to prove this result.

Lemma 2 : *If Assumption* 1 *holds and* T_n *is efficient and l.s.* (II) *then*

$$T_n(p^n) - \theta_0 = \frac{Z_n}{I} + \frac{1}{2}\sum\sum(p_i^n - \pi_i(\theta_0))(p_j^n - \pi_j(\theta_0))T^{ij} + o_E\left(\frac{1}{n}\right) \quad \dots \quad (2.27)$$

where T^{ij}*'s are the second order derivatives of* T *evaluated at* $\pi(\theta_0)$. *If moreover* T_n *is l.s.* (III) *then*

$$T_n(p^n) - \theta_0 = \frac{Z_n}{I} + \frac{1}{2}\sum\sum(p_i^n - \pi_i)(p_j^n - \pi_j)T^{ij} +$$

$$+\frac{1}{6}\sum\sum\sum(p_i^n - \pi_i)(p_j^n - \pi_j)(p_l^n - \pi_l)T^{ijl} + o_E(n^{-3/2}) \quad \dots \quad (2.27a)$$

where T^{ijl}*'s are the third order derivatives of* T *at* $\pi(\theta_0)$.

Proof : For $p^n \in U$, consider the Taylor expansion

$$T_n(p^n) - \theta_0 = T(p^n) - T(\pi(\theta_0))$$

$$= \Sigma\{p_j^n - \pi_j(\theta_0)\}T^j + \Sigma\Sigma(p_i^n - \pi_i(\theta_0))(p_j^n - \pi_j(\theta_0))T^{ij} + R(p^n) \quad \dots \quad (2.28)$$

where T^j's and T^{ij}'s are the first and second order derivatives and R is the remainder term. Note that

$$R(p^n) = \epsilon(p^n)\,\Sigma\Sigma(p_j^n - \pi_i(\theta_0))(p_i^n - \pi_j(\theta_0)T^{ij} \quad \dots \quad (2.29)$$

where $\epsilon(p) \to 0$ as $p \to \pi(\theta_0)$ and $|\epsilon(p)| < M$ on U for some suitable M. Fix $\eta > 0$ and choose $0 < \delta_1$ such that

$$U_1 = \{p; |p_i - \pi_i(\theta_0)| < \delta_1, \quad i = 1, \dots, k\} \subset U$$

and

$$|\epsilon(p)| < \eta \quad \text{if} \quad p \in U_1.$$

Then

$$E^U(R^2(p^n)) = E^{U_1}(R^2) + E^U(I_{U_1^c}R^2)$$

$$\leqslant \eta\Sigma\,\Sigma E_\theta\{|p_i^n - \pi_i| \cdot |p_j^n - \pi_j|\}|T^{ij}|$$

$$+ M\Sigma\,\Sigma E_{\theta_0}\{|p_i^n - \pi_i|\,|p_j^n - \pi_j|\,I_{U_1^c}\}|T^{ij}| \quad \dots \quad (2.30)$$

by (2.29). The second term in (2.30) is $M \cdot O(\rho^{n/2})$ by Lemma 1 which is obviously valid with U_1 in place of U. The first term in (2.30) is $\eta \cdot O(1/n^2)$. Since η is arbitrary, it follows that $E^U(R^2) = o(1/n^2)$. This fact together with (2.10) and (2.28) completes the proof.

The above proof is simple but details are given for completeness. In following pages when we assert that a random variable is $o_E(1/n)$, $O_E(1/n)$ etc. we shall not usually give a proof. But in each case justification is easy and involves an application of Lemma 1, aided perhaps by the Cauchy-Schwarz inequality.

Unless otherwise stated we shall take T_n always as efficient and l.s. (II). It is clear from Lemma 2, that

$$b(\theta_0) = \tfrac{1}{2} \Sigma\Sigma T^{ij} E_{\theta_0}(p_i' - \pi_i(\theta_0))(p_j' - \pi_j(\theta_0)). \qquad \ldots \quad (2.31)$$

If T_n is also l.s. (III) then by Proposition 2 b is continuously differentiable. Hence it can be shown that

$$\frac{b(T_n)}{n} = \frac{b(\theta_0)}{n} + (T_n - \theta_0)\,\frac{b'(\theta_0)}{n} + o_E(n^{-\frac{3}{2}})$$

$$= \frac{b(\theta_0)}{n} + \frac{Z_n}{I} \cdot \frac{b'(\theta_0)}{n} + o_E(n^{-\frac{3}{2}}) \qquad \ldots \quad (2.32)$$

applying (2.27) to $(T_n - \theta_0)$. When T_n is l.s. (III) we define

$$T_n^* = T_n - b(T_n)/n. \qquad \ldots \quad (2.33)$$

Then, by Proposition 2,

$$E^U(T_n^*) = \theta_0 + o(1/n). \qquad \ldots \quad (2.34)$$

We shall now calculate the covariance of $(T_n^* - \theta_0)I_U$ with Z_n, $Z_n W_n$ and Z_n^2 and show that these covariances are the same up to $o(1/n^2)$.

Lemma 3 : *If Assumption* 1′ *holds and* T_n *is efficient and l.s.* (III) *then*

$$E^U\{(T_n^*(p^n) - \theta_0)\, Z_n\} = \frac{1}{n} + o\left(\frac{1}{n^2}\right). \qquad \ldots \quad (2.35)$$

Proof : (2.34) follows from direct calculations using (2.27a), (2.31), (2.32) and Lemma 1.

Note that we can get (2.35) formally by differentiating (2.34) with respect to θ. The trouble in justifying this is that one has to show that on differentiating the $o(1/n)$ term in (2.34) one would get a term of order $o(1/n)$. The calculations needed for this are no less cumbrous than the direct proof of Lemma 3 given above.

To calculate the covariance with Z_n^2 and $Z_n W_n$ we shall need the following results which are well known and easy to derive. The same formulas occur in Rao (1961) but we shall use them in a different way. Let $(Y_{1i}, Y_{2i}, Y_{3i}, Y_{4i})$ be i.i.d. real vectors with zero expectations. Let $\overline{Y}_j = \sum_1^n Y_{ji}/n$. Then up to $o(1/n^2)$

$$E(\overline{Y}_1^4) = 3[\text{var}(Y_{11})]^2/n^2 \qquad \dots \quad (2.36)$$

$$\text{cov}(\overline{Y}_1^2, \overline{Y}_2\overline{Y}_3) = 2\,\text{cov}(Y_{11}, Y_{21})\,.\,\text{cov}(Y_{11}, Y_{31})/n^2 \qquad \dots \quad (2.37)$$

$$\text{cov}(\overline{Y}_1\overline{Y}_2, \overline{Y}_3\overline{Y}_4) = \{\text{cov}(Y_{11}, Y_{31})\,.\,\text{cov}\,(Y_{21}, Y_{41}) + \text{cov}(Y_{11}, Y_{41}).(Y_{21}, Y_{31})\}/n^2 \qquad \dots \quad (2.38)$$

$$E(\overline{Y}_1^3\overline{Y}_2) = 3\ \text{var}(Y_{11}).\text{cov}(Y_{11}, Y_{21})/n^2 \qquad \dots \quad (2.39)$$

$$E(\overline{Y}_1^2\overline{Y}_2^2) = \{\text{var}(Y_{11}).\text{var}(Y_{21}) + \text{w}(\text{cov}(Y_{11}, Y_{21}))^2\}/n^2. \qquad \dots \quad (2.40)$$

Lemma 4 : *Suppose Assumption 1 holds and T_n is efficient and l.s.*(II). *Let*

$$T_n^{**} = T_n(p^n) - \theta_0 - Z_n/I. \qquad \dots \quad (2.41)$$

Then

$$E^U\{T_n^{**}(Z_n^2 - I/n)\} = \frac{2\mu_{11}+J}{n^2 I} + o\left(\frac{1}{n^2}\right) \qquad \dots \quad (2.42)$$

and

$$E^U\{T_n^{**}(Z_n W_n - \mu_{11}/n)\} = \frac{\mu_{02}}{n^2 I} + \frac{\mu_{11}(J+\mu_{11})}{n^2 I^2} + o\left(\frac{1}{n^2}\right). \qquad \dots \quad (2.43)$$

Proof: $E^U\{T_n^{**}(Z_n^2 - I/n)\}$

$$= E^U\left\{\frac{1}{2}\,\Sigma\,\Sigma T^{ij}(p_i^n - \pi_i(\theta_0))(p_j^n - \pi_j(\theta_0)).(X_n^2 - I/n)\right\} + o\left(\frac{1}{n^2}\right) \quad \text{by (2.27)}$$

$$= E_{\theta_0}\left\{\frac{1}{2}\,\Sigma\,\Sigma T^{ij}(p_i^n - \pi_i(\theta_0))(p_j^n - \pi_j(\theta_0))\,.\,(Z_n^2 - I/n)\right\} + o\left(\frac{1}{n^2}\right) \text{ by Lemma 1}$$

$$= \frac{1}{2}\,\Sigma\,\Sigma T^{ij}\,\text{cov}_{\theta_0}\{Z_n^2, (p_i^n - \pi_i(\theta_i))(p_j^n - \pi_j(\theta_0))\} + o\left(\frac{1}{n^2}\right)$$

$$= \Sigma\,\Sigma T^{ij}\,\text{cov}_{\theta_0}\{Z_n, p_i^n\}\,\text{cov}_{\theta_0}\{Z_n, P_j^n\} + o\left(\frac{1}{n^2}\right) \text{ by (2.37)}$$

$$= \frac{1}{n}\,\Sigma\,\Sigma T^{ij}\,\text{cov}_{\theta_0}\{Z_n, p_i^n\{\pi_j'(\theta_0) + o\left(\frac{1}{n^2}\right)$$

$$= \frac{1}{n}\,\Sigma\,\text{cov}_{\theta_0}(Z_n, p)_i^n \frac{d}{d\theta}\left(\frac{\beta_i'}{I}\right)\Bigg|_{\theta_0} + o\left(\frac{1}{n^2}\right) \qquad \dots \quad (2.44)$$

since differentiating both sides of (2.10) with respect to θ and putting $\theta = \theta_0$, we get

$$\Sigma T^{ij}\pi_j'(\theta_0) = \frac{d}{d\theta}(\beta_i'/I)\bigg|_{\theta_0} = \left\{\frac{\beta_i''(\theta_0)}{I} - \frac{\beta_i'(\theta_0)I'(\theta_j)}{I^2}\right\}. \quad \dots \quad (2.45)$$

As noted earlier $I'(\theta)$, β_i' exists by Assumption 1.

Using (2.13), (2.14), (2.44) and (2.45) we get

$$E^U\{T_n^{**}(Z_n^2 - I/n)\} = \frac{I}{nI}\operatorname{cov}(Z_n, W_n) - \frac{I'}{nI^2}\operatorname{var}(Z_n) + o\left(\frac{1}{n^2}\right)$$

$$= \frac{\mu_{11}}{n^2 I} - \frac{I'}{n^2 I} + o\left(\frac{1}{n^2}\right)$$

$$= \frac{2\mu_{11}+J}{n^2 I} + o\left(\frac{1}{n^2}\right).$$

Since

$$-I'. = J+\mu. \quad \dots \quad (2.46)$$

This completes the proof of (2.42). Proceeding in the same way but using (2.38) in place of (2.37), we get

$$E^U\{T_n^{**}(Z_n W_n - \mu_{11}/n)\}$$

$$= \frac{1}{2}\Sigma\Sigma T^{ij}\operatorname{cov}_{\theta_0}\{Z_n W_n, (p_i^n - \pi_i)(p_j^n - \pi_j)\}$$

$$= \Sigma\Sigma T^{ij}\operatorname{cov}_{\theta_0}\{W_n, p_i^n\}.\operatorname{cov}\{Z_n, p_j^n\}$$

$$= \frac{1}{n}\Sigma\Sigma T^{ij}\operatorname{cov}_{\theta_0}\{W_n, p_i^n\}\pi_j'(\theta_0)$$

$$= \frac{1}{n}\Sigma\operatorname{cov}_{\theta_0}\{W_n, p_i^n\}\cdot\left\{\frac{\beta_i''(\theta_0)}{I} - \frac{\beta_i'(\theta_0).\,I'(\theta_0)}{I^2}\right\} \quad \dots \quad (2.47)$$

$$= \frac{\operatorname{var}(W_n)}{nI} - \frac{I'\operatorname{cov}(W_n, Z_n)}{nI^2}. + o\left(\frac{1}{n^2}\right)$$

$$= \frac{\mu_{02}}{n^2 I} + \frac{\mu_{11}(J+\mu_{11})}{n^2 I^2} + o\left(\frac{1}{n^2}\right)$$

by (2.46). This proves (2.43) and completes the proof of the lemma.

Lemma 5. *Suppose Assumption 1 holds and T_n is efficient and l.s.(III). Let*

$$T_n^{***} = T_n^{*} - \theta_0 - Z_n/I. \qquad \dots \quad (2.48)$$

Then $E^U\{T_n^{***}(Z_n^2 - I/n)\}$ and $E^U\{T_n^{***}(Z_n W_n - \mu_{11}/n)\}$ are given by the right hand sides of (2.42) and (2.43) respectively.

Proof: Lemma 5 follows from Lemma 4 if we note that

(i) $T_n^{**} - T_n^{***} = \dfrac{b(\theta_0)}{n} + o_E \dfrac{1}{n}$ by (2.27), (2.32) and (2.33)

(ii) $Z_n^2 - I/n = O_E\left(\dfrac{1}{n}\right)$ by (2.36)

(iii) $E^U(Z_n^2 - I/n) = E_{\theta_0}(Z_n^2 - I/n) + o\left(\dfrac{1}{n^2}\right)$ by Lemma 1

$= 0(n^{-2})$

(iv) $Z_n W_n - \mu_{11}/n = O_E\left(\dfrac{1}{n}\right)$ by (2.40)

(v) $E^U(Z_n W_n - \mu_{11}/n) = o\left(\dfrac{1}{n^2}\right)$ as in (iii).

Lemmas 4 and 5 are special cases of a more general result which expresses the covariance of $T_n^{**} I_U$ and $T_n^{***} I_U$ with $Z_n.\{\Sigma\alpha_i(p_i^n - \pi_i(\theta_0)\}$ as the covariance of $\{\Sigma\alpha_i(p_i^n - \pi_i(\theta_0))\}$ with $n^{-1}\{W_n I - Z_n I'\}/I^2$; here α_i's are constants. The proof of this more general result is similar to the proof of Lemmas 4 and 5. Another "formal" proof is given in the third remark after Theorem 1.

We are now in a position to prove Theorem 1.

Proof of Theorem 1 (i) and (ii) : We first prove (2.20)

$$0 = n^{-1} \frac{d \log L}{d\theta}\bigg|_{\hat\theta_n} = Z_n + (\hat\theta_n - \theta_0) n^{-1} \frac{d^2 \log L}{d\theta^2}\bigg|_{\theta_0}$$

$$+ \frac{1}{2} (\hat\theta_n - \theta_0)^2 \cdot n^{-1} \frac{d^3 \log L}{d\theta^3}\bigg|_{\theta_0} + o_E\left(\frac{1}{n}\right)$$

$$= Z_n + (\hat\theta_n - \theta_0)(W_n - I) + \frac{1}{2} (\hat\theta_n - \theta_0)^2 J + o_E\left(\frac{1}{n}\right).$$

$$\text{Hence } (\hat{\theta}_n-\theta_0) = \frac{Z_n}{I}+\frac{(\hat{\theta}_n-\theta_0)W_n}{I}+\frac{(\hat{\theta}_n-\theta_0)^2J}{2I}+o_E\left(\frac{1}{n}\right)$$

$$= \frac{Z_n}{I}+\frac{Z_nW_n}{I^2}+\frac{JZ_n^2}{2I^3}+o_E\left(\frac{1}{n}\right). \quad (2.49)$$

Since $\hat{\theta}_n-\theta_0 = Z_n/I+O_E(1/n)$ by Lemma 2.

We get from (2.49),

$$b_0(\theta_0) = \mu_{11}/I^2+J.I \quad \ldots \quad (2.50a)$$

$$(\theta_n-\theta_0)-\frac{b_0(\theta_0)}{n} = \frac{Z_n}{I}+\frac{(Z_nW_n-\mu_{11}/n)}{I^2}+\frac{J(Z_n^2-I/n)}{2I^3}+o_E\left(\frac{1}{n}\right) \quad \ldots \quad (2.50b)$$

$$= \frac{Z_n}{I}+S_n+o_E\left(\frac{1}{n}\right) \quad \ldots \quad (2.51)$$

which is (2.20).

We next prove (2.22). This is the crucial step. Suppose T_n is efficient and l.s. (II). Then by Lemma 4, the covariance of $T_n^{**}I_U$ and hence of $\left\{T_n^{**}-\frac{b(\theta_0)}{n}\right\}I_U$ with Z_n^2 and Z_nW_n does not depend on T_n. (In fact to prove this one needs only (2.44) and (2.47) rather than the more explicit formulas (2.42) and (2.43)). Hence we may write the regression equation

$$\left\{T_n^{**}-\frac{b(\theta_0)}{n}\right\}I_U = E_{\theta_0}\left[\left\{T_n^{**}-\frac{b(\theta_0)}{n}\right\}I_U\right]+\alpha_n(Z_n^2-I/n)$$
$$+\beta_n(Z_nW_n-\mu_{11}/n)+\eta_n \quad \ldots \quad (2.52)$$

where α_n, β_n do not depend on T_n, η_n has zero covariance with Z_n^2 and Z_nW_n up to $o\left(\frac{1}{n^2}\right)$. By Lemma 1 and Proposition 2(i) the first term on the RHS of (2.52) is $o\left(\frac{1}{n}\right)$. We may therefore lump this term and η_n and rewrite (2.52) as

$$\left\{T_n^{**}-\frac{b(\theta_0)}{n}\right\}I_U = \alpha_n(Z_n^2-I/n)+\beta_n(Z_nW_n-\mu_{11}/n)+\eta_n' \quad \ldots \quad (2.53)$$

where η_n' is orthogonal to Z_n^2 and Z_nW_n up to $o\left(\frac{1}{n^2}\right)$. We may write (2.53) as

$$T_n-\theta_0-\frac{Z_n}{I}-\frac{b(\theta_0)}{n} = \alpha_n(Z_n^2-I/n)+\beta_n(Z_nW_n-\mu_{11}/n)+\eta_n'' \quad \ldots \quad (2.54)$$

where $\eta_n' = \eta_n^* I_U$ and so η_n' is E^U-orthogonal to $(Z_n^2 - I/n)$ and $(Z_n W_n - \mu_{11}/n)$ up to $o\left(\frac{1}{n^2}\right)$. We can calculate α_n, β_n directly but it may be illuminating to get it in an indirect but somewhat easier method. Since $\hat{\theta}_n$ is efficient and l.s.(II) (in fact l.s.(III)) we get, on comparing (2.50b) and (2.54) that

$$\alpha_n = J/2I^3, \quad \beta_n = \frac{1}{I^2}. \qquad \dots \quad (2.55)$$

From (2.54) and (2.55) we get

$$T_n - \theta_0 = b(\theta_0)/n + S_n + R_n \qquad \dots \quad (2.56)$$

where R_n is E^U-orthogonal to Z_n^2, $Z_n W_n$ up to $o\left(\frac{1}{n^2}\right)$.

Since

$$R_n = (T_n - \theta_0) - S_n - b(\theta_0)/n \qquad \dots \quad (2.57)$$

it follows from Lemma 2, applied to T_n that R_n is $O_E\left(\frac{1}{n}\right)$. That $E^U(R_n) = o\left(\frac{1}{n}\right)$ follows from (2.57) and the definition of $b(\theta_0)$ and S_n. Thus R_n satisfies all the conditions stated in Theorem 1, completing the proof of (2.22).

For (2.23), recall that

$$E_2(\{T_n\}, \theta_0, \lambda) = \lim n^2 E^U \{Z_n - (T_n - \theta_0) I - \lambda (T_n - \theta_0)^2 - a_\lambda(\theta_0)/n\}^2$$

where $a_\lambda(\theta_0)$ is defined in Proposition 1. So

$$E_2(\{T_n\}, \theta, \lambda) = \lim n^2 E^U \left\{Z_n - \left(T_n - \theta_0 - \frac{b(\theta_0)}{n}\right) I - \frac{\lambda}{I^2}(Z_n^2 - I/n)\right\}^2$$

since $$T_n - \theta_0 = \frac{Z_n}{I} + o_E\left(\frac{1}{n}\right)$$

by Lemma 2 and $a_\lambda(\theta_0) = -b(\theta_0).\, I - \lambda/I$ from definition of a_λ and b. Using (2.54) and (2.55) we get now

$$E_2(\{T_n\}, \hat{\theta}_0, \lambda) = \lim n^2 E^U \left\{IS_n + \frac{\lambda}{I^2}(Z_n^2 - I/n)\right\}^2 + \lim n^2 E^U \left\{IR_n\right\}^2$$

since R_n is E^U-orthogonal, up to $o\left(\frac{1}{n^2}\right)$, to S_n and Z_n^2. So, finally using (2.20),

$$E_2(\{T_n\}, \theta_0, \lambda) = E_2(\{\hat{\theta}_n\}, \theta_0, \lambda) + \lim n^2 E^U (I R_n)^2$$

and $$E_2(\{T_n\}, \theta_0) \geqslant E_2(\{\hat{\theta}_n\}, \theta_0)$$

with equality iff $E^U(R_n^2) = o\left(\frac{1}{n^2}\right)$. This completes the proof of (2.23).

We now derive (2.21).

By (2.20), (2.32) and (2.33) we may write

$$\theta_n^*-\theta_0-Z_n/I = \frac{J}{2I^3}\left\{Z_n^2-\frac{I}{n}-\frac{\mu_{30}Z_n}{n\mu_{20}}\right\}+I^2\left\{Z_nW_n-\frac{\mu_{11}}{n}-\frac{\mu_{21}Z_n}{n\mu_{20}}\right\}+\hat{R}_n^* \quad \ldots \quad (2.58)$$

where
$$\hat{R}_n^* = \hat{R}_n+\frac{Z_n}{n}\left\{\frac{-b_0'(\theta_0)}{I}+J\mu_{30}/2I^4+I\mu_{21}\right\}+o_E(n^{-3/2})$$

$$= \hat{R}_n+o_E\left(\frac{1}{n}\right)$$

and so E^U-orthogonal to Z_n^2 and Z_nW_n up to $o\left(\frac{1}{n^2}\right)$. By Lemma 3 the LHS of (2.58) is E^U-orthogonal to Z_n' up to $o(n^{-2})$. By easy direct computation the same result is true of the first two terms on the RHS of (2.58). Hence $\hat{R}_n^*$ is also E^U-orthogonal to Z_n up to $o\left(\frac{1}{n^2}\right)$. So R_n^* has the properties asserted in Theorem 1 and (2.21) is nothing but (2.58) with some rearrangement of terms. This completes the proof (2.21).

By Lemmas 3 and 5 $(T_n^*\theta_0)I^U$ has same covariance up to $o(n^{-2})$ with Z_n, Z_nW_n and Z_n^2 as $(\hat{\theta}_n^*-\theta_0)I^U$. So (2.24) can be deduced from (2.21) in the same way as (2.22) was deduced from (2.20). It is easy to check that $R_n-R_n^*$ $= o_E\left(\frac{1}{n}\right)$.

Also, as in (i), using (2.21) and (2.24),

$$\psi(\{T_n^*\},\ \theta_0) = \psi(\{\hat{\theta}_n^*\},\ \theta_0)+\lim n^2E^U(R_n^*)^2 \quad \ldots \quad (2.59)$$

proving (2.25).

We next prove (2.26). From (2.59)

$$\psi(\{T_n^*\},\ \theta_0) = \psi(\{\hat{\theta}_n^*\},\theta_0)+\lim n^2E^U(R_n)^2. \quad \ldots \quad (2.60)$$

Since $R_n-R_n^*$ is $o_E\left(\frac{1}{n}\right)$ and $R_n = O_E\left(\frac{1}{n}\right)$. Clearly

$$\psi(\{\hat{\theta}_n^*\},\theta_0) = \lim n^2E^U(S_n^2) = \frac{1}{I^4}\{I\mu_{02}-\mu_{11}^2\}+\frac{2}{I^4}\left\{\frac{J}{2}+\mu_{11}\right\}^2. \quad \ldots \quad (2.61)$$

Also, from the proof of (2.23)

$$E_2(\{T_n\},\theta_0) = \lim n^2I^2E^U(R_n^2)+\lim n^2I^2E^U\left\{Z_nW_n-\mu_{11}/n)\frac{1}{I^2}-\gamma(Z_n^2-I/n)\right\}^2$$

where γ is the limiting regression coefficient of $I^{\frac{1}{2}}(Z_nW_n-\mu_{11}/n)$ on (Z_n^2-I/n) and is found to be μ_{11}/I^2 applying (2.36) and (2.39). So

$$E_2(\{T_n\},\theta_0) = \lim n^2I^2E\ (R_n^2)+\frac{1}{I^2}\{I.\mu_{02}-\mu_{11}^2\} \qquad \dots \quad (2.62)$$

using (2.40).

So by (2.60), (2.61) and (2.62)

$$E_2(\{T_n\},\theta_0) = I^2\psi(\{T_n\},\theta_0)-\frac{2}{I^2}\left\{\frac{J}{2}+\mu_{11}\right\}^2$$

proving (2.26).

(iii) Let T_n be efficient and l.s. (III). Let m and T_n' be as in the statement of Theorem 1 (iii). Let

$$\hat{\theta}_n' = \hat{\theta}_n-\{b_0(\hat{\theta}_n)-b(\hat{\theta}_n)-m(\hat{\theta}_n)\}/n$$

where $b(\theta_0)$ is defined in Proposition 2.

Then as in the proof of Lemma 3 it can be shown that $\hat{\theta}_n'$ and T_n' have the same covariance with Z_n up to $o\left(\frac{1}{n^2}\right)$. Since

$$[\hat{\theta}_n-\{b_0(\theta)-b(\theta)-m(\theta)\}/n-\hat{\theta}_n'] \text{ and } [T_n+m(\theta)/n-T_n']$$

are $o_B\left(\frac{1}{n}\right)$, we can apply Lemma 4 to conclude that $\hat{\theta}_n'$ and T_n' have the same covariance with Z_nW_n and Z_n^2 up to $o\left(\frac{1}{n^2}\right)$. It follows as in the proof of (2.25) that $\psi(\{T_n'\},\theta_0) \geqslant \psi(\{\hat{\theta}_n'\},\theta_0)$ which leads to the desired conclusion by Proposition 2. This completes the proof of Theorem 1.

Note that the main difference between (2.20) and (2.22) is that $\hat{R}_n$ is $o(n^{-1})$ whereas R_n is only $O(n^{-1})$. This is at the root of a result like (2.23). The main difference between $\hat{R}_n$ and $\hat{R}_n^*$ is that $\hat{R}_n^*$ is E^U-orthogonal to Z_n up to $o(n^{-2})$ but $\hat{R}_n$ is not. A similar remark applies to R_n and R_n^*. The significance of (2.24) for proving (2.25) should be self evident.

The expansions obtained above are not affected by the singularity of the dispersion matrix of Z_n, Z_nW_n, Z_n^2 up to $o(n^{-2})$ but it is worth pointing out that singularity up to $o(n^{-2})$ obtains iff there is a linear relation between Z_nW_n and Z_n^2 up to $o(n^{-1})$, which can be true iff there is a linear relation between Z_1 and W_1. If such a relation holds for all θ_0 in some open set then f_θ is essentially a one-dimensional exponential family. For, the hypothesis of linear relation between Z_1 and W_1 taken together with the linear independence of $p_1^1, \dots, p_k^1$ implies

$$\beta_i''(\theta)/\beta_i'(\theta) = g(\theta)$$

the solution of which can be written in the form $\beta_i(\theta) = d_i\beta_1(\theta)+c_i$ where d_i, c_i are constants. Hence $\log f_\theta = c(\theta)+\beta_1(\theta)\Sigma d_i p_i^1+\Sigma_i^k c_i$. Of course if f_θ is an one-dimensional exponential density then there is a linear relation between Z_1 and W_1.

Remarks (1) : Analogous results hold if there are two or more parameters to be estimated. Suppose θ is a vector with two real coordinates (θ^1, θ^2) and $T_n = (T_n^1, T_n^2)$ is an efficient estimator. Let

$$\Phi_1(\lambda, \{T_n\}, \theta_0) = n\left\{\frac{1}{n}\frac{\partial \log L}{\partial \theta^1}\bigg|_{\theta_0} -(T_n^1-\theta_0^1)I_{11}-(T_n^2-\theta_0^2)I_{12} -\Sigma\lambda_{ij}(T_n^i-\theta_0^i)(T_n^j-\theta_0^j)\right\}$$

$$\Phi_2(\gamma, \{T_n\}, \theta_0) = n\left\{\frac{1}{n}\frac{\partial \log L}{\partial \theta^2}\bigg|_{\theta_0} -(T_n^1-\theta_0^1)I_{12}-(T_n^2-\theta_0^2)I_{22} -\Sigma\gamma_{ij}(T_n^i-\theta_0^i)(T_n^j-\theta_0^j)\right\}.$$

where $[I_{ij}]$ is the 2×2 information matrix. Then ignoring as before contributions outside an open set around $\pi(\theta_0)$, it can be shown that the difference of the limiting dispersion matrix of $\Phi_1(\lambda, \{T_n\}, \theta_0)$, $\Phi_2(\gamma, \{T_n\}, \theta_0)$ and that of $\Phi_1(\lambda, \hat\theta_n, \theta_0)$, $\Phi_2(\gamma, \hat\theta_n, \theta_0)$ is positive semidefinite if T_n is l.s.(II) and efficient. The proof is exactly similar to the one-parameter case. A similar extension of Rao's theorem is also possible if one considers the expansion for the dispersion matrix of T_n^* and looks at the coefficient matrix of $1/n^2$. If one denotes this coefficient matrix by $\psi(\{T_n^*\}, \theta_0)$ one can prove $\psi(\{T_n^*\}, \theta_0)-\psi(\{\hat\theta_n^*\}, \theta_0)$ is positive semidefinite for l.s. (III) efficient estimators.

(2). $E_2(\{\hat\theta_n\}, \theta_0)=\frac{1}{I^2}(I\mu_{02}-\mu_{11}^2) = 0$ iff there is a linear relation between Z_1 and W_1. We have seen earlier that if this result holds for all θ_0 in some open set, then f_θ is essentially a 1-dimensional exponential density.

$$\psi(\{\hat\theta_n\}, \theta_0) = 0$$

if there is a linear relation between Z_1 and W_1 and, moreover, $\rho+2\mu_{11} = 0$.

(3). It may be illuminating to give a "formal" proof of the following result

$$I.E^U(T_n^*Z_nM_n) = \frac{1}{n}E^U(H_nM_n) \qquad \dots \quad (2.63)$$

where $M_n = \Sigma\alpha_i(p_i^n-\pi_i(\theta_0))$ and

$$H_n = (W_n-Z_nI')/I.$$

Now

$$\frac{1}{n}E^U\{(IT_n^*)M_n\} = o\left(\frac{1}{n^2}\right).$$

Hence differentiating this we get formally,

$$E^U\{(I\,T_n^*)Z_nM_n\} = -\frac{1}{n}\,E^U\left\{\frac{d}{d\theta}(IT_n^*)M_n\right\}$$

$$+\frac{1}{n}\,E^U(T_n^*)\Sigma\alpha_i\pi_i' + o\left(\frac{1}{n^2}\right)$$

$$= -\frac{1}{n}\,E^U\left\{\frac{d}{d\theta}(I\,T_n^*)M_n\right\} + o\left(\frac{1}{n^2}\right)$$

$$= \frac{1}{n}\,E^U(H_nM_n)$$

since $H_n + IT_n^* = o_B(1/\sqrt{n})$.

As pointed out earlier, Lemmas 3 and 4 are special cases of (2.63).

(4) The smaller the measures E_2 and ψ the better is the estimator, from the point of view of second order efficiency. They are really measures of deficiency as defined in a more general context by Hodges and Lehmann (1970). Deficiency of $\{T_n\}$ relative to $\{\hat\theta_n\}$ is $[\psi(\{T_n^*\}, \theta_0) - \psi(\{\hat\theta_n^*\}, \theta_0)]/I$.

(5) To facilitate comparison with Rao's (1961, 1963) results for the multinomial, let us denote by m_{ij} what Rao (1963) denotes as μ_{ij}. We shall show how to express m_{ij} in terms of our μ_{ij} and vice versa. Let the multinomial population consist of $k+1$ classes with probabilities $\pi_1(\theta), \ldots, \pi_{k+1}(\theta)$. Let

$U_i = 1$ if first observation falls in the i-th class

$= 0$ otherwise.

Hence $$Z_1 = \Sigma\left(\frac{\pi_i'}{\pi_i}\right)U_i,\quad W_1 = \Sigma\,\frac{d^2\log\pi_i}{d\theta^2}U_i + I.$$

and μ_{rs} was defined as $E_{\theta_0}(Z_1^rW_1^s)$. Let $Y = \Sigma\left(\frac{\pi_i''}{\pi_i}\right)U_i$. Then, following Rao (1963), $m_{ij} = E_{\theta_0}(Z_1^iY^j)$. Clearly

$$W_1 = Y - (Z_1)^2 + I.$$

Using this we can express m_{ij} in terms of μ_{ij} and vice versa. For example, $\mu_{11} = m_{11} - m_{30}$.

(6) Let us try to understand the calculations of second order efficiency by Fisher (1925). As Rao (1961, equation 5.14) has pointed out Fisher's

measure $E_2' = \lim\ (nI - I_{T_n})$ can be shown to equal the limit of the expectation of the conditional variance of nZ_n given T_n. Consider the expansion

$$Y_n = Z_n + (T_n - \theta_0)\ \frac{1}{n}\ \frac{d^2 \log L}{d\theta^2}\Bigg|_{\theta_0}$$

where
$$Y_n = \frac{1}{n}\frac{d \log L}{d\theta}\Bigg|_{T_n}$$

and T_n is an efficient l.s. (II) estimator. This expansion is not correct to $O(1/n)$ but the missing term of $O(1/n)$ is a function of T_n and its neglect does not cause any error in the calculation of conditional variance. Thus, a correct expansion is

$$Z_n = Y_n - \frac{Z_n W_n}{I} + (T_n - \theta_0)I + (T_n - \theta_0)^2 J/2 + o_E(n^{-1}). \quad \ldots \quad (2.64)$$

Fisher now takes the conditional expectation using the joint asymptotic normal distribution of the p_i^n's and replaces the condition "$T_n - \theta_0 =$ constant" by "$Z_n =$ constant". Let us note that these calculations lead to Rao's measure E_2. For, the "conditional expectation" of $Y_n - Z_n W_n/I$ is easily seen to be of the form $\lambda_0 Z_n^2 + c/n$ since by (2.64) and Lemma 2 $Y_n - Z_n$ can be written in the form $\Sigma \alpha_{ij}(p_i^n - \pi_i)(p_j^n - \pi_j)$. Also by (2.64)

$$(T_n - \theta_0) = \frac{Z_n}{I} - \frac{Y_n}{I} + \frac{Z_n W_n}{I^2} + \frac{Z_n^2 . J}{2I^2} + o_E(n^{-1})$$

so that by Theorem 1,

$$\frac{Y_n}{I} = -R_n + o_E(n^{-1}). \quad \ldots \quad (2.65)$$

Evaluating λ_0 and c one finds, the "conditional variance" of Z_n is the "expectation" of

$$\left\{Y_n - \frac{Z_n W_n}{I} - \lambda_0 Z_n^2 - c/n\right\}^2 \quad \text{which equals}$$

$$E^U\left[I^2\left\{-R_n - \frac{Z_n W_n}{I^2} - \lambda_0 Z_n^2 - c/n\right\}^2\right]$$

$$= E_2(\{T_n\}, \theta_0).$$

Thus the measure that Fisher calculates is exactly the measure E_2 of Rao.

Fisher seems to believe, wrongly as it turns out, that Y_n is independent of $(T_n - \theta_0)$ or Z_n up to second order terms. But it is true that the Y_n is E^U-orthogonal to Z_n^2 and $Z_n W_n$ up to $o(n^{-2})$. Thus

$$\text{Fisher's measure} = E_2 = \lim E^U(Y_n^2) + E_2(\{\hat{\theta}_n\}, \theta_0).$$

Fisher gets a wrong result for the minimum chi-square method because he substitutes the variance of

$$Y_n' = \frac{1}{2}\Sigma \frac{(p_i^n - \pi_i(\theta_0))^2}{\pi_i^2(\theta_0)} \pi_i'(\theta_0)$$

for the variance of

$$Y_n = \frac{1}{2}\Sigma \frac{(p_i^n - \pi_i(T_n))^2}{\pi_i^2(T_n)} \pi_i'(T_n).$$

(7). Suppose that instead of the criterion E_2 one considers

$$E_2' = \inf_{\lambda_1, \lambda_2} \lim n^2 E^U\{Z_n - (T_n - \theta_n)\lambda_1 - (T_n - \theta_0)^2 \cdot \lambda_2 - b_{\lambda_1, \lambda_2}(\theta_0)/n\}^2$$

where $$b_{\lambda_1, \lambda_2}(\theta_0) = \lim nE^U(Z_n - (T_n - \theta_0)\lambda_1 - (T_n - \theta_0)^2\lambda_2).$$

For E_2' we have the following result.

Suppose Assumption 1 holds and T_n is efficient and l.s.(II). Let $\hat{\theta}_n'$ be defined as in Theorem 1 (iii) with $m \equiv 0$. Then

$$E_2'(T_n, \theta_0) \geqslant E_2'(\hat{\theta}_n, \theta_0). \qquad \dots \quad (2.66)$$

The proof is similar to that of Theorem 1(iii).

We conclude this section by developing what may be called an asymptotic Bhattacharya bound for efficient estimators.

Suppose T_n is an estimator such that

$$E^U(T_n) = \theta + o\left(\frac{1}{n}\right) \qquad \dots \quad (2.67)$$

then

$$E^U\left(T_n \cdot \frac{1}{n^2}\frac{1}{L}\frac{d^2L}{d\theta^2}\right) = o\left(\frac{1}{n^2}\right). \qquad \dots \quad (2.68)$$

If T_n is efficient we may expect

$$E^U\{(T_n - \theta_0 - Z_n/I). W_n/n\} = o(1/n^2) \qquad \dots \quad (2.69)$$

For convenience let us say that T_n is regular if T_n is efficient and (2.67), (2.68) and (2.69) hold. If T_n is efficient and l.s.(III), then it can be shown easily using Lemma 3 that T_n is regular, since

$$\frac{1}{n^2L}\frac{d^2L}{d\theta^2} = W_n/n + (Z_n^2 - I/n). \qquad \dots \quad (2.70)$$

If T_n is regular then by (2.68), (2.69) and (2.70), $E^U\{(T_n-\theta_0-Z_n/I)(Z_n^2-I/n)\}$ can be shown to have the value given in lemma 3. Using the regression of $T_n I_U$ on Z_n and $\frac{d^2L}{d\theta^2}\times\frac{1}{n^2}$ we get

$$E^U(T_n-\theta_0)^2 \geqslant \frac{1}{nI}+\frac{1}{n^2I^4}\left\{\frac{J}{2}+\mu_{11}\right\} \quad \dots \quad (2.71)$$

It is easy to show that θ_n^* attains this bound iff there is a linear relation between Z_1 and W_1. The implication of this last relation has been discussed earlier.

3. A problem of Berkson

Suppose the dose d_i, $i = 1, \dots, k$, is given to n_i animals and $x_{ij} = 1$ or 0 according as the j-th animal dies or not $j = 1, \dots, n_i$. It is assumed that the probability of death at dose d_i is π_i where π_i lies on the logistic curve

$$\pi_i(\alpha, \beta) = \frac{1}{1+e^{-(\alpha+\beta d_i)}} \,. \quad \dots \quad (3.1)$$

To simplify calculations we assume the n_i's are all equal and β is known. The parameter to be estimated is α. Let $n = n_1+\dots+n_k$ and $p_i^n = \#\, x_{ij}$'s which are equal to one. Then the likelihood function is

$$L = \prod_{i=1}^{k} \pi_i^{np_i^n/k}(1-\pi_i)^{n(1-p_i^n)/k} .$$

The likelihood equation is

$$0 = \frac{d \log L}{d\alpha} = \sum_{i=1}^{k} \frac{n}{k}(p_i^n-\pi_i) \quad \dots \quad (3.2)$$

Let $\hat{\alpha}_n$ be the maximum likelihood estimator.

Let $L_i = \log \{\pi_i/(1-\pi_i)\} = \alpha+\beta\, di$ and $l_i^n = \log \{p_i^n/(1-p_i^n)\}$. Minimizing $\Sigma n_i p_i^n(1-p_i^n)(l_i^n-L_i))^2$ with respect to α one gets the minimum logit chi-square estimator T_n.T_n is to be found from

$$\Sigma n_i p_i^n(1-p_i^n)(l_i^n-L_i^*) = 0 \quad \dots \quad (3.3)$$

where $L_i^* = T_n+\beta d_i$.

This is a special case of the problem considered in Section 2. Here Assumption 1 holds and both T_n and $\hat{\alpha}_n$ are efficient and l.s.(III). The expansions for these estimators become, after some simplifications

$$\hat{\alpha}_n-\alpha_0-\frac{Z_n}{I} = \frac{\Sigma\pi_i(1-\pi_i)(2\pi_i-1)}{2I^3}Z_n^2+o_E(1/n) \quad \dots \quad (3.4)$$

$$\hat{T}_n-\alpha_0-\frac{Z_n}{I} = \Sigma(2\pi_i-1)(p_i^n-\pi_i)Z_n/I^2$$

$$-\frac{1}{2I}\Sigma\frac{(2\pi_i-1)}{\pi_i(1-\pi_i)}(p_i^n-\pi_i)^2+o_E(1/n) \quad \dots \quad (3.5)$$

where $I = \Sigma P_i\theta_i$ and $Z_n = \frac{1}{n}\frac{d\log L}{dx}$. If we consider the corresponding estimator $\hat{\alpha}_n^*$ and T_n^* which are corrected for bias upto $O(1/n)$, we get

$$E^U(\hat{\alpha}_n^*-\alpha_0)^2 = \frac{1}{nI}+\frac{[\Sigma\pi_i(1-\pi_i)(2\pi_i-1)]^2}{2n^2I^4}+o\left(\frac{1}{n^2}\right) \quad \dots \quad (3.6)$$

$$E^U(T_n^*-\alpha_0)^2 = \frac{1}{nI}+\frac{[\Sigma\pi_i(1-\pi_i)(2\pi_i-1)]^2}{n^2I^4}$$

$$+\frac{\Sigma(2\pi_i-1)^2}{2n^2I^2}-\frac{\Sigma\pi_i(1-\pi_i)(2\pi_i-1)^2}{n^2I^3}+o\left(\frac{1}{n^2}\right). \quad \dots \quad (3.7)$$

It follows from the theorem of Section 2 or can be checked directly that

$$E^U(T_n^*-\alpha_0)^2-E^U(\hat{\alpha}_n^*-\alpha_0)^2 = E^U(T_n^*-\hat{\alpha}_n^*)^2+o\left(\frac{1}{n^2}\right).$$

Here one has a complete sufficient statistic, namely Σp_i^n, but T_n^* is not a function of it. If one considers the so-called Rao-Blackwellized $T_n' = E(T_n^*|\Sigma p_i^n)$ then it is indistinguishable from $\hat{\alpha}_n$ up to $O_E(1/n)$.

Our second order expansions seemed to agree quite well with the Monte-Carlo values in a few examples of Berkson that we studied. In the examples T_n had lower bias as well as lower variance than $\hat{\alpha}_n$ but $b'(\alpha_0)$ for T_n was also smaller than the corresponding quantity $b_0'(\alpha_0)$ for $\hat{\alpha}_n$. This last fact explains why $\hat{\alpha}_n^*$ performs better than T_n^*, since

$$E^U(T_n^*-\alpha_0)^2 = E^U(T_n-\alpha_0-b(\alpha_0)/n)^2-\frac{2b'(\alpha_0)}{n^2I}+o\left(\frac{1}{n^2}\right).$$

Silverstone (1957) and Rao (1960) have defended the use of the maximum likelihood estimators from certain other points of view.

4. Bayesian approach to second order efficiency

In an important pioneering paper on Bayesian analysis, Lindley (1961) has considered an expansion for the a posteriori risk and obtained from it an expansion for a Bayes estimator in powers of $(1/n)$. In the discussion following Rao (1962) he seeks a Bayesian justification of Rao's (1962) results. Lindley considers a loss function, depending on observations, which is proportional to

$$l(d, \theta) = \left(Z_n - \frac{1}{n}\frac{d \log L}{d\theta}\bigg|_d\right)^2 \qquad \dots \quad (4.1)$$

and a uniform prior measure. Actually his terminology is slightly different. He considers the product of prior and loss function and calls it a weight function. Lindley shows that $\hat{\theta}_n$ is Bayes upto $o(1/n)$ for this prior and loss function. He claims that the loss function (4.1) is equivalent to the measure E_2 of Rao and that the Bayes property of $\hat{\theta}_n$ explains its second order efficiency. It seems to us that both these claims are unjustified.

Consider, for example, the special case of i.i.d. $N(\theta, 1)$ variables and note here (4.1) reduces to $(d-\theta)^2$ Then, presumably, one would evaluate, say an efficient estimator T_n by calculating $E_{\theta_0}\{T_n-\theta_0\}^2$ if the loss function (4.1) were used. This seems to have no relation with

$$E_0\{(\bar{X}_n-\theta_0)-(T_n-\theta_0)-\lambda(T_n-\theta_0)^2-a_\lambda(\theta_0)/n\}^2$$

which one would have to consider for Rao's measure. Moreover for the loss function (4.1) the Bayes property of $\hat{\theta}_n$ does not imply that for every efficient estimator T_n

$$E_{\theta_0}(T_n-\theta_n)^2 \geqslant E_{\theta_0}(\hat{\theta}_n-\theta_0)^2+o\left(\frac{1}{n^2}\right) \ \forall\ \theta_0. \qquad \dots \quad (4.2)$$

In fact it is easy to see (4.2) is false. Note that $\hat{\theta}_n = \bar{x}_n$, the sample mean and so if we take $T_n = \bar{x}_n+b(\bar{x}_n)/n$ where $b(\theta_0)$ and $b'(\theta_0) < 0$ then (4.2) is violated. If the prior is the Lebesgue measure then the approximate Bayes property for $\hat{\theta}_n$ becomes an exact one in the sense $\int[(d-\theta)^2 f_\theta(x_1) \dots f_\theta(x_n)]d\theta$ is minimized at $d = \hat{\theta}_n$. This result is known to be at the root of minimaxity and admissibility of $\hat{\theta}_n$ with respect to the loss (4.1) but it cannot imply any uniformly best property like (4.2).

The remarks regarding Lindley's loss for the special case considered above are true for the general problem with slight modification but we shall not pursue this matter further. Let us now proceed to show that a Bayesian proof of results on second order efficiency is indeed possible though not on the lines of Lindley outlined above. Our arrangements will be heuristic but we hope to provide a rigorous account at a later date.

We first approach Rao's result. Let $\{x_i\}$ be i.i.d. with density $f_\theta(x)$ and the loss function be $(d-\theta)^2$. Let the prior have a density $q(\theta)$ with respect to the Lebesgue measure and suppose $q(\theta)$ is twice continuously differentiable and positive everywhere. Then the Bayes solution is, using Lindley (1961),

$$B_n = \hat\theta_n + \frac{1}{2}\frac{L_3}{L_2^2} - \frac{1}{L_2}\,\frac{q'(\hat\theta_n)}{q(\hat\theta_n)} + o(1/n^{3/2}) \quad \dots \quad (4.3)$$

where $$L_i = \left.\frac{d^i \log L}{d\theta^i}\right|_{\hat\theta_n}.$$

Considering $$B_n' = \hat\theta_n + \frac{1}{2}\frac{L_3}{L_2^2} - \frac{1}{L_2}\,\frac{q'(\hat\theta_n)}{q(\hat\theta_n)} \quad \dots \quad (4.4)$$

as an approximate Bayes estimator we note that $E_\theta(B_n'-\theta)^2 = E_\theta(B_n-\theta)^2 + o(1/n^2)$.

From (4.4)

$$E_\theta(B_n') = \theta + \frac{b_0(\theta)}{n} + \frac{c(\theta)}{n} + o(1/n) \quad \dots \quad (4.5)$$

where $$E_\theta(\hat\theta_n) = \theta + b_0(\theta)/n + o(1/n) \quad \dots \quad (4.6)$$

and $$E_\theta(B_n' - \hat\theta_n) = c(\theta)/n + o(1/n). \quad \dots \quad (4.7)$$

We assume $c(\theta)$ to be continuously differentiable and consider

$$B_n^* = \hat\theta_n + c(\hat\theta_n)/n. \quad \dots \quad (4.8)$$

Then $$E_\theta(B_n^* - \hat\theta_n) = c(\theta)/n + o(1/n). \quad \dots \quad (4.9)$$

Clearly $B_n^* = B_n + o(1/n)$. We show this implies that B_n^* is Bayes upto $O(1/n^2)$ in the sense of (4.14) below. Now,

$$E_\theta(B_n'-\theta)^2 = E_\theta(\hat\theta_n-\theta)^2 + \frac{\{c(\theta)\}^2}{n^2} + 2c(\theta)E_\theta(\hat\theta_n-\theta)$$
$$+ 2E_\theta(\hat\theta_n-\theta)(B_n'-\hat\theta_n-c(\theta)/n) + o(1/n^2). \quad \dots \quad (4.10)$$

Similarly

$$E_\theta(B_n^*-\theta)^2 = E_\theta(\hat\theta_n-\theta)^2 + \frac{\{c(\theta)\}^2}{n^2} + 2c(\theta)E_\theta(\hat\theta_n)-\theta)$$
$$+ 2E_\theta(\hat\theta_n-\theta)(B_n^*-\hat\theta_n-c(\theta)/n) + o(1/n^2) \quad \dots \quad (4.11)$$

But

$$E_\theta(\hat{\theta}_n-\theta)\{(B'_n-\hat{\theta}_n-c(\theta)/n\}-E_\theta(\hat{\theta}_n-\theta)\{(B^*_n-\hat{\theta}_n-c(\theta)/n\}$$
$$= E_\theta\{Z_n/I)(B'_n-B^*_n)\}+o(1/n^2), \quad \ldots \quad (4.12)$$
$$= o(n^2)$$

since $B'_n-B^*_n = O(n^{-3/2})$ and $\hat{\theta}_n-\theta-Z_n/I = O(n^{-1})$ which is obtained by differentiating the relation

$$E_\theta(B'_n-B^*_n) = o(1/n). \quad \ldots \quad (4.13)$$

It follows now from (4.10), (4.11) and (4.12) that

$$E_\theta(B^*_n-\theta)^2 = E_\theta(B'_n-\theta)^2+o(1/n^2) = E_\theta(B_n-\theta)^2+o(1/n^2). \quad \ldots \quad (4.14)$$

We have arrived at a remarkable fact. From (2.18) of Lindley (1961) we notice that up to $O(1/n)$ the posterior depends on $\hat{\theta}_n$, L_2, L_3 and L_4; in a sense, therefore, they are sufficient to $O(1/n)$. (For the concept of sufficiency to $O(1/n)$ in a different sense see the next section.) Nevertheless (4.14) shows that for the loss function $(d-\theta)^2$ and all smooth priors B^*_n is a Bayes solution to the degree of accuracy specified in (4.14). Thus $\hat{\theta}_n$ alone is not sufficient to $O(1/n)$ but this Bayes solution B^*_n is a function $\hat{\theta}_n$ alone.[1] Incidentally, the Bayes property (4.14) would hold for any $T_n = B_n+o(1/n)$.

Consider now any efficient estimator T_n, such that

$$E_\theta(T_n) = \theta+b(\theta)/n \quad \ldots \quad (4.15)$$

Let
$$T'_n = T_n-\frac{b(T_n)+b_0(T_n)+c(T_n)}{n}. \quad \ldots \quad (4.16)$$

Then
$$E_\theta(T'_n-B_n) = O(1/n).$$

Hence

$$E_\theta(T'_n-\theta)^2-E_\theta(B^*_n-\theta)^2 = E_\theta(T^*_n-\theta)^2-E_\theta(\hat{\theta}^*_n-\theta)^2+o(1/n^2) \quad \ldots \quad (4.17)$$

where T^*_n and $\hat{\theta}^*_n$ are defined as in Section 2. Thus using (4.14), (4.17) and the definition of B_n, we get

$$\int E_\theta(T^*_n-\theta)^2q(\theta)d(\theta) \geqslant \int E_\theta(\hat{\theta}^*_n-\theta)^2q(\theta)d\theta+o(1/n^2). \quad \ldots \quad (4.18)$$

Since (4.18) is true for all q, Rao's result follows.

---

[1] We shall show elsewhere that this Bayes property of $\hat{\theta}_n$ is a consequence of asymptotic sufficiency up to $o(n^{-1})$ of $\hat{\theta}_n$ and $\left.\frac{d^2 \log L}{d\,\theta^2}\right|\hat{\theta}_n$.

We now turn to the Rao-Fisher result. Consider a fixed λ and a fixed efficient estimator T_n. Let $a_\lambda(\theta)$ be such that

$$E_\theta(Z_n-(T_n-\theta)I-\lambda(T_n-\theta)^2) = a_\lambda(\theta)/n+o(1/n). \quad \dots \quad (4.19)$$

Let $a_{0\lambda}(\theta)$ be defined similarly for $\hat{\theta}_n$.

Consider the loss function

$$l(d, \theta) = \left\{Z_n-(d-\theta)I-\lambda\frac{Z_n^2}{I^2}-a_\lambda(\theta)/n\right\}^2 \quad \dots \quad (4.20)$$

and a prior $q(\theta)$ satisfying the same restrictions as above. One can show as before an estimator of the form $B_n^* = \hat{\theta}_n+c(\hat{\theta}_n)/n$ satisfies

$$E_\theta\{l(B_n^*, \theta)\} = E_\theta\{l(B_n, \theta)\}+o(1/n^2) \quad \dots \quad (4.21)$$

where B_n is the Bayes solution for the loss function given in (4.20). Now it is easy to show that

$$E_\theta\{Z_n-(\hat{\theta}_n-\theta)I-\lambda(\hat{\theta}_n-\theta)^2-a_{0\lambda}(\theta)/n\}^2 = E_\theta\{l(B_n^*, \theta)\}+o(1/n^2). \quad \dots \quad (4.22)$$

and $$E_\theta\{Z_n-(\hat{\theta}_n-\theta)I-\lambda(\hat{\theta}_n-\theta)\} = \frac{a_{0\lambda}(\theta)}{n}+o(1/n). \quad \dots \quad (4.23)$$

Also

$$E_\theta\{Z_n-(T_n-\theta)I-\lambda(T_n-\theta)^2-a_\lambda(\theta)/n\}^2 = E_\theta\{l(T_n, \theta)\}+o(1/n^2). \quad \dots \quad (4.24)$$

Since $q(\theta)$ is arbitrary, we get from (4.21), (4.22) and (4.24), that

$$E_2(\{T_n\}, \theta, \lambda) \geqslant E_2(\{\hat{\theta}_n\}, \theta, \lambda) \text{ for all } \theta \quad \dots \quad (4.25)$$

which gives the Rao-Fisher result.

To justify these heuristic arguments one would of course need various restrictions on F_θ and T_n but one would expect that the restrictions would be much milder than those considered in Section 2.

5. Asymptotic sufficiency and other topics

Consider the same set-up as in the previous section. It will be shown elsewhere that under quite general conditions, $\hat{\theta}_n$, L_2, L_3 and L_4 are asymptotically sufficient up to $o(1/n)$ in the following sense. For each n there exist joint densities $q_{\theta n}$ such that

(1) $\hat{\theta}_n, L_2, L_3, L_4$ are sufficient for $q_{\theta n}$, $\theta \epsilon \Theta$

(2) $\int \dots \int \left| \prod_1^n f_\theta(x_i)-q_n(\theta) \right| \prod_1^n d\mu(x_i) = o(1/n)$.

In a similar sense $\hat{\theta}_n$ and L_2 are asymptotically sufficient up to $o(n^{-\frac{1}{2}})$.

This fact can be used to give another justification of the results on second order efficiency.

Another question of some importance is whether an expansion for $E^{U}(T_n-\theta_0)^2$ can be regarded as an expansion for the variance of the asymptotic distribution of T_n^*. Under certain conditions it can be shown that the distribution function of $\sqrt{nI}(T_n^*-\theta_0)$ has an Edgeworth type expansion[1] in powers of $n^{-\frac{1}{2}}$ and the expansion of $E^{U}(T_n^*-\theta_0)^2$ can be identified with the second moment about mean of the expansion for the distribution function. Under certain additional conditions the density of T_n^* has an Edgeworth type expansion. In this case Fisher's original measure E_2', defined in Section 1, and Rao's E_2 coincide.

So far we have considered second order efficiency of T_n mainly from the point of view of estimation. Suppose we wish to consider the same problem for testing of hypothesis. As in the problem of first order efficiency, one can consider tests of the form :

$$\text{If} \qquad (T_n-\theta_0) > c_n \text{ reject } H_0(\theta \leqslant \theta_0)$$
$$\leqslant c_n \text{ accept } H_0.$$

Expanding the power function of this test locally around θ_0, one could get a measure of second order efficiency.

It is possible to follow Berkson and Hodges (1961) and obtain an expansion for the minimax estimator for the loss function $(d-\theta)^2I(\theta)$ as well as the least favourable prior. One may also wish to know if $\hat{\theta}_n$ is admissible among all estimators of the form $\hat{\theta}_n+c(\hat{\theta}_n)/n$ up to the second order term.

There are many so-called non-regular problems, with the carrier of f depending on θ, where $\hat{\theta}_n$ is efficient. It may be possible to have second order results for such problems also.

We hope to return to these problems later.

---

[1] The leading term in the expansion is the distribution function of a standardized normal variate $N(0, 1)$.

Appendix

Proof of Proposition 1 :

(i) By Lemma 2,

$$\lim E^U n\{Z_n-(T_n-\theta_0)I-\lambda(T_n-\theta_0)^2\}$$
$$= \lim nE^U\{-\Sigma\Sigma T^{ij}(p_i^n-\pi_i)(p_j^n-\pi_j)-\lambda Z_n^2/I^2\}$$
$$= \lim nE_{\theta_0}\{-\Sigma\Sigma T^{ij}(p_i^n-\pi_i)(p_j^n-\pi_j)-\lambda Z_n^2/I^2\} \text{ (by Lemma 1)}$$
$$= -\Sigma\Sigma T^{ij}E_{\theta_0}(p_i^1-\pi_i)(p_j^1-\pi_j)-\frac{\lambda}{I}$$

(ii) By Lemma 2,

$$\lim n^2E^U\{Z_n-(T_n-\theta_0)I-\lambda(T_n-\theta_0)^2-a_\lambda(\theta_0)/n\}^2$$
$$= \lim n^2E^U\{-\Sigma\ \Sigma T^{ij}(p_i^n-\pi_i)(p_j^n-\pi_j)-\lambda Z_n^2/I^2-a_\lambda(\theta_0)/n\}^2$$
$$= \lim n^2E_{\theta_0}\{-\Sigma\ \Sigma T^{ij}(p_i^n-\pi_i)(p_j^n-\pi_j)$$
$$-\lambda Z_n^2/I^2-a_\lambda(\theta_0)/n\}^2 \quad \text{(by Lemma 1)}$$

which is easily seen to exist.

(iii) The required result follows since the limit obtained in the proof of (ii) does not depend on U.

Proof of Proposition 2 :

(i) By Lemma 2 and Lemma 1

$$b(\theta_0) = \Sigma\Sigma T^{ij}E_{\theta_0}(p_i^1-\pi_i)(p_j^1-\pi_j).$$

(ii) b is continuously differentiable if T^{ij}'s are since $E_\theta(p_i^1-\pi_i)(p_j^1-\pi_j)$ is differentiable.

(iii) Follows from Lemma 2 and an expansion of $b(T_n)$ around $b(\theta_0)$, which is possible since b is differentiable.

(iv) That $\qquad \psi(\{T_n'\}, \theta_0) = \lim n^2 \left\{E^U(T_n'-\theta_0)^2-\frac{1}{nI}\right\}$

exists follows from an argument similar to the proof of Proposition 1 (ii). However one needs an analogue of Lemma 3 in addition to Lemmas 1 and 2. Details are omitted.

(v) We can choose U_1 so small that

$$(T_n'-\theta_0)^2 < d \text{ if } p^n\epsilon U_1.$$

Then

$|E_{\theta_0}\{W(T_n', \theta_0)\}-E^{U_1}(T_n'-\theta_0)^2| \leqslant dP_{\theta_0}\{p^n\epsilon U_1^c\} = o(n^{-2})$ by Lemma 1. But

$|E^U(T_n'-\theta_0)^2-E^{U_1}(T_n'-\theta_0)^2| = o(n^{-2})$ by (iv). This completes the proof.

ACKNOWLEDGEMENT

We would like to thank the referee for improvements in the presentation and Dr. C. R. Rao for bringing to our notice the paper of Efron (1974). We are also indebted to Mr. C. Srinivasan for many helpful discussions.

REFERENCES

BERKSON, J. (1955): Maximum likelihood and minimum χ^2 estimates of the logistic function. *JASA*, 50, 130-126.

BERKSON, J. and HODGES, J. L. JR. (1961): A minimax estimator for the logistic function. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, 4, 77-86.

CHERNOFF, H. (1952): A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23, 493-507.

EFRON, B. (1974): Defining the curvature of a statistical problem (with applications to second order efficiency). (to be published).

FERGUSON, T. S. (1967): *Mathematical Statistics A Decision Theoretical Approach*, Academic Press, New York and London.

FISHER, R. A. (1925): Theory of statistical estimation. *Proceedings Cambridge Phil. Soc.*, 22, 700-725.

HAJEK, J. (1972): Local asymptotic minimax and admissibility in estimation. *Proceedings of the sixth Berkeley Symposium*, 1, 175-194.

HODGES, J. L., JR., and LEHMANN, E. L. (1970): Deficiency. *Ann. Math. Statist.*, 41, 783-801.

JOHNSON, R. A. (1970): Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.*, 41, 851-864.

KENDALL, M. G. (1946): *The Advanced Theory of Statistics*, First edition, 2, p. 30, Griffin and Co., London.

LINDLEY, D. V. (1961): The use of prior probability distributions in statistical inference and decision. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics*, 1, 453-468.

NANDI, H. K. (1956): Variance of a linear function of constrained variables. *Cal. Stat. Assoc. Bulletin*, 7, 41-43.

PFANZAGL, J. (1973): Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.*, 1, 993-1026.

RAO, C. R. (1960): Apprent anomalies and irregularities of the method of maximum likelihood. *Proceedings of the 32nd session of the Int. Stat. Inst.*, Tokyo. Reprinted with discussion in *Sankhyā*, 24, 73-102.

——— (1961): Asymptotic efficiency and limiting information. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1, 531-546.

——— (1962): Efficient estimates and optimum inference procedure in large samples, with discussion. *J.R.S.S.*, Ser. B, 24, 46-63.

——— (1963): Criteria of estimation in large samples. *Sankhyā*, Ser. A., 25, 189-206. Also included in *Contributions to Statistics* presented to Professor P. C. Mahalanobis on the occasion of his 70-th birthday.

ROBERTSON, C. A. (1972): On minimum discrepancy estimators. *Sankhyā*, Ser. A, 34, 133-144.

SILVERSTONE, H. (1957): Estimating the logistic curve. *JASA*, 52, 567-577.

Paper received: January, 1974.
Revised: May, 1974.