SANKHYA

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

Vol. 15	Part 4	1955

ON FISHER'S LOWER BOUND TO ASYMPTOTIC VARIANCE OF A CONSISTENT ESTIMATE

By G. KALLIANPUR

and

C. RADHAKRISHNA RAO

Indian Statistical Institute, Calcutta

1. Introduction

It is now more than 30 years since R. A. Fisher put forward the concepts of Consistency, Efficiency, Sufficiency and Likelihood in a paper entitled, 'The mathematical foundations of theoretical statistics' which is the first large scale attack on the problem of estimation. These ideas were further developed in another paper in 1925.

Suggesting that consistency, efficiency and sufficiency may be laid down as the common-sense criteria to be employed in problems of estimation, Fisher (1922) proposed the method of maximum likelihood (m.l.) as a formal solution to the problem of estimation. Besides some intuitive considerations which led the author to lay down the method of maximum likelihood as a 'primitive postulate', some reasons for its choice seem to be, in order of importance, as follows:

(a) "It supplies a method which for each particular problem will lead us automatically to the statistic by which the criterion of sufficiency is satisfied" (p.323, 1922).

(b) The method utilizes only the information supplied by the sample as observed data, without making any assumption about the apriori distribution of the unknown parameters (p.324, 1922).

(c) It leads to efficient (asymptotically minimum variance) estimates in the class of asymptotically normal estimates (p. 711, 1925).

Although no rigorous proofs or the exact conditions under which the above assertions are true were given by Fisher in 1922, 1925, the number and variety of new results the method of maximum likelihood disclosed and its universal applicability made it attractive more than any other method of estimation.

A number of workers investigated the exact conditions under which the above statements are true. Hotelling, Doob and Dugue provided the conditions for asymptotic normality, consistency and efficiency (see Lecam, 1953, for references). But according to Lecam (1953), 'The proofs are not rigorous and mistakes are apparent'. Wilks, Neyman, Barankin and Gurland (for references see Lecam, 1953) showed under some restrictions on the estimating function or the estimating equation that the m.l. provides efficient estimates.

While it is believed that some reasonable restrictions are necessary to avoid seemingly pathological estimates being offered as alternatives some authors produced ingenious examples intended to show that the assertions made by Fisher are not of a general nature and particularly that his concept of efficiency is void. For instance, Lecam (1953) quotes a method due to Hodges, by which any consistent estimate $T_n \to \theta$ in probability can be replaced by another which is not less efficient anywhere and more efficient at a large number of values of the parameter. This implies that there does not exist a best consistent estimate in the sense of possessing minimum asymptotic variance. The lower bound to the asymptotic variance is zero at all values of the parameter contradicting Fisher's inequality to asymptotic variance.

This is, no doubt, disturbing but not at all alarming because consistency in the above sense is by itself not a particularly attractive property of the estimate and some very reasonable properties which may be required of an estimate ensure the existence of a lower bound other than zero. The definition of consistency in the sense of a probability limit does not impose any restriction on the estimate however large n, the sample size, may be. For instance a rule of estimation of the form, 'do whatever you like up to $n = 10,000$ and for larger n choose $\bar{x}$ as the estimate of the mean of a normal population' satisfies the criterion of consistency in the above sense. Such a condition is useless because the practical interest chiefly lies in values of n not very large.

One restriction which has some merit is that the estimate should be unbiased for each n. Under this restriction it has been shown by Cramer, Darmois, Frechét and Rao that there is a positive lower bound to variance. This inequality was same as the inequality for asymptotic variance given by Fisher (1925) and therefore some writers were led to believe that a rigorous proof of Fisher's inequality was provided by the above authors.

It is relevant to examine at the outset the conditions under which Fisher made an assertion about the existence of efficient estimates because the concept of 'efficiency as defined by Fisher' has been the subject of criticism for some time. Let us first examine how far this criticism is due to misunderstanding of Fisher's statements.

The first is the definition of consistency. On the first page of Fisher's (p. 309, 1922) paper the following definition is found.

Consistency: 'A statistic satisfies the criterion of consistency, if, when it is calculated from the whole population, it is equal to the required parameter.' Later on page 316, 'That when applied to the whole population the derived statistic should be equal to the parameter'.

It is clear from this definition and examples given in the 1922 paper that Fisher was considering a statistic as a function of the observed distribution function. When the observed distribution function is same as the true one then the actual value of the parameter must be realised. If the statistic T_n is a continuous functional of the distribution function then it is easy to see that

$$\text{Prob}\{|T_n-\theta|>\epsilon\} \to 0 \quad \text{as} \quad n\to\infty$$

or $T_n\to\theta$ in probability, which is the definition of consistency without any restriction on the statistic that has come into current usage. The latter definition implies Fisher consistency (FC) if T_n is a continuous functional.

It is this second definition of probability limit which Fisher stressed in his 1925 paper and used in his subsequent papers. But some restriction on the statistic seems to have been tacitly assumed. In his 1938 paper, Fisher says (p.42) that if T tends to a limiting value 'it is easily recognised by inserting for the frequencies in our sample their mathematical expectation'. Later in this paper (p.45), consistency of the estimate obtained by equating a linear function of the frequencies to zero is verified by expressing the condition that when expected frequencies are substituted the equation is automatically satisfied. Again in 1938, Fisher states, 'The problem of estimation is to find from the sample point the most appropriate point on the curve of expectation. Thus every method of estimation is naturally equivalent to dividing up the space into what may be called equi-statistical regions such that every sample point on the same region leads to the same estimate. The criterion of consistency simply states that the equi-statistical region leading to any estimate of θ should actually cut the curve of expectation at the point corresponding to this value of θ'.

It is, thus, clear by consistency Fisher had in mind both the properties of the statistic tending to a limit in probability and the limiting value being attained by the statistic at the expected value of the frequencies. This would imply that the statistic need be continuous in probability only. This property alone may not be sufficient to ensure a limiting asymptotic variance. What we have shown in this paper is that Fisher's original definition of consistency together with Frechét differentiability ensures a lower limit to asymptotic variance. In proving this the pioneering work of von Mises (1947) on differentiable statistical functions has been extremely useful.

In a recent discussion with Fisher at the Indian Statistical Institute, the authors were told that he was considering only analytic functions of frequencies. This is more than what is actually needed to prove all the assertions about asymptotic efficiency and normality.

While it seems reasonable that we should consider only analytic functions critics may still maintain that Fisher has not explicitly stated this in his earlier papers nor can analyticity be inferred from other properties assumed such as asymptotic normality.

In this paper, we give some sufficient conditions under which there exists a lower bound to the asymptotic variance. Two different situations are considered. First the estimation of parameters based on the observed frequencies of a multinomial distribution. In this case the problem can be answered using the ordinary notions of continuity and differentiability with respect to several variables as shown in Rao (1955). Second, the estimation of parameters of a continuous distribution based on a set of observations on the random variable. At first it may appear that the continuous case can be treated as a multinomial with the number of class intervals tending to infinity. But to make the argument more rigorous we use the topological properties of function spaces in which the concepts of continuity and differentiability can be suitably defined. We refer to the pioneering work of von Mises (1947) on differentiable statistical functions but do not make use of the methods developed there. On the other hand, the concept of Frechet differentiability seems to provide a more fruitful approach to the problem at hand. To make the reader somewhat familiar with these concepts, some useful definitions and theorems on differentiable functionals are quoted.

In a forthcoming paper, a further use of these results will be made in proving some optimum properties of likelihood estimates similar to those given in the case of minimum chi-square by Rao (1955).

2. Estimation of parameters in a multinomial distribution

Let $n_1, \ldots, n_k, (\Sigma n_i = n)$ be the observed frequencies in k classes of a multinomial distribution with probabilities $\pi_1, \ldots, \pi_k$, depending on unknown parameters $\theta_1, \theta_2, \ldots$. The relative frequencies $(n_1/n), \ldots, (n_k/n)$, are represented by $p_1, \ldots, p_k$. According to Fisher, a statistic which is a function of the relative frequencies only is consistent (FC) if, when the true proportions are substituted the function identically reduces to θ or in other words $T(\pi_1, \ldots, \pi_k) \equiv \theta$. While the estimating function assumes the true value θ, when the true proportions are realised it is also reasonable to demand that when the observed proportions are close to the true ones the estimate should not be far from the true value. This implies continuity of the estimating function with respect to the relative frequencies. From Slutsky's theorem it follows that if $T(p_1, \ldots, p_k)$ is continuous then it converges in probability (CP) to $T(\pi_1, \ldots, \pi_k)$ which is identical with θ under (FC). Therefore in the case of a continuous function FC $\longleftrightarrow$ CP. We shall now prove a theorem containing a set of sufficient conditions under which there exists a lower bound to the asymptotic variance of a Fisher consistent (FC) statistic.

Theorem 1: *Let $\theta = \phi(\theta_1, \theta_2, \ldots)$ be the parametric function to be estimated. If*

(a) *$T(p_1, \ldots, p_k)$ is FC for θ,*

(b) *$T(p_1, \ldots, p_k)$ admits continuous partial derivatives,*

(c) *$\phi(\theta_1, \theta_2, \ldots)$ admits all partial derivatives and so also each π_i as a function of $\theta_1, \theta_2, \ldots$.*

then

(i) *$T(p_1, \ldots, p_k)$ has asymptotic normal distribution in the accepted sense of the term.*

(ii) *The asymptotic variance is*

$$V(T) = \left[\Sigma \pi_i \left(\frac{\partial T}{\partial \pi_i}\right)^2 - \left(\Sigma \pi_i \frac{\partial T}{\partial \pi_i}\right)^2\right] \div n$$

(iii)
$$V(T) \geqslant \Sigma \Sigma I^{rs} \frac{\partial \phi}{\partial \theta_r} \frac{\partial \phi}{\partial \theta_s}$$

where (I^{rs}) is the matrix inverse to (I_{rs}), the information matrix defined by Fisher.

All the conditions need be satisfied only in the neighbourhood of the true values.

Proofs of (i) and (ii) are well known. The proof of (iii) is as follows. The condition of FC, $T(\pi_1, \ldots, \pi_k) = \phi(\theta_1, \theta_2 \ldots)$ implies under regularity conditions

$$\left.\begin{aligned} &\sum \frac{\partial T}{\partial \pi_i} \frac{\partial \pi_i}{\partial \theta_j} = \frac{\partial \phi}{\partial \theta_j},\ j = 1, 2, \ldots \\ \text{or} \quad &\sum \pi_i \left(\frac{\partial T}{\partial \pi_i}\right)\left(\frac{1}{\pi_i}\frac{\partial \pi_i}{\partial \theta_j}\right) = \frac{\partial \phi}{\partial \theta_j},\ j = 1, 2, \ldots \end{aligned}\right\} \quad \ldots \quad (2.1)$$

By definition

$$\left.\begin{aligned} &\sum \pi_i \left(\frac{1}{\pi_i}\frac{\partial \pi_i}{\partial \theta_r}\right)\left(\frac{1}{\pi_i}\frac{\partial \pi_i}{\partial \theta_s}\right) = i_{rs} = I_{rs} \div n \\ \text{and} \quad &\sum \pi_i \left(\frac{\partial T}{\partial \pi_i}\right)^2 - \left(\Sigma \pi_i \frac{\partial T}{\partial \pi_i}\right)^2 = nV(T). \end{aligned}\right\} \quad \ldots \quad (2.2)$$

The equations (2.1), (2.2) imply that

$$\begin{vmatrix} V(T) & \frac{\partial \phi}{\partial \theta_1} & \frac{\partial \phi}{\partial \theta_2} & \ldots \\ \frac{\partial \phi}{\partial \theta_1} & I_{11} & I_{12} & \ldots \\ \frac{\partial \phi}{\partial \theta_2} & I_{12} & I_{22} & \ldots \\ . & . & . & \ldots \end{vmatrix} \geqslant 0 \quad \ldots \quad (2.3)$$

which on expansion by the bordered elements yields

$$V(T) \geqslant \Sigma \Sigma I^{rs} \frac{\partial \phi}{\partial \theta_r} \frac{\partial \phi}{\partial \theta_s}. \quad \ldots \quad (2.4)$$

Hence (iii) of Theorem 1. In the case of a single parameter

$$V(T) \geqslant 1/I \quad \ldots \quad (2.5)$$

where I is the information for a single parameter. Fisher (1925), in his original proof of this inequality replaces the conditions of Theorem 1 by FC and asymptotic normality without explicitly stating any analytic property of the statistic. It is, however, an interesting problem to examine to what extent analyticity of a statistic and normality of its asymptotic distribution are related. If FC is replaced by CP then the inequality may not hold as shown by Hodges.

The conditions of Theorem 1 can be relaxed in several ways without disturbing the lower bound. One such set of conditions is given in Theorem 2.

Theorem 2: *Let $\theta = \phi(\theta_1, \theta_2, \ldots)$ be the parametric function to be estimated and $T(p_1, \ldots, p_k, n)$ be the estimating function which may be an explicit function of n also. If*

(a) $T(\pi_1, \ldots, \pi_k, n) = \theta + b_n(\theta_1, \theta_2, \ldots)$ *where* $\sqrt{n}\, b_n \to 0$ *as* $n \to \infty$;

(b) $T(p_1 \ldots, p_k, n)$ *admits first partial derivatives continuous at* $\pi_1, \ldots, \pi_k$ *uniformly in* n;

(c) *same as condition (c) of Theorem* 1,

then $T(p_1, \ldots, p_k, n)$ is normally distributed about θ with an asymptotic variance not less than c/n where

$$c = \lim_{n \to \infty} n \sum \sum I^{rs} \left(\frac{\partial \phi}{\partial \theta_r} + \frac{\partial b_n}{\partial \theta_r}\right)\left(\frac{\partial \phi}{\partial \theta_s} + \frac{\partial b_n}{\partial \theta_s}\right).$$

The proof is similar to that of Theorem 1.

3. ESTIMATION OF PARAMETERS IN A CONTINUOUS DISTRIBUTION

When we turn to the problem of generalising the results of the previous section to the case of a general continuous distribution $F(x, \theta_1, \theta_2, \ldots)$ involving unknown parameters $\theta_1, \theta_2, \ldots$, the general class of estimating functions which naturally suggests itself is the class of functionals of the empirical distribution function $S_n(x)$. Many of the familiar statistics in use (e.g. the sample moments, median etc.) are functionals of S_n. If $f[S_n]$ is the estimating functional of $\theta = \phi(\theta_1, \theta_2, \ldots)$ then FC demands that

$$f[F(x, \theta_1, \theta_2, \ldots)] = \theta.$$

If f is a continuous functional, then $f[S_n(x)] \to \theta$ in probability since $S_n(x)$ strongly converges to $F(x, \theta_1, \theta_2, \ldots)$. Under the condition of continuity convergence in probability implies FC as in the discrete case.

To prove further results, it is necessary to consider the concept of differentiability of a statistical functional. This was developed by von Mises (1947) using the concept of Volterra derivative. Defining the first differential as a linear functional representable as an integral with respect to the increment, von Mises proceeds to prove asymptotic normality under some conditions. For a rigorous demonstration and as a natural extension of the condition (*b*) in Theorem 1, we require that the estimating functional should be differentiable in the sense of Frechét (1925).

Some basic notions:

(1) Any transformation $f(x)$ which has for its domain a subset D of a normed, linear space E and the set of real numbers for its range is called a (real-valued) functional.

(2) A functional f is continuous at $x_0 \epsilon D$ if $\lim f(x) = f(x_0)$ as $\|x-x_0\| \to 0$.

(3) A functional f having E for its domain is said to be *additive* if for any two elements x_1 and x_2

$$f(x_1+x_2) = f(x_1)+f(x_2).$$

(4) An additive, continuous functional f (defined over E) is called *linear*.

It is easy to see that if an additive functional f is continuous at a single point x_0, it is continuous everywhere.

Furthermore, a linear functional f is homogeneous; i.e., $f(\lambda x) = \lambda f(x)$, λ being any scalar.

(5) Let f be any functional with domain $D \subseteq E$. Then f is said to possess a *Frechèt differential* (F-differential) at x_0, if for $h \epsilon E$ so that $(x_0+h) \epsilon D$ there exists a functional $L(x_0, h)$ linear in h such that

$$\lim_{\|h\| \to 0} \frac{|f(x_0+h)-f(x_0)-L(x_0\, h)|}{\|h\|} = 0\,.$$

Remark: The continuity of L (implied by its linearity) is a natural requirement because it is desirable that functionals F-differentiable at x_0 should *a fortiori* be continuous at x_0. That this is the case follows from the relation

$$f(x_0+h)-f(x_0) = L(x_0, h)+\rho(h),\ \rho(h) = o(\|h\|) \qquad \dots \quad (3)$$

which shows that f is continuous at x_0 if and only if L is continuous at 0, i.e. if and only if L is a continuous functional.

F-Differentiable, FC-Statistical functionals:

Let E be the normed, linear space of functions $V(x)$ of bounded variation in $(-\infty, \infty)$ with norm $\|V\| = \sup_{-\infty<x<\infty} |V(x)|$ and let D be the set of continuous distribution functions $F_\theta(x)$, (involving a single parameter $\theta \epsilon$ a set C of possible values) and empirical d.f.'s $S_n(x)$ for all n. $F_\theta(x)$ is assumed to have a derivative $p_\theta(x)$ which satisfies the regularity conditions which validate differentiating with respect to the parameter under the integral sign.

In this section, we shall consider FC-functionals which possess a Frechét differential at the true point $F_\theta(x)$. In other words, we make the following assumptions about the estimating functionals:

(*i*) $$f[F_\theta] = \theta \qquad \dots \quad (3.1)$$

(*ii*) f possesses an F-differential at F_θ,

that is, $$f[G]-f[F_\theta] = L(F_\theta, h)+\rho(h) \qquad \dots \quad (3.2)$$

where $h = G - F_\theta$, $G \epsilon D$,

$$\frac{\rho(h)}{\|h\|} \to 0 \quad \text{as} \quad \|h\| \to 0.$$

and (*iii*)

$$L(F_\theta, h) = \int_{-\infty}^{+\infty} \phi(F_\theta, x) dh(x). \qquad \dots \quad (3.3)$$

Equation (3.3) is an assumption about the specific form of the linear functional L and is suggested by von Mises' definition of the Volterra derivative. Here ϕ is assumed to be a Baire function of x.

Theorem 3: *Let $f[S_n]$ be a functional defined over D and satisfying conditions* (3.1)-(3.3). *Then $\sqrt{n}\{f[S_n]-\theta\}$ has an asymptotic normal distribution if and only if*

$$\int_{-\infty}^{+\infty} \phi^2(F_\theta, x) dF_\theta(x) < \infty. \qquad \dots \quad (3.4)$$

Proof: From the assumption of F-differentiability we have

$$f[F_\theta + h] - f[F_\theta] = \int \phi(F_\theta, x) dh(x) + \rho(h)$$

where

$$\lim_{\|h\| \to 0} \frac{\rho(h)}{\|h\|} = 0.$$

We now use an argument similar to one of Barankin and Gurland (1950). Given $\delta > 0$, let $\mu(\delta)$ be the supremum of all μ for which

$$\|h\| < \mu \quad \text{implies} \quad |\rho(h)| < \delta \|h\|. \qquad \dots \quad (3.5)$$

Then $\mu(\delta) \to 0$ implies $\delta \to 0$. Taking $h(x) = S_n(x) - F_\theta(x)$ we shall show that

$$\text{prob.} \lim_{n \to \infty} \sqrt{n} \rho(h) = 0. \qquad \dots \quad (3.6)$$

Clearly

$$P(\sqrt{n}|\rho(h)| > \epsilon) = P\left(\frac{\sqrt{n}|\rho(h)|}{\|h\|} > \frac{\epsilon}{\|h\|}\right)$$

$$= P\left(\frac{|\rho(h)|}{\|h\|} > \frac{\epsilon}{\|h\|\sqrt{n}}\right)$$

$$< P\left\{\|h\| > \mu\left(\frac{\epsilon}{\|h\|\sqrt{n}}\right)\right\} \qquad \dots \quad (3.7)$$

from (3.5). For those h satisfying the inequality within the braces of the right hand side of (3.7) we must have for every positive constant $\lambda, \sqrt{n}\|h\| > \lambda$ for all sufficiently large n. Otherwise there exists a sequence $\{n_k\}$, $n_k \to \infty$ and a constant λ_0 such that $\sqrt{n_k}\|h_{n_k}\| < \lambda_0$ implying $\|h_{n_k}\| \to 0$ which in turn implies that $\mu(\epsilon/\sqrt{n_k}\|h_{n_k}\|) \to 0$ which is a contradiction. Hence

$$P(\sqrt{n}\,|\rho_n| > \epsilon) < P(\sqrt{n}\,\|h\| > \lambda). \quad \ldots \quad (3.8)$$

Now

$$\sqrt{n}\|h\| = \sqrt{n} \sup_{-\infty < x < \infty} |S_n - F_\theta|$$

has, by Kolmogorov's theorem, a limiting distribution (if $F_\theta(x)$ is continuous), say $K(\lambda)$. Thus (3.8) gives

$$\limsup_{n \to \infty} P(\sqrt{n}\,|\rho_n| > \epsilon) < 1 - K(\lambda).$$

Since the r.h.s. of the above can be made arbitrarily small making λ large enough, (3.6) follows.

If

$$\int \{\phi(F_\theta, x)\}^2 dF_\theta(x) < \infty$$

then

$$\frac{\sqrt{n}}{\sigma_f} \int \phi(F_\theta, x) dh(x)$$

has a limiting normal distribution. This fact combined with (3.6) implies that

$$\frac{\sqrt{n}}{\sigma_f} \left\{ f[S_n(x)] - f[F_\theta(x)] \right\}$$

has also the same asymptotic normal distribution, with zero mean and unit variance, where

$$\sigma_f^2 = \int \phi^2 dF_\theta - \left(\int \phi dF_\theta \right)^2.$$

On the other hand, the asymptotic normality of $\sqrt{n}(f[S_n] - f[F_\theta])$ together with (3.6) implies that of

$$\sqrt{n} \int \phi(F_\theta, x) dh(x)$$

which is of the form

$$\frac{Y_1 + \ldots + Y_n}{\sqrt{n}} - \sqrt{n} \int \phi dF_\theta$$

where

$$Y_i = \phi(F_\theta, x_i)$$

are independent, identically distributed r.v.'s. Hence by Khinchine's theorem (see Gnedenko and Kolmogorov, 1949), $E(Y^2)$ i.e.

$$\int \{\phi(F_\theta, x)\}^2 dF_\theta(x)$$

is finite.

We now proceed to the main result of this section, which obtains Fisher's lower bound for the asymptotic variance of FC and asymptotically normal statistical functionals.

Theorem 4: *Let f satisfy (3.1)-(3.3) and in addition suppose that*

$$\sqrt{n}\{f[S_n]-\theta\} \quad \ldots \quad (3.9)$$

is asymptotically normally distributed with mean zero and variance σ_f^2. In view of Theorem 3, (3.9) can be replaced by the equivalent condition

$$\int_{-\infty}^{+\infty} \{\phi(F_\theta, x)\}^2 dF_\theta(x) < \infty. \quad \ldots \quad (3.10)$$

Then

$$\sigma_f^2 \geqslant \frac{1}{E_\theta\left(\frac{\partial \log p}{\partial \theta}\right)^2}. \quad \ldots \quad (3.11)$$

Proof: We may assume from now on without loss of generality that

$$\int_{-\infty}^{+\infty} \phi(F_\theta, x) dF_\theta(x) = 0.$$

Let ϵ be so small that $\theta+\epsilon$ belongs to C. (C can be assumed to be an open set.) Taking $h = F_{\theta+\epsilon}-F_\theta$ we have

$$f[F_{\theta+\epsilon}]-f[F_\theta] = \int \phi(F_\theta, x) d(F_{\theta+\epsilon}-F_\theta)+\rho(\epsilon).$$

The left side of the above relation is ϵ since f is FC. Hence

$$1 = \int \phi(F_\theta, x) \left\{\frac{p_{\theta+\epsilon}(x)-p_\theta(x)}{\epsilon p_\theta(x)}\right\} p_\theta(x)dx + \frac{\rho(\epsilon)}{\epsilon}.$$

The conditions imposed on the density $p_\theta(x)$ imply that the first term on the right side of the above equation tends to

$$\int \phi(F_\theta, x) \frac{\partial \log p}{\partial \theta} p_\theta(x) dx$$

while

$$\left|\frac{\rho(\epsilon)}{\epsilon}\right| = \frac{|\rho(\epsilon)|}{\|h\|} \frac{\|F_{\theta+\epsilon}-F_\theta\|}{|\epsilon|}.$$

Since $\lim\limits_{\epsilon\to 0} p_{\theta+\epsilon}(x) = p_\theta(x)$ for a.e.x and the limit $p_\theta(x)$ is also a density function we have

$$\lim_{\epsilon\to 0}\int_{-\infty}^{+\infty} |p_{\theta+\epsilon}(x)-p_\theta(x)|dx = 0$$

and hence

$$\|F_{\theta+\epsilon}-F_\theta\| \leqslant \int_{-\infty}^{+\infty} |p_{\theta+\epsilon}(x)-p_\theta(x)|\, dx \to 0.$$

Thus

$$\frac{\rho(\epsilon)}{\|h\|} \to 0 \quad \text{as} \quad \epsilon \to 0 .$$

Further it is easy to verify that

$$\limsup_{\epsilon \to 0} \frac{\| F_{\theta+\epsilon} - F_\theta \|}{\epsilon} < \int_{-\infty}^{+\infty} \left| \frac{\partial \log p}{\partial \theta} \right| p_\theta(x) dx < \infty .$$

Hence

$$\lim_{\epsilon \to 0} \frac{\rho(\epsilon)}{\epsilon} = 0 \qquad \text{and} \qquad \int \phi(F_\theta, x) \frac{\partial \log p}{\partial \theta} p_\theta(x) dx = 1 .$$

An application of Schwarz's inequality gives

$$\sigma_f^2 = \int \phi^2(F_\theta, x) p_\theta(x) dx \geqslant \frac{1}{E_\theta\left(\frac{\partial \log p}{\partial \theta}\right)^2}$$

which is the lower bound given by Fisher.

4. Simultaneous estimation of parameters

Let $T = (T_1, T_2, \ldots)$ be the vector of FC estimates of the vector of parametric functions $\Phi = (\Phi_1, \Phi_2; \ldots)$. Let each T_i satisfy the regularity conditions of theorems 1 or 4 (according as the distribution is multinomial or continuous). Then it is easy to see that the statistic

$$t = l_1 T_1 + l_2 T_2 + \ldots$$

where l_i are arbitrary, is FC for $\phi = l_1 \Phi_1 + l_2 \Phi_2 + \ldots$ and under the regularity conditions assumed it follows that $V(t)$, the asymptotic variance of t, satisfies the relation

$$V(t) \geqslant \Sigma \Sigma I^{rs} \frac{\partial \phi}{\partial \theta_r} \frac{\partial \phi}{\partial \theta_s}$$

or expressing in terms of l_i the inequality becomes

$$\Sigma \Sigma l_i l_j \operatorname{cov}(T_i, T_j) \geqslant \Sigma \Sigma l_i l_j \Sigma \Sigma I^{rs} \frac{\partial \Phi_i}{\partial \theta_r} \frac{\partial \Phi_j}{\partial \theta_s}$$

or

$$\Sigma \Sigma l_i l_j \left(C_{ij} - \Sigma \Sigma I^{rs} \frac{\partial \Phi_i}{\partial \theta_r} \frac{\partial \Phi_j}{\partial \theta_s} \right) \geqslant 0 .$$

Hence the matrix

$$\left(C_{ij} - \Sigma \Sigma I^{rs} \frac{\partial \Phi_i}{\partial \theta_r} \frac{\partial \Phi_j}{\partial \theta_s} \right)$$

is non-negative definite, where C_{ij} is to be interpreted as asymptotic covariance. We thus arrive at a series of results similar to that for unbiased estimates (Rao, 1945).

References

Barankin, E. W. and J. Gurland (1950): On asymptotically normal efficient estimates. University of California publication in statistics, 1, 89-129.

Fisher, R. A. (1922): On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.*, 222, 309-368.

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

——— (1938): *Statistical Theory of Estimation*. Calcutta University lectures.

Frechet, M. (1925): La notion de differentielle dans l'analyse generale. *Ann. Ecole. Norm. Sup.*, 42, 293.

Gnedenko, B. V. and Kolmogorov, A. N. (1954): *Limit distributions of sums of independent random variables*. Translation by K. L. Chung, Addison-Wesley Publishing Company Inc.

Lecam, L. (1953): On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. University of California publication in Statistics, 1, 277.

Rao, C. R. (1945): Information and accuracy attainable in the estimation of statistical parameters. *Bull. Cal. Math. Soc.*, 37, 81-91.

——— (1955): Theory of the method of estimation by minimum chi-square. *Proc. Int. Stat. Conferences*, 1955.

von Mises, R. (1947): On the asymptotic distribution of differentiable statistical functions. *Ann. Math Stat.*, 18, 309-348.

Paper received : August, 1955.