# ASYMPTOTIC RISK OF MAXIMUM LIKELIHOOD ESTIMATES

by

V. M. Joshi
University of North Carolina, Chapel Hill
(on leave from Maharashtra Government, Bombay)

Institute of Statistics Mimeo Series No. 471

May 1966

This research was supported by the Mathematics Division of the Air Force Office of Scientific Research Contract No. AF-AFOSR-760-65.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
Chapel Hill, N. C.

1. Introduction.

The main result proved in this paper is a hypothesis conjectured by Chernoff (1956), that with increase in sample size, the risk, suitably normalized, of a maximum likelihood estimate (m.l.e.) converges to a limit equal to the variance of the asymptotic distribution of the estimate. The hypothesis is shown to hold generally subject to mild conditions. As the asymptotic variance is a lower bound for the risk function for all estimates, the result establishes a new optimum propety of the m.l.e. viz that it is asymptotically of minimum risk.

Chernoff's above mentioned hypothesis has also been briefly referred to in a recent paper of Yu. V. Linnik and N. M. Mitro Fanova (1965).

The efficiency of estimates in general is considered in the last section in connection with superefficient estimates and a revised definition of asymptotic efficiency suggested.

2. Main Result.

We first reproduce the relevant formulae from the above mentioned paper of Chernoff (1956). X is a random variable with a frequency function $f(x, \theta)$, with one unknown parameter $\theta$. We consider a sequence of estimates of $\theta$,

$$T_n = T_n(X_1, X_2, \ldots, X_n)$$

where

$$(1) \qquad T_n - \theta = O_p\left(\frac{1}{\sqrt{n}}\right)$$

A sequence of loss functions is assumed, given by

$$(2) \qquad L_n(t, \theta) = c_{on}(\theta) + c_{2n}(t-\theta)^2$$

where $c_{2n} > o$,

and of normalized loss functions,

$$(3) \qquad L_n^*(t, \theta) = n \left[\frac{L_n(t, \theta) - c_{on}(\theta)}{c_{2n}}\right]$$

where we assume,

(4) $$L_n^*(t,\theta) = n[(t-\theta)^2 + o(t-\theta)^2]$$

in which $o$ is assumed to hold uniformly in n as $t \to \theta$.

Then, c being an arbitrary constant > 0,

$$\liminf_{n \to \infty} \frac{E\, L_n^*(T_n, \ )}{n\, E\{\min[(T_n-\theta)^2, \frac{c^2}{n}]\}} \geq 1$$

so that,

(5) $$\lim_{c \to \infty} \ \lim_{n \to \infty} \ \inf \frac{E\, L_n^*(T_n,\theta)}{n\, E\{\min[(T_n-\theta)^2, \frac{c^2}{n}]\}} \geq 1$$

If $\sqrt{n}\,(T_n-\theta)$ has a limiting distribution with second moment $\sigma^2(\theta)$, then

(6) $$\lim_{c \to \infty} \ \lim_{n \to \infty} \ n\, E\{\min[(T_n-\theta)^2, \frac{c^2}{n}]\} = \sigma^2(\theta)$$

so that $\sigma^2(\theta)$ is a lower bound for the risk function

(7) $$R_n(T_n,\theta) = E\{L_n^*(T_n,\theta)\}$$

On the other hand if,

(8) $$P\{|T_n-\theta| > c\} = o(\frac{1}{n})$$

for each c, it is possible to show that

(9) $$\liminf_{n \to \infty} \frac{n\, E\{(T_n-\theta)^2\}}{E\{L_n^*(T_n,\theta)\}} \geq 1$$

so that the normalized risk is sandwiched between the real variance and the asymptotic variance.

Chernoff remarks that in accordance with the axioms of Neumann and Morgenstein the loss function requires to be bounded above and (8) indicates that he assumes an upper bound C for the loss function in (2) and hence an upper bound n C for the normalized loss function in (3), so that

(10) $$L_n^*(t,\theta) \leq n\, C$$

Chernoff's conjecture then is that if $T_n$ is the maximum likelihood estimate, the standard derivations of the asymptotic normal distribution of $T_n$, can, without unreasonable modifications be used to show that

(11) $\lim_{n \to \infty} E\{L_n^*(T_n, \theta)\}$ = asymptotic variance of $T_n$.

We now give a proof of this conjecture. We assume tha the frequency function $f(x,\theta)$ satisfies the following conditions, sufficient for asymptotic normality, as stated in Cramer's Mathematical Statistics, (1946, §.33.2). Writing f for $f(x,\theta)$.

(i) For almost all (Lebesgue measure) x the derivates $\frac{\partial \log f}{\partial \theta}$, $\frac{\partial^2 \log f}{\partial \theta^2}$, and $\frac{\partial^3 \log}{\partial \theta^3}$ exist for every $\theta$ belonging to a non-degenerate interval A;

(ii) For every $\theta$ in A, we have $\frac{\partial f}{\partial \theta} < F_1(x)$, $\left|\frac{\partial^2 f}{\partial \theta^2}\right| < F_2(x)$ and $\left|\frac{\partial^3 \log f}{\partial \theta^3}\right| < H(x)$ the functions $F_1$ and $F_2$ being (Lebesgue) integrable over $(-\infty, \infty)$ while

$$\int_{-\infty}^{\infty} H(x)\, f(x,\theta)\, dx < M$$

where M is independent of $\theta$,

(iii) For every $\theta$ in A, the integral $\int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \theta}\right)^2 f\, dx$ is finite and positive.

These conditions are sufficient for asymptotic normality of $T_n$. For proving our result, we assume that f satisfies the following further condition.

FURTHER CONDITION: (iv) For every $\theta$ in A, the integrals $\int_{\infty}^{\infty} \left(\frac{\partial^2 \log f}{\partial \theta^2}\right)^2 f\, dx$ and $\int_{-\infty}^{\infty} (H(x))^2 dx$ are finite, or in other words the random variables $\frac{\partial^2 \log f}{\partial \theta^2}$ and $H(x)$ have finite variances.

We now use the same notation as in Cramer (1946) except that we denote the parameter by $\theta$, instead of $\alpha$ and the maximum likelihood estimate by $T_n = T_n(X_1, X_2, \ldots, X_n)$ instead of $\alpha^*$. $\theta_o$ denotes the unknown true value of the parameter $\theta$, and $\theta_o$

is assumed to be an interior point of A. We write $f_i$ in place of $f(x_i, \theta)$ and use the subscript o to indicate that $\theta$ is to be put equal to $\theta_o$. Then putting

$$B_o = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial \log f_i}{\partial \theta}\right)_o, \quad B_1 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial^2 \log f_i}{\partial \theta^2}\right)_o$$

(12) $$B_2 = \frac{1}{n} \sum_{i=1}^{n} H(x_i)$$

and

$$E\left(\frac{\partial^2 \log f}{\partial \theta^2}\right)_o = - E\left(\frac{\partial \log f}{\partial \theta}\right)^2 = - k^2$$

it is shown as in Cramer (1946), that

(13) $$k\sqrt{n}(T_n - \theta_o) = \frac{\frac{1}{k\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\partial \log f}{\partial \theta}\right)_o}{- \frac{B_1}{k^2} - \frac{1}{2} \alpha B_2 (T_n - \theta_o)/k^2}$$

where $0 \leq |\alpha| < 1$.

The asymptotic normality of $(T_n - \theta)$ follows from (13). In (13), we now put for brevity

$$u_n = k\sqrt{n}\,(T_n - \theta)$$

(14) and

$$v_n = - \frac{B_1}{k^2} - 1 - \frac{1}{2} \alpha B_2 (T_n - \theta_o)/k^2$$

so that the denominator in the r.h.s. of (13) is $1+v_n$. Hence from (13), and (14)

$$u_n^2(1 + v_n)^2 = \frac{1}{k^2 n} \left[ \sum_{i=1}^{n} \left(\frac{\partial \log f}{\partial \theta}\right)_o \right]^2$$

so that

(15) $$u_n^2 + 2u_n^2 v_n \leq \frac{1}{k^2 n} \left[ \sum_{i=1}^{n} \left(\frac{\partial \log f}{\partial \theta}\right)_o \right]^2$$

Let $F_n = F_n(x_1, x_2, \ldots, x_n)$ denote the d.f. of the $X_i$, $i = 1, 2, \ldots, n$. We now integrate both sides of (15), with respect to the probability measure determined by $F_n$, on the subset D of the sample space $R_n$, given by

(16) $$D = \{x: \ |T_n - \theta_o| \leq \delta\}$$

where $x = (x_1, x_2, \ldots, x_n)$ denotes a point in $R_n$ and $\delta$ is some positive constant, whose value shall be determined later. We thus have from (15),

(16) $$\int_D u_n^2 \, d F_n + 2 \int_D u_n^2 \ v_n \, d F_n \leq \frac{1}{k^2 n} \int_D \Big[ \sum_{i=1}^{n} \Big(\frac{\partial \log f_i}{\partial \theta}\Big)_o \Big]^2 d F_n.$$

Now in the r.h.s. of (16), for each i, i = 1, 2, ..., n,

$$E\Big(\frac{\partial \log f_i}{\partial \theta}\Big)_o = o$$

and by (12)

$$E\Big(\frac{\partial \log f_i}{\partial \theta}\Big)_o^2 = k^2$$

Hence since the variables $x_i$, i = 1, 2, ..., n, are independently distributed,

$$E\Big[ \sum_{i=1}^{n} \Big(\frac{\partial \log f_i}{\partial \theta}\Big)_o \Big]^2 = \int_{R_n} \Big[ \sum_{i=1}^{n} \Big(\frac{\partial \log f_i}{\partial \theta}\Big) \Big]^2 = nk^2$$

and hence

(17) r.h.s. of (16) $\leq 1$

Now by the assumed condition (iii) $E\{H(x)\}_o$ exists and is finite. Let

(18) $$E\ \{H(x)\}_o = \mu(\theta_o)$$

Then using (14) and (18)

(19) $$\text{l.h.s. of (16)} = \int_D u_n^2 \, d F_n - 2 \int_D u_n^2 \Big(\frac{B_1}{k^2} + 1\Big) d F_n$$
$$- \int_D u_n^2 \ \alpha \ \mu(\theta_o) \frac{(T_n - \theta_o)}{k^2} \, d F_n - \int_D u_n^2 \ \alpha \ [B_2 - \mu(\theta_o)] \frac{(T_n - \theta_o)}{k^2} \, d F_n$$
$$= \text{say } g_1 + 2g_2 + g_3 + g_4$$

where $g_i$, i = 1, 2, ..., 4 is written for brevity for the ith term in the r.h s. of (19).

We now derive upper bounds for $|g_2|$ , $|g_3|$ and $|g_4|$ , by applying Schwarz' inequality and using the fact that for $x \in D$, by (14) and (16)

$$|u_n| \leq k\sqrt{n}\ \delta.$$

We have by Schwarz' inequality

$$(20) \qquad g_2^2 \leq \int_D u_n^4 \, d\, F_n \int_D \left(\frac{B_1}{k^2} + 1\right)^2 d\, F_n$$

Now in (20), using (12),

$$(21) \qquad \frac{B_1}{k^2} + 1 = \frac{1}{nk^2} \sum_{i=1}^{n} \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right]_o$$

where by (12), for each i, i=1, 2, ..., n

$$E \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right] = o$$

and by the assumed further condition (iv)

$$E \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right]^2 = \sigma_1^2(\theta_o) \text{ say}$$

Hence since the $x_i$ are distributed independently

$$E \left\{ \sum_{i=1}^{n} \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right] \right\}^2 = \int_{R_n} \left\{ \sum_{i=1}^{n} \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right] \right\}^2 d\, F_n$$

$$= n\, \sigma_1^2\, (\theta_o)$$

so that,

$$(22) \qquad \int_D \left\{ \sum_{i=1}^{n} \left[ \left( \frac{\partial^2 \log f_i}{\partial \theta^2} \right)_o + k^2 \right] \right\}^2 \leq n\, \sigma_1^2\, (\theta_o)$$

Hence in (20), by (21) and (22),

$$(23) \qquad \int_D \left( \frac{B_1}{k^2} + 1 \right)^2 d\, F_n \leq \frac{\sigma_1^2\, (\theta_o)}{n\, k^4}$$

Again in the first integral in the r.h.s. of (20), since by (16) and (14)

$$u_n^2 \leq k^2\, n\, \delta^2$$

(24) $$\int_D u_n^4 \, d\, F_n \leq k^2\, n\, \delta^2 \int_D u_n^2 \, d\, F_n = k^2\, n\, \delta^2\, g_1$$

Combining (23) and (24) with (20), we have

$$g_2^2 \leq \frac{\delta^2\, \sigma_1^2\, (\theta_o)}{k^2}\; g_1$$

so that

(25) $$|g_2| \leq \frac{\delta \;\; \sigma_1(\theta_o)}{k}\, (g_1)^{\frac{1}{2}}$$

Next in $g_3$ in (19),

$$|\alpha| \leq 1,$$

by assumed condition (ii), $|\mu(\theta_o)| \leq M$,

and by (16) $|T_n - \theta_o| \leq \delta$

and hence

(26) $$|g_3| \leq \frac{M\,\delta}{k^2} \int_D u_n^2 \, d\, F_n = \frac{M\,\delta\; g_1}{k^2}$$

Lastly in $g_4$, since $|\alpha| \leq 1$ and $|T_n - \theta| \leq \delta$

(27) $$|g_4| \leq \frac{\delta}{k^2} \int_D u_n^2 \; |B_2 - \mu(\theta_o)| \; d\, F_n$$

etting

$$g_5 = \int_D u_n^2 \; |B_2 - \mu(\theta_o)| \; d\, F_n$$

we again apply Schwarz' inequality, and have

(28) $$g_5^2 \leq \int_D u_n^4 \, d\, F_n \int_D [B_2 - \mu(\theta_o)]^2 \, d\, F_n$$

In the second integral in the r.h.s. of (28), by (12)

(29) $$B_2 - \mu(\theta_o) = \frac{1}{n} \sum_{i=1}^{n} [H(x_i) - \mu(\theta_o)]$$

where for each i, i = 1, 2, ..., n

by (14) $$E[H(x_i) - \mu(\theta_o)] = o$$

and by assumed further condition (iv)

$$E[H(x_i) - \mu(\theta_o)]^2 \text{ is finite } = \sigma_2^2(\theta_o) \text{ say.}$$

Hence since the $x_i$ are distributed independently, in (29)

$$E[B_2 - \mu(\theta_o)]^2 = \frac{\sigma_2^2(\theta_o)}{n}$$

so that in (28),

(30) $$\int_D [B_2 - \mu(\theta_o)]^2 \, d\,F_n \leq \frac{\sigma_2^2(\theta_o)}{n}$$

Now in the first integral in the r.h.s. of (28), since by (12) and (16)

$$u_n^2 \leq n\,k^2\delta^2$$

(31) $$\int_D u_n^4 \, d\,F_n \leq n\,k\,\delta^2 \int_D u_n^2 \, d\,F_n = n\,k^2\delta^2\,g_1$$

so that combining (27), (28) (30) and (31)

(32) $$|g_4| \leq \frac{\delta^2 \; \sigma_2(\theta_o)}{k} \; (g_1)^{\frac{1}{2}}$$

Now in the r.h.s. of (19), obviously

(33) $$g_1 + 2g_2 + g_3 + g_4 \geq g_1 - 2|g_2| - |g_3| - |g_4|$$

and hence collecting together the results in (25), (26), and (32), we have from (16), (17) and (19), using (33),

(34) $$g_1\left(1 - \frac{M\,\delta}{k^2}\right) - g_1^{\frac{1}{2}}\left[\frac{2\delta \;\; \sigma_1(\theta_o)}{k} + \frac{\delta^2 \; \sigma_2(\theta_o)}{k}\right] \leq 1$$

and denoting the coefficient of $g_1$ in (34) by 2 $\kappa$, where $\kappa \geq o$ we write (34) as

(35) $$g_1(1 - \frac{M\,\delta}{k^2}) - 2\kappa\, g_1^{\frac{1}{2}} \leq 1$$

We now assume that $\kappa$ is sufficiently small so that the coefficient of $g_1$ in the l.h.s. of (35) is positive; say,

(36) $$\delta \leq \frac{k^2}{2\,M}$$

Then solving the quadratic in $g_1^{\frac{1}{2}}$

$$g_1(1 - \frac{M\,\delta}{k^2}) - 2\kappa\, g_1^{\frac{1}{2}} = 1$$

(35) is seen to imply

$$g_1^{\frac{1}{2}} \leq \frac{\kappa + \sqrt{\{1 - \frac{M\,\delta}{k^2} + \kappa^2\}}}{1 - \frac{M\,\delta}{k^2}}$$

$$< \frac{\kappa + \sqrt{\{1 + \frac{M\,\delta}{k^2} + \kappa^2\}}}{1 - \frac{M\,\delta}{k^2}}$$

$$< \frac{\kappa + 1 + \frac{M\,\delta}{2k^2} + \kappa}{1 - \frac{M\,\delta}{k^2}}$$

so that substituting the value of $\kappa$ from (34), we have

$$g_1^{\frac{1}{2}} < \frac{1}{1 - \frac{M\,\delta}{k^2}} \left\{1 + \delta\left[\frac{2\,\sigma_1(\theta_o)}{k} + \frac{\delta^2\,\sigma_2(\theta_o)}{k^2} + \frac{M}{2\,k^2}\right]\right\}$$

which using (36), can be reduced to the form

(37) $$g_1 < 1 + A(\theta_o)\,\delta$$

where $A(\theta_o)$ is independent of $\delta$

Returning to the loss function in (4), we now write it in the form

(38) $$L_n^*(t, \theta_o) = n(t-\theta)^2[1 + h(t, \theta_o)]$$

where by assumption, given any arbitrary number $\epsilon > 0$, we can find a $\delta_o$, such that

(39) $$|h(t,\theta)| \leq \epsilon \text{ for all } t, \text{ such that } |t - \theta_o| \leq \delta_o$$

We now take $\delta = \min[\delta_o, \frac{k^2}{2M}]$, so that $\delta$ satisfies both (36) and (39)

Now

(40) $$E\{L_n^*(T_n, \theta_o)\} = \int_{R_n} L_n^*(T_n, \theta_o)\, d F_n$$

$$= \int_D L_n^*(T_n, \theta_o) d F_n + \int_{R_n - D} L_n^*(T_n, \theta_o) d F_n$$

D being the subset of $R_n$ defined by (16)

Now the first integral in the r.h.s. of (40)

$$= \int_D n(T_n - \theta_o)^2 d F_n + \int_D n(T_n - \theta_o)^2\, h(T_n, \theta_o) d F_n$$

which by (39)

$$\leq (1 + \epsilon) \int_D n(T_n - \theta_o)^2 d F_n$$

Hence noting the definitions of $g_1$ and $u_n$ in (19) and (14)

(41) $$\int_D L_n^*(T_n, \theta) d F_n \leq \frac{(1 + \epsilon) g_1}{k^2}$$

Then consider the second integral in the r.h.s. of (40). Here we note that since the distribution of $T_n$ is asymptotically normal, by a well known property of the Normal frequency function (8) holds for $T_n$. Hence for given $\delta$, and $\epsilon$, by making n sufficiently large, we can make

(42) $$P[|T_n - \theta_o| > \delta] = P[R_n - D] \leq \frac{\epsilon}{n\, C}$$

where C is the upper bound in (10). Then from (10) and (42)

$$(43) \qquad \int_{R_n - D} L_n^* (T_n, \theta_o)\, d F_n \leq \epsilon$$

Now collecting the results in (40), (41), and (43) and using (37), we have

$$(44) \qquad E\{L_n^* (T_n, \theta_o)\} = \int_{R_n} L_n^*(T_n, {}_o)\, d F_n$$

$$\leq (1 + A(\theta_o)\delta) \frac{(1 + \epsilon)}{k^2} + \epsilon$$

for all sufficiently large n.

Since $\epsilon$ and $\delta$ can be made arbitrarily small, it follows that

$$(45) \qquad \limsup_{n \to \infty} E\{L_n^*(T_n, \theta_o)\} \leq \frac{1}{k^2}$$

But $\frac{1}{k^2}$ is the asymptotic variance of $T_n$, and hence from (5) and (6)

$$(46) \qquad \liminf_{n \to \infty} E\{L_n^*(T_n, \theta_o)\} \geq \frac{1}{k^2}$$

(45) and (46) together imply

$$(47) \qquad E\{L_n^* (T_n, \theta_o)\} = \frac{1}{k^2}$$

which was to be proved.

3. Asymptotic Efficiency.

Here we incidentally point out another optimum property of a m.l.e., which is implied by a relation given in Cheroff (1956), but which has a special significance in relation to superefficient estimates, which is stressed here. According to Fisher's original concept of asymptotic efficiency (A.E. for short), the A.E. of any estimate $T_n$ is given by

(48) $$\text{A. E.} = \frac{1}{\lim_{n \to \infty} \{n\, I(\theta)\, E[T_n - \theta]^2\}}$$

where $I(\theta)$ is Fisher's measure of information; evaluated at $\theta$,

(49) $$I(\theta) = E_\theta \{ \frac{1}{f^2} (\frac{\partial f}{\partial \theta})^2 \}$$

It was assumed that this A. E. has an upper bound of 1. This assumption has however turned out to be false, and in fact the A. E. in (48) has no upper bound, so that given any estimate we can construct another uniformly superior to it and there exists no most efficient estimate. The Fisherian concept of A. E is thus void.

A superefficient estimate is one whose A. E. is not less than 1 for any $\theta$ and exceeds 1 for at least one $\theta$. An example of such an estimate was first presented by J. L Hodges, Jr. in 1951. The example relates to a Normae population $N(\theta,1)$, and is

(50) $$T_n = \alpha \bar{x}_n \text{ if } |\bar{x}_n| \leq \frac{1}{n^{\frac{1}{4}}}$$
$$T_n = \bar{x}_n \text{ if } |\bar{x}_n| > \frac{1}{n^{\frac{1}{4}}}$$

where $\alpha$ is some constant, such that $0 \leq |\alpha| < 1$ and $\bar{x}_n$ is the sample mean.

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Further examples of superefficient estimates were later given by Le Cam (1953) who constructed an example where the set of superefficiency, i.e. the set of values of $\theta$ for which the A. E. exceeds 1, is non-denumerable and everywhere dense. It was also proved by him that the set of superefficiency must necessarily be of Lebesgue measure zero.

It should be noted here, that though the A. E. of the superefficient estimate $T_n$ in (48) is for all $\theta$ higher than that of m.l.e. $\bar{x}_n$, $T_n$ is not 'superior' to $\bar{x}_n$, even if we take the M.S.E. (mean squared error) as the sole criterion. While $T_n$ has a lower M.S.E. at $\theta = 0$, for every n, however large, there exist values of $\theta$ (in fact an interval of values) in which $\bar{x}_n$ has a lower M.S.E. The observation made sometimes, that according to a 'behaviourist viewpoint' the superefficient estimate $T_n$ has to be preferred to $\bar{x}_n$ is thus not correct. As pointed out by Le Cam (1953) the same thing occurs in the case of all other superefficient estimates.

The author considers that this reveals a defect in the definition (48) as the possession of a higher A. E. does not imply the possession of a higher degree of some desirable property. The defect in this definition (48) arises from the fact that it takes into consideration the performance of the estimate $T_n$ for each value of n at some single point $\theta = \theta_o$. The true value of $\theta$ is however not known, though as n increases we can locate $\theta$ more and more closely with a given degree confidence. This suggests that a revised definition of A. E. should take into account the performance of $T_n$, not at a single point $\theta_o$ but in some neighbourhood of it, With these considerations a modified definition of A. E. is proposed below.

Let $c_n$ be a sequence of positive numbers such that

$$(51) \qquad \begin{aligned} c_n &\to o \\ \sqrt{n}\, c_n &\to \infty \end{aligned}$$

Then the suggested revised definition is

$$(52) \qquad \text{A.E.} = \liminf_{n \to \infty} \frac{1}{\sup\limits_{\theta_o - c_n < \theta < \theta_o + c_n} [n\, I(\theta)\, E_\theta (T_n - \theta)^2}$$

The following result is proved in Chernoff(1956, Theorem 1)

$$(53) \quad \lim_{k\to\infty} \liminf_{n\to\infty} \sup_{-\frac{k}{4\sqrt{n}} < \theta < \frac{k}{4\sqrt{n}}} \{ n\, I(\theta)\, E_\theta(\min[(T_n-\theta)^2, \frac{k^2}{n}])\} \geq 1$$

From the proof of the theorem, it is seen that the following conditions are assumed,

(a) that the frequency function $f(x,\theta)$ and the estimate $T_n$, satisfy the 'regularity' conditions sufficient for the Cramer-Rao inequality, such as those given by Wolfourtz

(b) $I(\theta)$ is bounded away from zero in some neighbourhood of $\theta = 0$.

For (52) we now replace condition (b) by the condition

(c) $I(\theta)$ bounded away from zero in some neighborhood of every point $\theta_o$, in the range of values assumed by $\theta$.

It is then easily seen to be a consequence of (53), that the A. E defined by (52), has an upper bound of 1. It is further seen that this upper bound is attained for a m.l.e. while for the superefficient estimate $T_n$ the A. E. by the revised definition is nil.

Acknowledgement

The relation given in ( 53 ) was pointed out to me by a referee of the Annals in connection with another paper. I have also to thank Mrs. Kay Herring for doing the typing work neatly and speadily.

References

(1) Chernoff H (1956) Large Sample Theory. Ann. Math Statist. V. 27 1-23.

(2) Cramer Harold (1946). Mathematical Methods of Statistics Princeton University Press. p. 500-503

(3) Le Cam (1953) On some asymptotic properties of maximum likelihood estimates and related Bayes estimates. University of California Pubn Stats Vol I, 1953 p 277-330

(4) Neumann von J and Morgenstein (1947) Theory of Yarnes and Economic Behavior. 2d ed. Princeton University Press, Princeton N. J. 1947, p. 641

(5) Linnik Yu. V. and Mitrofanova N. M. (1965) Some asymptotic expansions for the distribution of the maximum likelihood estimate. Sankhya Series A. Vol. 27 Part 1, p 73

(6) Wolfowitz J (1947) The efficiency of sequential estimates and Walds' equation for sequential processes Ann. Math. Statist. Vol. 18 p 215-230