J Statistics
1994 (1), 37-52

EFFICIENCY AND SUPER – EFFICIENCY OF ESTIMATES

BY

MUJAHID RASUL

(Department Of Statistics, Islamia University, Bahawalpur, Pakistan)

ABSTRACT

Asymptotic theory of estimation based on the criteria of consistency and efficiency is considered and certain methods are shown to yield estimators satisfying this criteria. some results with examples are discussed in which the idea of super – efficiency is not statistically important. We also discuss asymptotic properties of some methods by considering two – piece normal distribution.

1. INTRODUCTION

Much of the work on efficiency was given by an attempt to understand how well the maximum likelihood estimate performs as n goes to infinity. The principle of maximum likelihood proposed by Fisher consists in adopting as an estimate of a parameter θ, the particular value of the parameter which maximizes the probability of the facts actually observed.

Let $X_1, X_2, \ldots\ldots X_n$ be i. i. d. according to a density $f(x\,;\theta)$ and suppose that $T_n = T_n(\underline{x})$ be an estimator of θ. Our object is to examine the asymptotic properties of the estimator T_n as $n \to \infty$. The existence of the first two derivatives of $f(x;\theta)$ with respect to θ allows us to introduce Fisher's amount of information. Let us define

$$I(\theta) = E_\theta\left\{-\frac{\partial^2 \log f(x;\theta)}{\partial^2\theta^2}\right\} \quad \ldots\ldots\ldots\ldots \quad (1.1)$$

as the Fisher information function associated with density $f(x;\theta)$. If the density $f(x;\theta)$ satisfies suitable regularity conditions, the celebrated Cramer – Rao inequality states that the variance of any unbiased estimator T_n of $g(\theta)$ satisfies

$$Var_\theta(T_n) \geq \{g'(\theta)\}^2/n\, I(\theta) \quad \ldots\ldots\ldots\ldots \quad (1.2)$$

Suppose that $T_n(\underline{x})$ is asymptotically normal :

$$n^{1/2}\{T_n - g(\theta)\} \rightarrow N\{0, v(\theta)\},\ v(\theta) > 0 \quad \ldots\ldots \quad (1.3)$$

with $v(\theta) \geq \{g(\theta)\}^2 / I(\theta)$;

If (1.3) holds, T_n will be said to satisfy the Fisher's idea of efficiency. A sequence (T_n) satisfying (1.3) with

$$v(\theta) = \{g'(\theta)\}^2 / I(\theta) \quad \ldots\ldots\ldots\ldots \quad (1.4)$$

is said to be asymptotically efficient. For $g(\theta) = \theta$,

we have

$$v(\theta) \geq I^{-1}(\theta) \quad \ldots\ldots\ldots\ldots \quad (1.5)$$

For a long time, it was believed that for a consistent asymptotically normal (CAN) estimator the asymptotic variance $v(\theta)$ satisfied (1.5) subject only to regularity conditions on the density $f(x;\theta)$. Unfortunately, as everyone knows today that the result (1.5) is not true in general as shown by the examples due to Hodges (Le Cam, 1953). Therefore, there is no lower bound to the asymptotic variance of a CAN estimator, so there does not exist any best CAN estimator, without any further conditions on the estimator Le Can (1953) proved the remarkable result that (1.3) does entail $v(\theta) \geq I^{-1}(\theta)$ almost everywhere (with respect to Lebesgue measure).

Example 1.1

Let $X_1, \ldots . X_n$ be i. i. d. according to the normal distribution $N(\theta, 1)$. In this case $I(\theta) = 1$, and equation (1.5) reduces to $v(\theta) \geq 1$. On the other hand, consider the sequence of estimators,

$$T_n = \begin{pmatrix} \overline{X} & |\overline{X}| > n^{-1/4} \\ C\overline{X} & |\overline{X}| > n^{-1/4} \end{pmatrix}$$

Where $\overline{X}$ is the average of the n observations and C is an arbitrary constant, Then

$$n^{1/2}(T_n - \theta) \to N\{0, v(\theta)\},$$

with $v(\theta) = 1$ when $\theta \neq 1$ and $v(\theta) = C^2$ when $\theta = 0$. If $C < 1$, inequality (1.3) is therefore violated at $\theta = 0$.

2. SUPER – EFFICIENCY

For a long time it was believed that for consistent asymptotically normal (CAN) estimator the asymptotic variance $v(\theta)$ satisfied (1.3) subject only to regularity conditions on the density $f(x \; ; \; \theta)$. Unfortunately, this is not strictly true without any restrictions on the estimating function. This belief was finally exploded by the examples of Hodges to show that the results (1.3) is not true in general. We can call an estimator super–efficient if for all parameter values the estimator is asymptotically normal around the true value with a variance never exceeding and sometimes less than the Cramer – Rao lower bound.

Since Hodge's examples became known many attempts have been made to rescue the Fisher program. All the serious attempts fall under the following two approaches: In the first approach the authors prove that , for any competing estimators (T_n) which satisfy (1.3), the set of points of super – efficiency have Lebesgue measure Zero. That the points of super – efficiency constitute a set of measure zero was stated in LeCam (1953). In the second approach one imposes conditions on the competing estimators which (conditions) are sufficient to eliminate the super – efficient estimators from competition. A number of attempts have been made to rule out super – efficient estimators by imposing regularity conditions which the competing estimators must satisfy.

Rao (1963) established the results that the asymptotic variance of consistent uniformly asymptotically normal (CUAN) estimator has Fisher's lower bound $I^{-1}(\theta)$ when the probability density satisfies some regularity conditions. It appears then that in the examples of Hodges and LeCam, super – efficiency in the sense of having asymptotic variance less than $I^{-1}(\theta)$ has been achieved at the sacrifice of uniform convergence.

Since there is no lower bound to the asymptotic variance of a CAN estimator, it may be thought that an improvement is possible by constructing a statis T_n with a uniformly lower asymptotic variance and thereby increasing the concentration at every value of the parameter, as at $\theta = 0$ in example (1.1). It is no doubt true that an estimator having a higher concentration than another for every value of θ is more useful in drawing inferences on θ from an observed estimate. From the point of view of a statistician, the most appealing sufficient condition is the one due to Rao (1963) who strengthens (1.3) by requiring that it be uniform on compacts. Assuming $n^{1/2}(T_n - \theta)$ is A. N. $\{0, v(\theta)\}$ uniformly on compacts, Rao proves (1.5).

Rao's proof by using Neyman – peasons lemma states that for testing the simple hypothesis $\theta = \theta_0$ against the simple alternative $\theta=\theta_0+n^{1/2}$, the test based on the m. l. estimator is asymptotically most powerful.

LeCam(1953) has shown that the set of super-efficiency must be of Lebesgue measure zero. However, no every set of measure zero can be a set of super efficiency. For certain classes of loss functions and probability densities, he has proved that if an estimate is super-efficient at a given parameter value θ_0, then there must exist an infinite sequence$\{\theta_n\}$ of values at which this estimate is worse than the M.L estimate. Let $\{\sigma_n^2(\theta)\}$ be the asymptotic variance of the M.L. estimate of the parameter θ and $\{\sigma_n^*(\theta)$ that of an alternative estimate T_n. Then T_n is super-efficient if

$$n \xrightarrow{\lim} \infty \, SUP \frac{\sigma_n^*(\theta)}{\sigma_n(\theta)} < 1 \qquad (2.1)$$

or equivalently with slightly stronger condition,

$$\lim_{n \to \infty} \frac{\sigma_n^*(\theta)}{\sigma_n(\theta)} \leq 1-\delta \qquad (2.2)$$

where $\delta > 0$. If an estimate T_n satisfies condition (2.2), then we shall say that T_n is uniformly super-efficient.

Example (2.1)

Let $\{T_n\}$ be a consistent asymptotically normal (CAN) estimate of θ. Let $\{\sigma_n^2(\theta)\}$ be the asymptotic variance of T_n. If there exists a non-negative sequence of numbers $\{\beta_n\}$ satisfying the conditions:

$$\left.\begin{array}{ll} (i) & \lim_{n \to \infty} \beta_n = 0 \\ (ii) & \lim_{n \to \infty} \dfrac{\beta_n}{\sigma_n(\theta)} = \infty \end{array}\right\} \qquad (2.3)$$

and let X be a normal random variable with an unknown mean θ and with variance 1. Then the M.L. estimate of θ based on n independent observations is the arithmatic mean, say $\overline{X}_n$ of the observations. Its variance is $\sigma_n^2 = 1/n$. To produce an estimate T_n uniformly super-efficient, it is sufficient to produce a sequence $\{\beta_n\}$. Because of the particular form of the variance σ_n^2, it is obvious that we can take $\{\beta_n\} = n^{-1/4}$. Suppose that T_n is an estimate for which (2.3) are satisfied. Then they are also satisfied for any estimate which is more efficient than T_n in Fisher's sense. Hence the existence of at-least one estimate T_n satisfying the conditions (2.3) implies the inexistance of an estimate efficient in Fisher's sense.

According to Cox and Hinkley (1974), super-efficiency is not a statistically important because such estimators give no improvement when regarded as test statistics. For any fixed n the reduction in mean square error for parameter points near to the point of super-efficiency is balanced by an increase in mean square error at points a moderate distance away.

3. EFFICIENCY OF M L ESTIMATORS

Asymptotic properties of ML estimates have been discussed in the literature under two separate lines. Some authors, including Cramer (1946) and Gurland (1954), have considered the roots of the likelihood equation, while others, including Wald (1949) and Wolfowitz (1949), have discussed the parameter value which yields the absolute maximum of the likelihood function. Cramer has proved that, under certain regularity conditions, the MLE of θ is consistent and asymptotically efficient. One of his conditions is that for every θ ε Θ

$$\left.\begin{aligned} &\left|\frac{\partial^3 \log f(x;\theta)}{\partial \theta^3}\right| < G(x) \quad and \\ &\int_{-\infty}^{\infty} G(x)\, f(x;\theta)\,dx < \infty \end{aligned}\right\} \qquad (3.1)$$

Kulldorff (1957) replaced this condition with much weaker one. He introduces the following two conditions in the replacement of (3.1) and shows that MLE is asymptotically efficient under the set of new conditions.

(i) There exists a positive function g(θ)

and θ_1, θ_2 ε Θ such that

$$\frac{1}{\theta_1-\theta_2}\int_{-\infty}^{\infty}\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)_{\theta=\theta_1} \quad f(x;\theta_2) < 0$$

where $0 < |\theta_1 - \theta_2| < g(\theta_2)$ (3.2)

(ii) There exists a positive differentiable function p(θ), then for every θ ε θ

$$\frac{\partial}{\partial\theta}\left\{P(Q)\frac{\partial \log f(x;\theta)}{\partial\theta}\right\}$$

is continuous function of uniformly in x. (3.3)

Example (3.1)

Consider the following density function

$$f(x;\theta_1,\theta_2) = (2\pi\theta_2)^{1/2} \quad \exp\left\{-\frac{(x-\theta_1)^2}{2\theta_2}\right\}$$

we have the unique roots for θ_1 , and θ_2 are

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \quad and \quad \hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

These estimates are consistent, asymptotically normal and jointly asymptotically efficient. Now let the random variable X, is normally distributed with mean zero and variance θ. We have likelihood function.

$$L = (2\pi\theta)^{-n/2} \quad \exp\left(-\sum_{i=1}^{n}\frac{x_i^2}{2\theta}\right)$$

and the likelihood equation for θ has the unique root $\hat{\theta}_n = \frac{1}{n}\sum_{i-1}^{n} x_i^2$.

This estimate is consistent and asymptotically efficient. However, we find that the expression

$$\frac{\partial^3 \log f(x;\theta)}{\partial\theta^3} = -\frac{1}{\theta^3} + \frac{3x^2}{\theta^4}$$

tends to infinity as $\theta \to 0$, and is not bounded in the open interval $0 < \theta < \infty$ showing that condition (3.1) is not fulfilled. Now checking conditions (3.2) and (3.3), we find that

$$\frac{1}{\theta_1 - \theta_2} \int_{-\infty}^{\infty} \left(\frac{\partial \log f(x;\theta)}{\partial\theta} \right) \theta = \theta_1 \; f(x;\theta_2) dx = -\frac{1}{2\theta_1^2}$$

which is less than zero and so satisfies condition (3.2). Now introducing positive and differentiable function $p(\theta) = \theta^2$.

We have

$$\frac{\partial}{\partial\theta} \left\{ p(\theta) \frac{\partial \log f(x;\theta)}{\partial\theta} \right\} = -\frac{1}{2}$$

which is continuos of θ uniformly in x, and condition (3.3) is also satisfied. DOSS (1962) also presented a set of conditions which are satisfied by the above example. He shows that under these conditions, $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent, unique, asymptotically normal and jointly asymptotically efficient.

Most of the authors discuss large sample estimation which requires regularity conditions on the second derivative of the likelihood for the MLE to be asymptotically efficient. However, cases are known which are not covered by these regularity conditions. For example, in the density function $f(x;\theta) = \frac{1}{2} e^{-|x-\theta|}$, the sample median is MLE of θ, which is asymptotically normal with variance equal to Cramer-Rao lower bound. But $\partial \log f(x;\theta) / \partial\theta$ is discontinuous and

$\frac{\partial^2 \log f(x;\theta)}{\partial \theta^2} =$ for almost all x. Daniels (1961) applied some weaker conditions for asymptotic efficiency. He proved asymptotic efficiency of $\hat{\theta}$ under the following conditions:

(i) *log f(x;θ)* is continuous in θ through out θ . At every θ_0 there is a neighbourhood such that for all θ, θ` in it,

$$|\log f(x;\theta) - \log f(x;\theta')| < S(x,\theta_o)\,|\theta - \theta_o'| \quad and$$
$$E(S^2/\theta_o) < \infty$$

(ii) At every $\theta, \partial \log f(x;\theta)/\partial\theta$ exists and is continuous for almost all x. It is not almost everywhere zero. It is a no where increasing and some where decreasing function of θ.

According to Daniels, it is enough to apply the condition that the function *log f(x;θ)* is convex in θ for almost all x. A sequence of ML estimates is then under θ asymptotically distributed according to $N\{\theta, I^{-1}(\theta)\}$. This result is of interest because it states the asymptotic normal distribution of a sequence of ML estimates without explicit assumptions on the existence of the second derivative of the function *log f(x;θ)*. On the other hand, it is well known that the second derivative of a continuous and convex function exists upto a set of Lebesgue-measure zero.

Example (3.2)

A random variable is said to have the two-piece normal (TPN) distribution with parameters $\theta_1, \theta_2, \theta_3 > 0$ if it has probability density function

$$f(x;\theta_1,\theta_2,\theta_3) = \begin{cases} \left(\frac{2}{\Pi}\right)^{\frac{1}{2}} (\theta_2+\theta_3)^{-1} \exp\left\{\frac{-(x-\theta_1)^2}{2\theta_2^2}\right\}, & x \le \theta_1 \\ \left(\frac{2}{\Pi}\right)^{\frac{1}{2}} (\theta_2+\theta_3)^{-1} \exp\left\{\frac{-(x-\theta_1)^2}{2\theta_3^2}\right\}, & x > \theta_1 \end{cases} \quad (3.4)$$

This distribution was introduced as the joined half-Gaussian by Gibbons and Mylroie (1973) who found it to be a very good fit to impurity profiles data in ion-implantation research. John (1982) discusses estimation of the parameters of the TPN distribution by the method of maximum likelihood and the method of moments. In this paper we examine the asymptotic properties of some methods for TPN distribution.

Let $X_1, X_2, \ldots\ldots\ldots\ldots, X_n$ be a random sample from the TPN distribution with density (3.4). For the MLE's $\hat{\theta}_1, \hat{\theta}_2$ *and* $\hat{\theta}_3$ of θ_1, θ_2 *and* θ_3 respectively, differentiating log-likelihood of (3.4) with respect to θ_1, θ_2 *and* θ_3 we have MLE's of θ_1, θ_2 *and* θ_3

$$\hat{\theta}_1 = \frac{\Sigma x_i}{n}$$

$$\hat{\theta}_2 = \left\{\frac{\Sigma(x_i-\bar{x})^2}{n-\Sigma(x_i-\bar{x})^2}\right\}^{\frac{1}{2}}$$

$$\hat{\theta}_3 = n\left(\frac{\left\{\Sigma\left(x_i-\bar{x}\right)^2\right\}^{\frac{1}{2}} - \left\{\Sigma\left(x_i-\bar{x}\right)^2\right\}^{\frac{3}{2}}}{\left\{n-\Sigma\left(x_i-\bar{x}\right)^2\right\}^{\frac{3}{2}}}\right)$$

Large sample approximations for the variance-covariance matrix of the MLE's are obtained by the inverse Fisher information matrix. The sample information matrix is

$$B_n = \begin{pmatrix} n(\theta_2^{-2}+\theta_3^{-2}) & 2\theta_2^{-3}\Sigma(x_i-\theta_1) & 2\theta_3^{-3}\Sigma(x_i-\theta_1) \\ 2\theta_2^{-3}\Sigma(x_i-\theta_1) & 3\theta_2^{-4}\Sigma(x_i-\theta_1)^2 - n(\theta_2+\theta_3)^{-2} & -n(\theta_2+\theta_3)^{-2} \\ 2\theta_3^{-3}\Sigma(x_i-\theta_1) & -n(\theta_2+\theta_3)^{-2} & 3\theta_3^{-4}\Sigma(x_i-\theta_1)^2 - n(\theta_2+\theta_3)^{-2} \end{pmatrix}$$

Large sample variance approximations of the MLE's are most useful when the estimators are asymptotically normal. This is not immediately obvious in the case of the TPN distribution. To show that the MLE's are asymptotically normal it is sufficient to consider the case of estimating θ_1 with θ_2 and θ_3 known. In this example, we see that all the regularity conditions given in the earlier sections for the asymptotic normality of $\hat{\theta}_1$ are satisfied except that involving the third derivative of the log-likelihood of (3.4).

According to John (1982), to prove this normality result it is sufficient to check that

$$d_n = n^{-1}\left[\ell''(\theta_1 + kn^{-\frac{1}{2}}\ \ell''(\theta_1)\right] \xrightarrow{P} 0$$

as $n \to \infty$, where k is a constant and ρ" denotes the second derivative with respect to θ_1 of the log-likelihood. Thus $d_n \to 0$ in probability and hence

$n^{-\frac{1}{2}}(\hat{\theta}_1 - \theta_1) \xrightarrow{D} N[o, I^{-1}(\theta)]$ *as* $n \to \infty$, where I(θ) in this case is equal to $(\theta^2\theta^3)^{-1}$.

The MM estimates of θ_1, θ_2 and θ_3 are

$$\theta_1^+ = \bar{x} - (2/\pi)^{\frac{1}{2}} (\theta_3 - \theta_2), \qquad \theta_2, \theta_3 \text{ known}$$

$$\theta_2^+ = \theta_3 + (\pi/2)^{\frac{1}{2}} (\bar{x} - \theta_1), \qquad \theta_1, \theta_3 \text{ known}$$

By using large sample properties of MLE's, the efficiencies of these estimators relative to the MLE's are

$$E_1 = \frac{Var(\hat{\theta}_1)}{Var(\theta_1^+)} = \left[1 + \frac{(\pi+2)}{\pi}\left\{\frac{\theta_2}{\theta_3} - 2 + \frac{\theta_3}{\theta_2}\right\}\right]^{-1}$$

$$E_2 = \frac{Var(\hat{\theta}_2)}{Var(\theta_2^+)} = 2\left(\frac{\theta_3}{\theta_2} + 1\right)^2\left(\frac{3\theta_3}{\theta_2} + 2\right)^{-1} \left[(\pi - 2)\left(\frac{\theta_3}{\theta_2} - 1\right)^2 + \frac{\pi\theta_3}{\theta_2}\right]^{-1}$$

Numerical values of these efficiencies are obtain by some given parameter values. The plots of the efficiencies against some parameters values are shown in figure-1.

REFERENCES

1) **LeCam, L. (1953).** *On some asymptotic properties of maximum likelihood estimates and related Baye's estimates.* University of California Publication in Stat., 1,277-330.

2) **Rao, C.R.(1963).** *Criteria of estimation in large samples.* Sankhya, 25 A, 189-206

3) **Cox, D.R and Hinkley, D.V.(1974).** *Theoretical Statistics.* Chapman & Hall, London.

4) **Cramer, H. (1946).** *Mathematical methods of statistics,* Princeton University Press, Princeton.

5) **Gurland, J.(1954).** *On regularity conditions for maximum likelihood estimators.* Skand. Aktuarietidskr, Vol. 37, pp.71-76

6) **Wald, A.(1949).** *Note on the consistency of maximum likelihood estimate,* Ann. Math. Statist. 20,595-601

7) **Wolfowitz, J. (1949).** *On Wald's proof of the consistency of the maximum likelihood estimate.* Ann.Math.Statist. Vol.20,pp.601-602

8) **Kuldorff, G.(1957).** *On the conditions for consistency and asymptotic efficiency of maximum likelihood estimators,* Skand. Akluarietidskr., Vol.40,pp.130-144.

9) **Doss, S.A.D.C.(1962)** *A note on consistency and Asymptotic Efficiency of maximum likelihood estimates in multi-parametric problems.* Calcutta Statist. Association Bull 11, 85-93

10) **Daniels, H.E (1961).** *The asymptotic efficiency of a maximum likelihood estimator.* Proc. Fourth Berkeley Symp. *on Math. Statist. and Prob.* 1,Vol.1,151-163.

11) **Gibbons, J.F and Mylories, S.(1973)** *Estimation of impurity profiles in ion-implanted amorphous targets using joined half-Gaussian distributions.* Appl.Phys.Lett. 22,568-569.

12) **John, S.(1982)** *The three-parameter two-piece normal family of distributions and its fitting.* Common. Statist. Theor. Meth., 11,879-885.

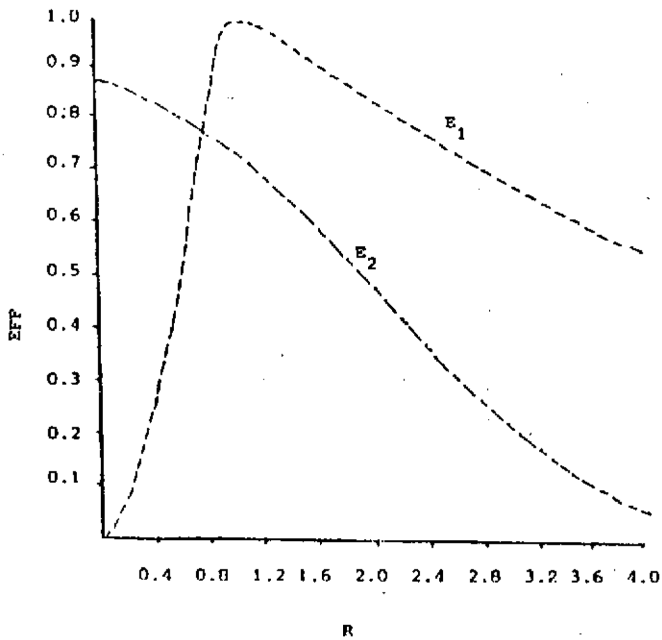

Fig.1 Plots for the efficiencies of MM estimators relative to the MLE's against some paramaters values ($R = \theta_2/\theta_3$).