Journal of the Korean Statistical Society
Vol. 20, No. 2, 1991

A Short Note on Superefficiency

Youngjo Lee* and Byeong U. Park**

ABSTRACT

In Le Cam's earlier work on superefficiency, it is proved that if an estimate is superefficient at a given parameter value θ_0, then there must exist an infinite sequence $\{\theta_n\}$ of values(converging to θ_0) at which this estimate is worse than M.L.E. for certain classes of loss functions. For one-dimensional cases, these classes of loss functions include squared error loss. However, for multi-dimensional cases, they do not. This note is to give an example where a superefficient estimator of a multi-dimensional parameter is not inferior to M.L.E. along any sequence $\{\theta_n\}$ converging to the point of superefficiency with respect to the squared error loss.

1. Introduction

The two main points related to *superefficient* estimators which have been explored by Le Cam (1953), are as follows.

1. The set of superefficiency must be of Lebesque measure zero.
2. If an estimate is superefficient at a given parameter value θ_0, then there must exist an infinite sequence $\{\theta_n\}$ converging to θ_0 at which this estimate is worse than M.L.E. for certain classes of loss functions.

Here 2 means that good performance at one point θ_0 entails certain unpleasant properties of the risk of the estimators in the neighborhood of θ_0.

The classes of loss functions considered in 1 and 2 above are broad enough to include every reasonable loss functions in one-dimensional cases. In particular, we only need to require that the loss function $L(x)$ satisfies

1) L is bounded below
2) As a function of λ

$$\int L(x)\exp\{-\frac{1}{2}c(x-\lambda)^2\}dx$$

* Department of Statistics, Hallym University, Chunchon
** Department of Computer Science and Statistics, Seoul National University, Seoul

has a minimum at $\lambda=0$ for any positive number c contained in some closed interval.
Remark The squared error loss function $L(x)=x^2$ satisfies the above two conditions.
In the case of a multi-dimensional parameter, additional conditions should be imposed on the loss function to keep 2 valid. One of sufficient conditions which entail 2 is that $L(x)$ depends only on a linear combination of the coordinates of x(c.f. Le Cam 1953, page 327).
In this paper we give an example which shows that 2 is invalid if we use squared error loss in the case of a multi-dimensional parameter. In particular, we show that the shrinkage estimator is superefficient at some point and the limit of its risk(with respect to the squared error loss) along any sequence of parameter values converging to the point of superefficiency is not greater than that of M.L.E.

2. Superefficiency of the Shrinkage Estimator

For a p-dimensional parameter θ, a superefficient estimate of θ is defined as follows(cf. Le Cam, 1953).

Definition Let $A(T_1, T_2, \theta)=\lim_{n\to\infty}[R_n(T_{n1}, \theta)-R_n(T_{n2}, \theta)]$ where $R(T_n, \theta)=E_\theta L_n(T_n, \theta)$ and L_n is a loss function which may depend on n. An estimate T_n is called *superefficient* with respect to $\{L_n\}$ if $A(T, \delta_M, \theta)\leq 0$ for every $\theta\in\Theta$ and $A(T, \delta_M, \theta')<0$ for at least one value $\theta'\in\Theta$ where δ_M is the M.L.E. of $\mu(\theta)$.

In Theorem 14 of Le Cam (1953), it is asserted that for one-dimensional parameters, if T_n is superefficient at a point θ_0, then there must exist an infinite sequence $\{\theta_n\}$ which converges to θ_0 and

$$\varliminf_{n\to\infty}[R_n(T_n, \theta_n)-R_n(\delta_{M_n}, \theta_n)]>0$$

for certain classes of loss functions. This tells us that the reduction in risk at one point is balanced by an increase in risk in the neighborfood of that point. However, as indicated in Le Cam(1953), this result can not be extended to multi-dimensional case in obvious way. In particular, some severe restrictions should be imposed on the loss functions. In this section, we will see that for the squared error loss

$$L_n(T_n, \theta)=n\sum_{i=1}^{p}(T_{n_i}-\mu_i(\theta))^2 \tag{2.1}$$

the result is no longer valid for multi-dimensional cases.
First we will introduce some techniques to handle the expectation of functions of noncentral χ^2 random variables. This is based on the work by Peixoto(1982). Let $X\sim N_p(\theta, I_p)$. Then $\|\underline{X}\|^2$ is distributed as a noncentral χ^2 with p degrees of freedom and noncentrality parameter $\lambda=\|\underline{\theta}\|^2/2$.
Define Z to be a poisson random variable with parameter λ. The following three lemmas will be useful to calculate the expected values of some functions of the normal random variable X through those of the poisson random variable Z. Assume p is greater than two. Proofs of the following results are given in Peixoto(1982).

Lemma 2.1

$$E[f(\|\underline{X}\|^2)]=E[g(Z)]$$

where $g(Z)=E\{E[f(\chi^2_{p+2Z}) \mid Z]\}$.

Therefore, we can have

$$E_0(1/\|\underline{X}\|^2)=E_\lambda[1/(p+2Z-2)]$$

and

$$E_0[(1/\|\underline{X}\|^2)^2]=\lambda E_\lambda[1/(p+2Z-2)(p+2Z-4)].$$
$$1/(p+2Z-2)(p+2Z-q)].$$

Lemma 2.2

$$E_\lambda[Zg(Z)]=\lambda E_\lambda[g(Z+1)].$$

Using Lemma 2.1, it is not so hard to get the following lemma.

Lemma 2.3

$$E_0[X_i\, f(\|\underline{X}\|^2)]=\theta_i\, E_\lambda[g(Z+1)],$$

$$E_0[X_i^2\, f(\|\underline{X}\|^2)]=E_\lambda[g(Z+1)]+\theta_i^2 E_\lambda[g(Z+2)],$$

and for $i\neq j$

$$E_\theta[X_i\, X_j\, f(\|\underline{X}\|^2)]=\theta_i\theta_j E_\lambda[g(Z+2)].$$

If $\lambda=0$, then each θ_i is zero. Therefore, the above expectation will become zero. With the help of the above three lemmas, we can obtain the following formula:

$$E_0(X_i/\|\underline{X}\|^2)=\theta_i\, E_\lambda[1/(p+2Z)], \tag{2.2}$$

$$E_0[X_i/\|\underline{X}\|^4]=\theta_i\, E_\lambda[1/(p+2Z)(p+2Z-2)], \tag{2.3}$$

$$E_0[X_i^2/\|\underline{X}\|^2]=\begin{cases} 1/p & if\, \lambda=0 \\ E_\lambda[(1+\theta_i^2\, Z/\lambda)/(p+2Z)] & if\, \lambda\neq 0, \end{cases} \tag{2.4}$$

$$E_0[X_i^2/\|X\|^4]=\begin{cases} 1/p(p-2) & if\, \lambda=0 \\ E_\lambda[(1+\theta_i^2\, Z/\lambda)/(p+2Z)(p+2Z-2)] & if\, \lambda\neq 0, \end{cases} \tag{2.5}$$

$$E_0[X_i\, X_j/\|\underline{X}\|^2]=\theta_i\, \theta_j E_\lambda[1/(p+2Z+2)], \tag{2.6}$$

$$E_0[X_i\, X_j/\|\underline{X}\|^4]=\theta_i\, \theta_j E_\lambda[1/(p+2Z)(p+2Z+2)], \tag{2.7}$$

Here the equations (2.6) and (2.7) hold for $i \neq j$. Now let $\{\underline{\bar{X}}_n\}$ be a sequence of random variables such that

$$\underline{\bar{X}}_n \sim N_p(\underline{\theta},\ I_p/n).$$

Then the James-Stein estimator of $\underline{\theta}$ based on $\underline{\bar{X}}_n$ is

$$\delta_s^{(n)}(\underline{\bar{X}}_n) = \underline{\bar{X}}_n - (p-2)\underline{\bar{X}}_n/(n\|\underline{\bar{X}}_n\|^2),$$

where

$$n\|\underline{\bar{X}}_n\|^2 \sim \chi^2(p, n\|\underline{\theta}\|^2/2).$$

Let $\delta_{s.i}^{(n)}$ be the i-th coordinate of $\delta_s^{(n)}$. Then

$$E_\theta\{n^{1/2}[\delta_{s.i}^{(n)}(\underline{\bar{X}}_n) - \theta_i] \cdot n^{1/2}[\delta_{s.j}^{(n)}(\underline{\bar{X}}_n) - \theta_j]\}$$
$$= Cov(n^{1/2}\delta_{s.i}^{(n)}(\underline{\bar{X}}_n),\ n^{1/2}\delta_{s.j}^{(n)}(\underline{\bar{X}}_n))$$
$$+ E_\theta[(p-2)n^{1/2}\bar{X}_{n.i}/(n\|\underline{\bar{X}}_n\|^2)] \cdot E_\theta[(p-2)n^{1/2}\bar{X}_{n.j}/(n\|\underline{\bar{X}}\|^2)].$$

By (2.2) and the fact that

$$E_\lambda[1/(p+2Z)] \leq (2\lambda)^{-1}(1-e^{-\lambda})$$

where Z is poisson random variable with prameter λ, we have for $\|\underline{\theta}\|^2 \neq 0$

$$|E_\theta[n^{1/2}\bar{X}_{n.i}/(n\|\underline{\bar{X}}_n\|^2)]| \leq (2\lambda_n)^{-1}n^{1/2}\ |\theta_i|\ (1-e^{-\lambda_n}) \to 0$$

as $n \to \infty$ where $\lambda_n = n\|\underline{\theta}\|^2/2$. Therefore

$$\lim_{n\to\infty} E_\theta\{n^{1/2}[\delta_{s.i}^{(n)}(\underline{\bar{X}}_n) - \theta_i] \cdot n^{1/2}[\delta_{s.j}^{(n)}(\underline{\bar{X}}_n) - \theta_j]\}$$
$$= \lim_{n\to\infty} Cov(n^{1/2}\delta_{s.i}^{(n)}(\underline{\bar{X}}_n),\ n^{1/2}\delta_{s.j}^{(n)}(\underline{\bar{X}}_n)).$$

Furthermore since

$$n^{1/2}\bar{X}_{n.i}/(n\|\underline{\bar{X}}_n\|^2) = O_p(n^{-1/2}) \text{ if } \|\underline{\theta}\|^2 \neq 0,$$

we have

$$\lim_{n\to\infty} E_\theta\{n[\delta_s^{(n)}(\underline{\bar{X}}) - \underline{\theta}][\delta_s^{(n)}(\underline{\bar{X}}_n) - \underline{\theta}]^t\} = I_p \qquad \text{if } \|\underline{\theta}\|^2 \neq 0.$$

When $\|\underline{\theta}\|^2 = 0$, $n^{1/2}\delta_s^{(n)}(\underline{\bar{X}}_n)$'s have the same distribution for all n since $n^{1/2}\underline{\bar{X}}_n \sim N_p(0,\ I_p)$ for each n. For this case note that

$$\lim_{n\to\infty} E[n^{1/2}\delta_{s,i}^{(n)}(\underline{\bar{X}}_n)]\,[n^{1/2}\delta_{s,j}^{(n)}(\underline{\bar{X}}_n)] = \begin{cases} 0 & \text{if } i\neq j \\ 2/p & \text{if } i=j \end{cases}$$

since

$$E_0(n^{1/2}\underline{\bar{X}}_{n,i}\cdot n^{1/2}\underline{\bar{X}}_{n,j}) = \begin{cases} 0 & \text{if } i\neq j \\ 1 & \text{if } i=j, \end{cases}$$

$$E_0[n^{1/2}\underline{\bar{X}}_{n,i}\cdot n^{1/2}\underline{\bar{X}}_{n,j}/(n\|\underline{\bar{X}}_n\|^2)] = \begin{cases} 0 & \text{if } i\neq j \\ 1/p & \text{if } i=j, \end{cases} \tag{2.8}$$

and

$$E_0[n^{1/2}\underline{\bar{X}}_{n,i}\cdot n^{1/2}\underline{\bar{X}}_{n,j}/(n(\|\underline{\bar{X}}_n\|^2)^2] = \begin{cases} 0 & \text{if } i\neq j \\ 1/p(p-2) & \text{if } i=j. \end{cases} \tag{2.9}$$

Here (2.8) follows from (2.4) and (2.6), and (2.9) from (2.5) and (2.7). Therefore,

$$\lim_{n\to\infty} E_\theta\{n^{1/2}[\delta_s^{(n)}(\underline{\bar{X}}_n)-\underline{\theta}]\cdot n^{1/2}[\delta_s^{(n)}(\underline{\bar{X}}_n)-\underline{\theta}]'\} = \begin{cases} I_p & \text{if } \|\underline{\theta}\|^2\neq 0 \\ 2I_p/p & \text{if } \|\theta\|^2=0. \end{cases}$$

From the above equation, we can see that the James-Stein estimator is a superefficient estimator with respect to the squared error loss defined in (2.1) when $p\geq 3$.

Now consdier any sequence of nonzero p-dimensional vectors $\{\underline{\theta}_n\}$ converging to $\underline{0}$. When $n^{1/2}\underline{\theta}_n\to\underline{c}(\underline{c}\neq\underline{0})$ or $n\|\underline{\theta}_n\|^2\to\infty$, then by similar arguments as in the case of fixed $\underline{\theta}$ with $\|\underline{\theta}\|\neq 0$, we can get

$$\lim_{n\to\infty} E_\theta\{n[\delta_s^{(n)}(\underline{\bar{X}}_n)-\underline{\theta}_n][\delta_s^{(n)}(\underline{\bar{X}}_n)-\underline{\theta}_n]'\}=I_p. \tag{2.10}$$

when $n^{1/2}\underline{\theta}_n\to\underline{0}$, we have

$$\lim_{n\to\infty} E_{\theta_n}\{n^{1/2}[\delta_{s,j}^{(n)}(\underline{\bar{X}}_n)-\theta_{ni}]n^{1/2}[\delta_{s,j}^{(n)}(\underline{\bar{X}}_n)-\theta_{nj}]\} = \begin{cases} 0 & \text{if } i\neq j \\ d & \text{if } i=j. \end{cases} \tag{2.11}$$

Here $d<1$ if $p\geq 3$. The identity for the case $i\neq j$ in (2.11) follows from applying (2.6) and (2.7). The identity for the case $i=j$ follows from the facts

$$\lim_{n\to\infty} E_{\theta_n}[n\underline{\bar{X}}_{n,i}^2]=1,$$

$$\lim_{n\to\infty} E_{\theta_n}[\underline{\bar{X}}_{n,i}^2/(n\|\underline{\bar{X}}_n\|^4)]=0$$

and

$$\lim_{n\to\infty} E_{\theta_n}[\underline{\bar{X}}_{n,i}^2/\|\underline{\bar{X}}_n\|^2] = \lim_{n\to\infty} E_{\lambda_n}[(1+n\theta_{ni}^2 Z/\lambda_n)/(p+2Z)] > \lim_{n\to\infty} E_{\lambda_n}[p+2Z]^{-1}>0$$

where $\lambda_n=n\|\underline{\theta}_n\|^2/2$. Thus from (2.10) and (2.11)

$$\lim_{n\to\infty}[R_n(\delta_s^{(n)}, \underline{\theta}_n)-R_n(\delta_{M_n}, \underline{\theta}_n)]\leq 0$$

for any sequence $\{\underline{\theta}_n\}$ converging to $\underline{0}$.

References

1. Le Cam. L(1953). On the asymptotic properties of a maximum likelihood estimates and related Bayes estimates. University of California Publications in Statist., 1, 277-300.
2. Peixoto, J.O.(1982). Estimation of random effects in the balanced one way classification. Thesis, Iowa State University.
3. Zacks, S.(1971). *The Theory of Statistical Inference*, Wiley, New York.