韓國數學敎育學會誌〈數學敎育〉
1974. 12. Vol. XIII No. 2

Journal of the Korea Society of Mathematical Education
Dec. 1974. Vol. XIII No. 2

SOME OBSERVATIONS ON ROBUST ESTIMATION

by

Han Shick Park

Seoul National University, Seoul Korea

§1. INTRODUCTION AND SUMMARY

The problem of finding location estimators which are robust against deviations from normality has received increasing attention in the last several years. See, for example, Huber [10], and papers cited therein.

Hardly anybody realized how bad the classical estimates could be in slightly nonnormal situations. E.S.Pearson may have been the first to note the high sensitivity to deviations from normality of some standard procedures (tests for equality of variances): incidentally, in connection with the same test problems, G.E.P.Box [3] later coined the term robustness.

From the begining, robustness has been a rather vague concept; for example, Box and Anderson [4] had introduced the notion as follows: Procedures are required which are robust (insensitive to changes in extraneous factors not under test) as well as powerful (sensitive to specific factors cunder test).

But if one wants to choose in a rational fashion between different robust competitors to a lassical procedure, one has to make precise the goals one wants to achieve. Unfortunately, a consensus has not been reached at least five or six conflicting ones, and we does not think that all of them should be called by the same name robust.

Hence we will set up the concept of robust as follows. The robust estimator should possess a small asymptotic variance over some neighborhood of one shape, in particular of the normal one [9] or that the distribution of estimate should change little under arbitrary small variances of the underlying distribution F, and this uniformly in the sample size n[7].

During the last several years various approaches have been proposed to deal with the lack of robustness of the sample mean as an estimate of the population mean when the distribution sampled is contaminated by gross errors, i.e., has heavier tails than the normal distribution. First, Tukey and the Statistical Reserch Group at Princeton suggested and investigated the properties of trimmed and Winsorized means [13].

More recently, Hodges and Lehmann [8], proposed estimates related to the robust Wilcoxon and normal scores tests, among others. Huber in [9] considered essentially the class of maximum likelihood estimates and found those members of this class which minimize the maximum variance over various classes of contaminated distributions. For a review of work in these directions in estimation the reader is referred to [1], [10].

Most of the estimators commonly studied are, under suitable regularity conditions, asymptotically normal about the center of symmetry, with asymptotic variance depending on the underlying distribution. We thus have a simple criterion, the asymptotic variance, for comparing the perform ance of different estimators for a given underlying distribution, and of a given estimator for diffe rent underlying distributions. Huber [9] has formulated and solved some minimax problems, in which the estimators are judged by their asymptoic variance.

In this paper we show that some robust location estimators has a same property asymptotically under some conditions.

In section 2 we define and state asymptotic variances which have been found for the three most commonly studied types of location estimators, i.e., M-estimator, L-estimator and R-estimator.

In section 3 we demonstrate some relationships among the three types of estimators, i.e., under some conditions the asymptotic variances of these estimators are equal to each other and M-estimator and the corresponding L-estimator is very closer.

§ 2. ESTIMATORS

Let $X_1, X_2, \cdots, X_n$ be independent random variables with common distribution $F(x-\theta)$, where F is symmetric, that is $F(x)=1-F(-x)$.

We assume F has a density f. The problem is to estimate the unknown parameter θ and judge the quality of an estimator by its asymptotic variance. Each of the three types of estimators defined below is, under general regularity conditions, asymptotically normal with mean θ and asymptotic variance as given below. Since we shall be dealing with translation-invariant statistics, we shall henceforth assume that $\theta=0$.

Note that the estimators defined here are actually sequences of estimators indexed by n, the sample size, but we shall simply think of them as single estimators, without an index, whenever no confusion will arise thereby.

2-1 M-estimators

Maximum likelihood type estimators, which we shall call M-estimators. Let ρ be a real valued function of real parameter, with derivative $\psi=\rho'$.

Define M either by

$$\sum_{i=1}^{n}\rho(X_i-M)=inf_m \sum_{i=1}^{n}\rho(Xi-m)$$

or by

$$\sum_{i=1}^{n}\psi(X_i-M)=0$$

If $\psi(x)$ is monotonic, M is essentially uniquely determined. And $\sqrt{n}M$ is asymptotically normal with asymptotic variance

$$\sigma_M^2(F)=\frac{E(\psi^2)}{(E(\psi'))^2}=\frac{\int_{-\infty}^{\infty}\psi^2(x)f(x)dx}{\left(\int_{-\infty}^{\infty}\psi'(x)f(x)dx\right)^2}$$

See Huber [9]

2-2 L-estimators

An L-estimator, or linear combination of order statistics, is defind as follows:

Let $h(t)$ be defined on $[0,1]$ and such that

$$\int_0^1 h(t)\,dt=1 \qquad \text{and} \qquad h(1-t)=h(t).$$

Let $X_{(1)}\leqq\cdots\leqq X_{(n)}$ be the order statistics derived from the sample.

Let

$$i^*=\frac{i}{n+1}$$

Define L as

$$L=\frac{1}{n}\sum_{i=1}^{n}h(i^*)X_{(i)}$$

Under some regularity conditions, $\sqrt{n}L$ is asymptotically normal with asympotic variance

$$\sigma_L^2(F)=\int_0^1 U^2(t)\,dt,$$

where

$$U(t)=\int_{\frac{1}{2}}^{t}\frac{h(u)}{f[F^{-1}(u)]}\,du$$

See Chernoff, Gastwirth and Johns [5], and Jaeckel [11].

2-3 R-estimators

Estimators derived from rank tests, which we shall call R-estimators.

Let $J(t)$ be such that $J(1-t)=-J(t)$.

For any given r, form the $2n$ numbers

$$X_1-r,\cdots,\ X_n-r;\ -X_1+r,\cdots,\ -X_n+r.$$

Order these $2n$ numbers and let $V_i=1$ if the ith smallest is of the form X_j-r, and $V_i=0$ otherwise.

Form the sum

$$W(r)=\sum_{i=1}^{2n}J\left(\frac{i}{2n+1}\right)V_i.$$

Define R as a solution of the equation $W(R)=0$.

If J is monotonic, R is essentially uniquely determined. The asymptotic variance of $\sqrt{n}R$ is

$$\sigma_R^2(F)=\frac{\int_{-\infty}^{\infty}J^2(t)\,dt}{\left(\int_{-\infty}^{\infty}\frac{d}{dx}J\{F[(x)]\}f(x)\,dx\right)^2}$$

See Hodges and Lehmann [8] and Jaeckel [11].

§3. RELATIONSHIPS AMONG THE ESTIMATORS

For any given symmetric F satisfying appropriate regularity conditions, there is a three-way correspondence among the three types of estimators which preserves the asymptotic variance under F.

Let $\psi(x)$ be such that

$$\psi(-x)=-\psi(x)$$

for ψ defining an M-estimator.

And we defining an L-estimator and an R-estimator as follows.

Let $t=F(x)$, so that $x=F^{-1}(t)$.

We assume for simplicity that F is strictly increasing.

Let

$$h(t)=\psi'[F^{-1}(t)]=\psi'(x),$$

where

$$\psi'(x)=\frac{d}{dx}\psi(x).$$

Let

$$J(t)=\psi[F^{-1}(t)]=\psi(x).$$

The symmetry conditions on h and J,

$$h(1-t)=h(t), \qquad J(1-t)=-J(t)$$

are satisfied as belows.

$$\begin{aligned} h(1-t)&=\psi'[F^{-1}(1-t)]=\psi'(-x)\\ &=\psi'(x) \qquad [\because\ \psi(-x)=-\psi(x)]\\ J(1-t)&=\psi[F^{-1}(1-t)]=\psi(-x)\\ &=-\psi(x)=-J(t) \end{aligned}$$

And we may assume that

$$\int_{-\infty}^{\infty}\psi'(x)f(x)\,dx=1.$$

This condition implies that

$$\begin{aligned} \int_0^1 h(t)\,dt&=\int_{-\infty}^{\infty}h[F(x)]f(x)\,dx\\ &=\int_{-\infty}^{\infty}\psi'(x)f(x)\,dx=1 \end{aligned}$$

THEOREM 1. *Assuming the asymptotic variance formulas hold, we have, for the estimators defined above,*

$$\sigma_M^2(F)=\sigma_L^2(F)=\sigma_R^2(F).$$

PROOF: Since

$$\int_{-\infty}^{\infty}\psi'(x)f(x)\,dx=1,$$

$$\sigma_M^2(F)=\int_{-\infty}^{\infty}\psi^2(x)f(x)\,dx.$$

For the L-estimator we have

$$\begin{aligned} U(t)&=\int_{\frac{1}{2}}^{t}\frac{h(u)}{f[F^{-1}(u)]}du\\ &=\int_0^{F^{-1}(t)}\frac{h[F(x)]}{f(x)}f(x)\,dx\\ &=\int_0^{F^{-1}(t)}\psi'(x)\,dx=\psi[F^{-1}(t)], \end{aligned}$$

since $\psi(0)=0$. Therefore,

$$\sigma_L^2(F)=\int_{-\infty}^{\infty}U^2(t)\,dt=\int_{-\infty}^{\infty}\psi^2[F^{-1}(t)]\,dt$$
$$=\int_{-\infty}^{\infty}\psi^2(x)f(x)\,dx=\sigma_M^2(F).$$

Since the denominator for the variance of the R-estimator is the square of

$$\int_{-\infty}^{\infty}\frac{d}{dx}\{J(F(x))\}f(x)\,dx=\int_{-\infty}^{\infty}\left[\frac{d}{dx}\psi(x)\right]f(x)\,dx=1,$$

we have

$$\sigma_M^2(F)=\int_{-\infty}^{\infty}J^2(t)\,dt$$
$$=\int_{-\infty}^{\infty}\psi^2(x)f(x)\,dx=\sigma_M^2(F).\qquad\text{(Q.E.D.)}$$

If we introduce the function ψ, displaying explicitly the parameter k occurring in it

$$\psi(x)=x \qquad \text{for } |x|<k$$
$$=k\,\mathrm{sign}(x) \qquad \text{for } |x|\geqq k,$$

we have

$$h(t)=\text{constant} \qquad \text{for } F(-k)<t<F(k,)$$
$$=0 \qquad \text{otherwise.}$$

This is the trimmed mean with trimming proportion $\alpha=F(-k)$ in the sense of Tukey. That is: Tukey defines the α-trimmed mean of the sample by

$$\overline{X}_\alpha=\frac{1}{n-2[\alpha n]}\sum_{i=[\alpha n]+1}^{n-[\alpha n]}X_{(i)}$$

where $[\alpha n]$ is the greatest integral in αn for $0\leqq\alpha<\frac{1}{2}$.

The equality of variances in this case was recognized by Bickel [2].

The corresponding R-esti mattor is defined by

$$J(t)=F^{-1}(t) \quad \text{for} \quad F(-k)<t<F(k)$$
$$=k \quad \text{for} \quad t\geqq F(k)$$
$$=-k \quad \text{for} \quad t\leqq F(-k)$$

If F is the contaminated normal distribution considered by Huber, this $J(t)$ defines a sort of truncated Van der Waerden test. See Lindgren [12] and Gastwirth [6].

It follows from the theorem that if any one of three estimators is asymptotically optimal for F among translation-invariant estimators, then all three are. Correspondences of this type have been ngive in the asymptotically optimal case by Gastwirth [6].

Jaeckel [11] have investigated that under some more restrictive conditions the relationship between the M-estimator and the corresponding L-estimator is even closer than that indicated above. We restate some definition, a lemma and a theorem from Jaekel [11]

DEFINITION. (1) The sequence of random variables $\{Z_n\}$ is bounded in probability if

$$\forall\delta>0\ \exists B,N \text{ such that } \forall n\geqq N\colon P\{|Z_n|\leqq B\}\geqq 1-\delta.$$

(2) The sequence of random variables $\{Z_n\}$ is bounded in probability uniformly in k if

$$\forall\delta>0\ \exists B,N \text{ such that } \forall n\geqq N\colon P\{|Z_{nk}|\leqq B,\ \forall k\}\geqq 1-\delta.$$

(3) The sequence $\{Z_n\}$ ($\{Z_{nk}\}$) is $O(n^3)$ in probability (uniformly in k) if the sequence $\left\{\frac{Z_n}{n^a}\right\}$ $\left(\left\{\frac{Z_{nk}}{n^a}\right\}\right)$ is bounded in probability (unformly in k).

DEFINITION. The sample distribution function is

$$\begin{aligned} F_n(x) &= \frac{i}{n} && \text{for } X_{(i)} < x < X_{(i+1)},\ i=1,\ \cdots,\ n-1 \\ &= 0 && \text{for } x < X_{(1)} \\ &= 1 && \text{for } x > X_{(n)} \\ &= i^* = \frac{i}{n+1} && \text{for } x = X_{(i)},\ i=1,\ \cdots,\ n. \end{aligned}$$

LEMMA. *Suppose F has a density $f(x)$, and there are number $\alpha_0>0$, $\epsilon_0>0$, and $f_0>0$ such that $f(x)\geqq f_0$ for all x such that $\alpha_0-\epsilon_0\leqq F(x)\leqq 1-(\alpha_0-\epsilon_0)$. Then $X_{(i)}-F^{-1}(i^*)$ is $O\left(\frac{1}{\sqrt{n}}\right)$ in probability uniformly in $i=[\alpha_0 n]+1,\cdots,\ n-[\alpha_0 n]$ That is, for all $\delta>0$ there exist D and Nsuch that for all $n\geqq N$:*

$$P\{|X_{(i)}-F^{-1}(i^*)|\leqq \frac{D}{\sqrt{n}f_0} \text{ for } i=[\alpha_0 n]+1,\cdots,n-[\alpha_0 n]\}\geqq 1-\delta.$$

PROOF. The statistic

$$K_n = \sqrt{n}\,sup|F_n(x)-F(x)|$$

has a limiting distribution which was found by Kolmogorov. For a given δ, we can therefore choose a D so that for sufficiently large n,

$$P\{K_n\leqq D\}\geqq 1-\delta.$$

Suppose $K_n\leqq D$; that is,

$$|F_n(x)-F(x)|\leqq \frac{D}{\sqrt{n}} \qquad \text{for all } x.$$

Then

$$|i^*-F(x_{(i)})|\leqq \frac{D}{\sqrt{n}} \qquad \text{for } i=1,2,\cdots,n.$$

Since $([\alpha_0 n]+1)^*\to\alpha_0$ and $(n-[\alpha_0 n])^*\to 1-\alpha_0$,

and $\frac{D}{\sqrt{n}}<\frac{1}{2}\epsilon_0$ for sufficiently large n,

we have, for lrage n,

$$\alpha_0-\epsilon_0<F(X_{(i)})<1-(\alpha_0-\epsilon_0) \text{ for } i=[\alpha_0 n]+1,\cdots,\ n-[\alpha_0 n].$$

Since, for $\alpha_0-\epsilon_0<t<1-(\alpha_0-\epsilon_0)$

we have

$$\frac{d}{dt}F^{-1}(t)=\frac{1}{f(F^{-1}(t))}\leqq\frac{1}{f_0},$$

we can apply the mean value theorem to $F^{-1}(t)$, obtaining

$$|F^{-1}(i^*)-F^{-1}(F(X_{(i)}))|\leqq\frac{1}{f_0}|i^*-F(X_{(i)})|\leqq\frac{D}{\sqrt{n}f_0}$$

for $i=[\alpha_0 n]+1,\cdots,\ n-[\alpha_0 n]$. (Q.E.D.)

THEOREM 2. *Suppose E has a bounded density f and satisfies the conditions of Lemma for some α_0. Suppose $\psi(x)$ and $h(t)$ are related by*

$$h(t)=\psi'(F^{-1}(t))=\psi'(x),$$

the asymptotic variance formulas apply, and

$$h(t)=\psi'(x)=0 \qquad \text{for } t<\alpha_0 \text{ and } t>1-\alpha_0.$$

Finally, suppose that ψ is continuous, and tha', at all but a finite number of points ψ' is defined and bounded and has a bounded derivative. Then

$$\sqrt{n}\,(M-L)\to_p O.$$

In fact, $M-L$ is $O\left(\frac{1}{n}\right)$ in probability.

PROOF. If we write

$$X_{(i)}=F^{-1}(i^*)+e_i,$$

we have, by Lemma, e_i is $O\left(\frac{1}{\sqrt{n}}\right)$ in probability uniformly in i such that $\alpha_0\leqq i^*\leqq 1-\alpha_0$.

Since

$$F^{-1}(1-i^*)=-F^{-1}(i^*) \text{ and } h(1-i^*)=h(i^*),\ \sum h(i^*)F^{-1}(i^*)=0.$$

L is therefore defined by

$$nL=\sum_i h(i^*)X_{(i)}=\sum_i h(i^*)\ [F^{-1}(i^*)+e_i]$$

$$=\sum_i h(i^*)e_i=\sum_i \psi'\,[F^{-1}(i^*)]\,e_i.$$

M is defined by $\sum\psi(X_{(i)}-M)=0$. We can expand each term of this sum as follows;

(1) $$\psi(X_{(i)}-M)=\psi[F^{-1}(i^*)+(e_i-M)]$$

$$=\psi[F^{-1}(i^*)]+(e_i-M)\psi'\,[F^{-1}(i^*)]+r_i.$$

Let $r=\sum r_i$. These remainder terms will be dealt with later. Summing over i in (1) we get

$$0=\sum_i\psi(X_{(i)}-M)$$

$$=\sum_i\psi[F^{-1}(i^*)+\sum_i e_i\psi'\,[F^{-1}(i^*)]-M\sum\psi'\,[F^{-1}(i^*)]+r.$$

Since $\psi(-x)=-\psi(x)$, $\sum\psi[F^{-1}(i^*)]=0$.

Since $\int h(t)\,dt=1$, $\sum\psi'\,[F^{-1}(i^*)]=\sum h(i^*)=n+O(1)$;

the excess here may be absorbed into the remainder term r.

Therefore,

$$O=O+nL-nM+r$$

or

$$M-L=\frac{r}{n}.$$

We must now examine the remainder. First suppose $\alpha_0\leqq i^*\leqq 1-\alpha_0$. If no bad point of ψ' lies between $X_{(i)}-M$ and $F^{-1}(i^*)$, then by(1),

$$|r_i|\leqq\frac{1}{2}(e_i-M)^2\ sup\ \psi'',$$

which is $O\left(\frac{1}{n}\right)$ in probability uniformly in i. The contribution of these terms to r is therefore O (1) in probability. If a bad point of ψ' lies between $X_{(i)}-M$ and $F^{-1}(i^*)$, then r_i is $O\left[\frac{1}{\sqrt{n}}\right]$ in probability uniformly in i, since ψ is continous and ψ' is bounded. Since f is bounded, the number of such i is $O(\sqrt{n})$ in probability, so their contribution to r is $O(1)$ in probablity.

Now Suppose $i^*<\alpha_0$. Since $\psi(x)$ is constant for $x<F^{-1}(\alpha_0)$, (1) implies

$$r_i=0 \quad \text{if} \quad X_{(i)}-M<F^{-1}(\alpha_0).$$

If $X_{(i)}-M\geqq F^{-1}(\alpha_0)$, we have, letting j be the smallest i such that $i^*\geqq\alpha_0$,

$$\begin{aligned}O&\leqq X_{(i)}-M-F^{-1}(\alpha_0)\leqq X_{(j)}-M-F^{-1}(\alpha_0)\\&=F^{-1}(j^*)-F^{-1}(\alpha_0)+e_j-M,\end{aligned}$$

which is $O\left(\frac{1}{\sqrt{n}}\right)$ in probability. Thus the intrusion of $X_{(i)}-M$ into the non-constant part of the domain of ψ is $O\left(\frac{1}{\sqrt{n}}\right)$ in probability uniformly in i. Hence, by (1), r_i is $O\left(\frac{1}{\sqrt{n}}\right)$ in probability uniformly in i. Since the number of such i is $O(\sqrt{n})$ in probability, their contribution to r is $O(1)$ in probability. A similar argment holds for $i^*>1-\alpha_0$. Therefore, r is bounded in probability, and the theorem follows. (Q.E.D.)

REFERENCES

[1] Andrews, Bickel, Hampel, Huber, Rogers, Tukey(1972). Robust estimates of location. Princeton University Press.

[2] Bickel, Peter J. (1965). On some robust estimates of location. Ann. Math. Stat., 36, pp. 847-858

[3] Box, G.E.P. (1953). Non-normality and tests on variances. Biometrika, 40, pp. 318-335

[4] Box, G.E.P. and Anderson, S.L. (1955). Permutation theory in the derivation of robust. criteria and the study of departures from assumption. J.Roy. Statist. Soc. Sert B 7. pp. 1-34

[5] Chernoff, Herman; Gastwirth, Joseph L. and Jones, M.V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. Ann. Math. Stat., 38. pp. 52-72.

[6] Gastwirth, Joseph L. (1966). On robust procedures. Journal of the Amer. Statisti Assoc., 61, pp. 929-948

[7] Hampel, Frank R. (1971). A general qualitative definition of robustness. Ann. Math. Stat. 42, pp. 1887-1896

[8] Hodges, J.L. and Lehmann, E.L. (1963). Estimates of Location based on rank tests. Ann. Math. Stat. 34, pp. 598-611

[9] Huber, Peter J. (1964). Robust estimation of a location parameter. Ann. Math. Stat., 35, pp. 73-101

[10] Huber, Peter J. (1972). Robust statistics: A review. Ann. Math. Stat., 43, pp. 1041-1067

[11] Jaeckel, Louis A. (1971). Robust estimates of location; Symmetry and asymmetric contamination. Ann. Math. Stat., 42, pp. 1020-1034

[12] Lindgren, B.W. (1962), Statistical Theory. The Macmillan Co. New York.

[13] Tukey, J.W. (1962). The future of data analysis. Ann. Math. Stat., 33, pp. 1-67