Ann. Inst. Statist. Math.
34 (1982), Part A, 225-239

SOME ASYMPTOTIC DISTRIBUTIONS IN THE LOCATION-SCALE MODEL

LOUIS-PAUL RIVEST

(Received Jan. 16, 1980; revised Oct. 27, 1981)

Summary

Scale and location estimators defined by the equation

$$\sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n)/\hat{V}_n]=0$$

are introduced. Their asymptotic distribution is derived. If the underlying distribution is known, a large number of estimators is shown to be efficient. Step versions of these estimators are also studied. Hampel's (1974, *J. Amer. Statist. Ass.*, **69**, 383–393) concept of influence curve is used. All the asymptotic results presented in this paper are derived from a general theorem of Rivest (1979, *Tech. Rep.*, Univ. of Toronto).

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be a random sample from a distribution $F(x)$, let $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ be the corresponding ordered sample.

With the modern emphasis on robustness (see Huber [8]), two classes of estimators of the location parameter have been widely investigated: The M-estimator $\hat{T}_n$ defined as a solution of

$$\sum_{i=1}^{n} \psi[(X_i-\theta)/\hat{V}_n^*]=0$$

where $\hat{V}_n^*$ is a scale estimator.

The L-estimator $\hat{T}_n$ defined as

$$\hat{T}_n=n^{-1}\sum_{i=1}^{n} J[i/(n+1)]X_{(i)}$$

where J satisfies $\int_0^1 J(t)dt=1$.

Key words: M-estimator, L-estimator, influence curve, Robust estimation, step estimator.

In M-estimation an observation is weighted according to its magnitude while in L-estimation it is weighted according to its rank in the sample. In Section 2 the asymptotic behavior of L-M-estimators which weight an observation according to both its magnitude and its rank is investigated. The findings are compared with known results about L-estimators (Stigler [12]) and M-estimators (Huber [6], [7]).

The third section is devoted to the study of step estimators. If the estimating equation is of the type

$$l(\theta, \hat{V}_n^*)=0$$

where $\hat{V}_n^*$ is a scale parameter, a one step estimator is defined as

$$\hat{T}_n^{(1)}=\hat{T}_n^*-l(\hat{T}_n^*, \hat{V}_n^*)/l_x(\hat{T}_n^*, \hat{V}_n^*)$$

where l_x is the partial derivative of $l(x, y)$ with respect to x, $\hat{T}_n^*$ and $\hat{V}_n^*$ are a location and a scale estimator given a priori. In Section 3, the asymptotic distribution of L–M step estimators is derived under minimal regularity conditions.

For the estimators defined in Section 2 and their step versions studied in Section 3 it is shown that

$$\left[\hat{\theta}_n-\theta-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\theta, X_i)\right] \text{ is } o_p(n^{-1/2})$$

where $\mathrm{IC}\,(\mu, x)$ is Hampel [5] influence curve.

NOTATION. The superscript "$*$" will denote estimators given a priori, independently of the estimation procedure under consideration.

2. Asymptotic behavior of L-M-estimators

As mentioned in the introduction, the L- and the M-estimators can be subsumed in the following class.

DEFINITION (L-M-estimators). Let $J(t)$ be a weight function defined in $[0, 1]$ and $\psi(x)$ be a function defined in R then the L-M-estimator of location $\hat{T}_n$, is defined as a solution of:

$$\sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-\theta)/\hat{V}_n^*]=0 \tag{2.1}$$

while the L-M-estimator of scale, $\hat{V}_n$, is defined as a solution of

$$\sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n^*)/\theta]=0\,.$$

If $J(t)=1$ the L-M-estimator reduces to M-estimators while if $\psi(x)=x$,

$$\hat{T}_n = \sum_{i=1}^{n} J[i/(n+1)]X_{(i)} \Big/ \sum_{i=1}^{n} J[i/(n+1)]$$

which is equivalent to the L-estimator of location and if $\psi(x)=|x|^\alpha - 1$,

$$\hat{V}_n = \left[\sum_{i=1}^{n} J[i/(n+1)] |X_{(i)} - \hat{T}_n^*|^\alpha \Big/ \sum_{i=1}^{n} J[i/(n+1)] \right]^{1/\alpha}$$

which is equivalent to the L-estimator of scale defined by Bickel and Lehmann [2].

The asymptotic results of this section will be derived from the following theorem:

THEOREM 1. *Let $J(t)$ be a bounded variation function defined in $[0, 1]$ and $\psi(x)$ be a function defined in R which can be written as*

$$\sum_{j=1}^{n_0} b_i \psi_i(x)$$

where $b_i \in R$, $i=1, 2, \cdots, n_0$ and $\{\psi_i\}_{i=1}^{n_0}$ is a sequence of increasing functions. Let $\hat{T}_n^*$ and $\hat{V}_n^*$ be consistent estimators of μ and γ then under the assumptions

A1) *$J(t)$ and $\psi[F^{-1}(t)]$ are not discontinuous together, ψ is continuous at $F^{-1}(t)$ for almost all t.*

And either

A2) i) *$(\hat{T}_n^* - \mu)$ and $(\hat{V}_n^* - \gamma)$ are $o_p(1)$*

ii) *There exists $\delta \in (0, 1/2)$ such that $J(t)=0$, $t \notin (\delta, 1-\delta)$ or there exists $B>0$ such that $|\psi(x)|<B$, $x \in R$.*

Or

A3) i) *$(\hat{T}_n^* - \mu)$ and $(\hat{V}_n^* - \gamma)$ are $O_p(n^{-1/2})$*

ii) *$\lambda(x, y)$ and $\lambda_H(x, y)$ are continuously diffentiable in a neighborhood of (μ, γ) where*

$$\lambda(x, y) = \int_0^1 J(t)\psi[(F^{-1}(t)-x)/y]dt$$

$$\lambda_H(x, y) = \mathrm{E}[\psi_H[(X-x)/y)]]$$

and

$$\psi_H(x) = \int_0^x J[F(y)]d\psi(y) - \mathrm{E}\left[\int_0^{(X-\mu)/\gamma} J[F(y)]d\psi(y)\right]$$

iii) *There exist $\eta>0$, M_0 in N such that $|J(t)-J(s)|<M_0|t-s|$ for both s and t in $[0, \eta]$ or in $[1-\eta, 1]$.*

There exist M_1, M_2 in N such that F is absolutely continuous in $\{x \in R: |x|>M_1\}$ and $f(x)$, the density of F, satisfies $f(x)$ and $|xf(x)|<M_2$ for $|x|>M_1$

iv) $\mathrm{E}[\psi_H^2[(X-x)/y]]$ *is finite in a neighborhood of* (μ, γ).
Then the following is true:

$$n^{-1}\sum_{i=1}^{n}\{J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*] \\ -\lambda(\hat{T}_n^*, \hat{V}_n^*)-\psi_H[(X_{(i)}-\mu)/\gamma]\} \textit{ is } o_p(n^{-1/2}) .$$

The proof of this result is technical. A sketch of the proof is contained in the appendix while a formal proof is derived in Rivest [11].

Remarks. 1) If $\psi(x)=x$, $\lambda(x, y)=\left[\int_0^1 J(t)(F^{-1}(t)-x)dt\right]\Big/y$ and if $\hat{T}_n^*$ is the L-estimator corresponding to $J(t)$, Theorem 1 implies that (taking $\hat{V}_n^*=\gamma=1$):

$$n^{1/2}\left\{\hat{T}_n^*-\mu-n^{-1}\sum_{i=1}^{n}\left[\int_0^{X_i-\mu} J[F(y)]dy-\mathrm{E}\left[\int_0^{X_i-\mu} J[F(y)]dy\right]\right]\right\} \text{ is } o_p(1) .$$

This result has been proved by Stigler [12]. It implies the asymptotic normality of L-estimators of location.

2) If $J(t)=1$ and if $\hat{T}_n^*$ is a consistent root of (2.1), Theorem 1, under assumptions A1) and A2) implies that

$$n^{1/2}\left\{\lambda(\hat{T}_n^*, \hat{V}_n^*)-n^{-1}\sum_{i=1}^{n}\psi[(X_i-\mu)/\gamma]\right\} \text{ is } o_p(1) .$$

This is a special case of a theorem of Huber [7] used to establish the asymptotic normality of maximum likelihood estimators under nonstandard conditions.

3) Define $\nu(F)=\int_0^1 J(t)\psi[(F^{-1}(t)-\mu(F))/\gamma(F)]dt$ where μ and γ are the functionals corresponding to $\hat{T}_n^*$ and $\hat{V}_n^*$. After some algebra the influence curve (Hampel [5]) of ν, $\mathrm{IC}(\nu, x)$, is shown to be equal to:

$$\psi_H[(x-\mu)/\gamma]+\mathrm{IC}(\mu, x)\lambda_x(\mu, \gamma)+\mathrm{IC}(\gamma, x)\lambda_y(\mu, \gamma)$$

where λ_x and λ_y denote the partial derivatives of λ with respect to x and y respectively and $\mathrm{IC}(\mu, x)$, $\mathrm{IC}(\gamma, x)$ are the influence curves of μ and γ respectively.

Now assuming $\left[\hat{T}_n^*-\mu-n^{-1}\sum_{i=1}^{n}\mathrm{IC}(\mu, X_i)\right]$ and $\left[\hat{V}_n^*-\gamma-n^{-1}\sum_{i=1}^{n}\mathrm{IC}(\gamma, X_i)\right]$ are $o_p(n^{-1/2})$, $(\hat{T}_n^*-\mu)$ and $(\hat{V}_n^*-\gamma)$ are $O_p(n^{-1/2})$, therefore

$$\lambda(\mu, \gamma)-\lambda(\hat{T}_n^*, \hat{V}_n^*)+(\hat{T}_n^*-\mu)\lambda_x(\mu, \gamma)+(\hat{V}_n^*-\gamma)\lambda_y(\mu, \gamma) \text{ is } o_p(n^{-1/2})$$

since λ is differentiable at (μ, γ). With the influence curve the conclusion of Theorem 1 can be reformulated as

$$\left[n^{-1}\sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]-\nu(F)-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\nu, X_i)\right] \text{ is } o_p(n^{-1/2})\,.$$

Filippova [4] has established this type of result for several statistics.

4) The assumption ψ can be written as a weighted sum of increasing functions is not too restrictive. It is easily shown (see Rivest [11]) that any function with a finite number of minima and maxima can be decomposed in such a way. All the functions ψ used in robust estimation (see Andrews et al. [1]) are of that type.

THEOREM 2 (*Asymptotic normality of L-M-estimators of location*). *Under the assumptions*

i) *ψ is increasing and J is positive,*

ii) *$\lambda_x(\mu, \gamma)\in(-\infty, 0)$ where μ is defined as the solution of $\lambda(x, \gamma)=0$,*

iii) *$\hat{V}_n^*$, the scale estimator, satisfies:*

$$n^{1/2}\left[\hat{V}_n^*-\gamma-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\gamma, X_i)\right] \text{ is } o_p(1)\,,$$

iv) A1) *and* A3) *of Theorem* 1,

the L-M-estimator $\hat{T}_n$ based on J and ψ satisfies

$$n^{1/2}\left[\hat{T}_n-\mu-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\mu, X_i)\right] \text{ is } o_p(1)$$

where

$$\mathrm{IC}\,(\mu, x)=-\{\psi_H[(x-\mu)/\gamma]+\lambda_y(\mu, \gamma)\,\mathrm{IC}\,(\gamma, x)\}/\lambda_x(\mu, \gamma)$$

is Hampel's influence curve for μ.

The theorem is also true under assumptions A1) and A2) of Theorem 1 provided J is 0 near 0 and 1 or ψ is bounded.

Note that this result implies that $n^{1/2}(\hat{T}_n-\mu)$ is asymptotically $N[0, \mathrm{E}\,[\mathrm{IC}^2\,(\mu, X)]]$.

PROOF. For any $g\in R$,

$$\mathrm{P}\,[n^{1/2}\lambda(\hat{T}_n, \gamma)<g]=\mathrm{P}\,(\hat{T}_n>k_n)$$

where k_n is defined by $n^{1/2}\lambda(k_n, \gamma)=g$. Since λ is differentiable at (μ, γ), $n^{1/2}(k_n-\mu)$ is $O(1)$. As in Huber [6], $\mathrm{P}\,(\hat{T}_n>k_n)$ and

$$\mathrm{P}\left[n^{-1/2}\sum_{i=1}^{n}\{J[i/(n+1)]\psi[(X_{(i)}-k_n)/\hat{V}_n^*]-\lambda(k_n, \gamma)\}\geqq -g\right]$$

reach the same limit as $n\to\infty$. Applying Theorem 1 under the assumptions A1) and A3)

$$n^{-1/2}\sum_{i=1}^{n}\{J[i/(n+1)]\psi[(X_{(i)}-k_n)/\hat{V}_n^*]-\lambda(k_n,\hat{V}_n^*)-\psi_H[(X_i-\mu)/\gamma]\}\text{ is }o_p(1)\,.$$

Therefore

$$\lim_n \mathrm{P}\,[n^{1/2}\lambda(\hat{T}_n,\gamma)<g]$$
$$=\lim_n \mathrm{P}\left[n^{-1/2}\sum_{i=1}^{n}\{\psi_H[(X_i-\mu)/\gamma]+\lambda(k_n,\hat{V}_n^*)-\lambda(k_n,\gamma)\}\geqq -g\right].$$

This shows that $n^{1/2}\lambda(\hat{T}_n,\gamma)$ is asymptotically normal. Since $\lambda_x(\mu,\gamma)$ is nonzero $n^{1/2}(\hat{T}_n-\mu)$ is asymptotically normal by Slutsky's Theorem. Applying Theorem 1 with $\hat{T}_n$ and $\hat{V}_n^*$ yields

$$n^{1/2}\left[\lambda(\hat{T}_n,\hat{V}_n^*)+n^{-1}\sum_{i=1}^{n}\psi_H[(X_i-\mu)/\gamma]\right]\text{ is }o_p(1)$$

which is equivalent to:

$$n^{1/2}\left\{\hat{T}_n-\mu+n^{-1}\sum_{i=1}^{n}[\psi_H[(X_i-\mu)/\gamma]+(\hat{V}_n^*-\gamma)\lambda_y(\mu,\gamma)]/\lambda_x(\mu,\gamma)\right\}\text{ is }o_p(1)$$

since λ is differentiable at (μ,γ). Replacing $n^{1/2}(\hat{V}_n^*-\gamma)$ by $n^{-1/2}\cdot\sum_{i=1}^{n}\mathrm{IC}\,(\gamma,X_i)$ concludes the proof. Q.E.D.

Remarks. 5) The assumption ψ is increasing and J is positive implies that the L-M-estimator is uniquely defined. If this assumption is not met, one has to use the method of Huber [7] to prove the asymptotic normality: first find a consistent solution to (2.1) then Theorem 1 under the assumptions A1) and A2) yields the asymptotic normality of this solution.

6) If F, J and ψ are symmetric, $\lambda_y(\mu,\gamma)=0$ and the influence curve of $\hat{T}_n$ is an odd function. If the influence curve of $\hat{V}_n^*$ is even, as is usually the case, $\hat{T}_n$ is asymptotically independent of $\hat{V}_n^*$.

7) For M-estimator, Carroll [3] has shown that $n(\log n)^{-1}\left[\hat{T}_n-\mu-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\mu,X_i)\right]$ is $O(1)$ almost surely provided ψ is a smooth function.

8) If two L-M-estimators are estimating the same parameter and have the same influence curve, their difference is $o_p(n^{-1/2})$ as conjectured by Hampel [5], see also Jaeckel [9].

Along the lines of Huber [6], one proves:

COROLLARY 1 (*Efficient estimation*). *Assuming that F is a symmetric distribution and that $\hat{V}_n^*$ is a consistent estimator of γ, for any* *strictly positive function* $J(t)$, *symmetric about* $1/2$ *with bounded variation, there exists a function* ψ,

$$\psi(y)=\int_0^y [J(F(x))]^{-1}d\left(-\frac{f'(x)}{f(x)}\right)$$

such that the L-M-estimator $\hat{T}_n$ *based on* J *and* ψ *is efficient for* μ.

Example 1. Let F be logistic, i.e. $F(x)=(1+e^{-x})^{-1}$, then

$$-\frac{f'}{f}(x)=(e^x-1)/(e^x+1)\,.$$

If $J(t)=1$ and $\psi(x)=(e^x-1)/(e^x+1)$, the efficient M-estimator is obtained. If $J(t)=t(1-t)$ and $\psi(x)=x$, this is the efficient L-estimator. If

$$J(t)=\begin{cases} t^2 & t<1/2 \\ (1-t)^2 & t\geqq 1/2 \end{cases}$$

and

$$\psi(x)=\begin{cases} 1-e^{-x} & x\geqq 0 \\ e^x-1 & x<0 \end{cases}$$

the L-M-estimator based on J and ψ is efficient.

For scale estimators, the same reasoning yields:

THEOREM 3 (*Asymptotic normality of L-M-estimators of scale*). *If the L-M-estimator* $\hat{V}_n$ *based on* ψ *and* J *is uniquely defined, under assumptions similar to the ones of Theorem* 2,

$$n^{1/2}\left[\hat{V}_n-\gamma-n^{-1}\sum_{i=1}^n \mathrm{IC}(\gamma, X_i)\right] \textit{ is } o_p(1)$$

where

$$\mathrm{IC}(\gamma, x)=\{-\psi_H[(x-\mu)/\gamma]-\lambda_x(\mu, \gamma)\,\mathrm{IC}(\mu, x)\}/\lambda_y(\mu, \gamma)$$

is the influence curve of γ.

Remark. 9) As for location estimators one can find an infinity of efficient L-M-estimator of scale. Under symmetry it is easily shown that $\hat{V}_n$ is asymptotically independent of $\hat{T}_n^*$ (compare with Bickel and Lehmann [2]).

Example 2 (The median deviation). If

$$\psi(x)=\begin{cases} -1 & |x|<1 \\ 1 & |x|>1 \end{cases}$$

and if $\hat{T}_n^*$ is the median the M-estimator of scale $\hat{V}_n$ is the median deviation. Here

$$\lambda(x, y)=\mathrm{P}\,(|X-x|/y>1)-\mathrm{P}\,(|X-x|/y<1)\,.$$

Assuming that $F(x)$ is symmetric with respect to μ, $\gamma(F)=F^{-1}(3/4)-\mu$,

$$\lambda_y(\mu, \gamma)=-4f[F^{-1}(3/4)]$$

and

$$\mathrm{IC}\,(\gamma, x)=\begin{cases} -1/4f[F^{-1}(3/4)] & |x|<1 \\ 1/4f[F^{-1}(3/4)] & |x|>1\,. \end{cases}$$

According to Theorem 1, under assumptions A1) and A2), $\hat{V}_n$ is asymptotically normal provided $f(\mu)$ and $f[F^{-1}(3/4)]$ exist and are nonzero.

3. Step estimators

Consider now a one step L-M-estimator of location

$$\hat{T}_n^{(1)}=\hat{T}_n^*-l(\hat{T}_n^*, \hat{V}_n^*)/l_x(\hat{T}_n^*, \hat{V}_n^*)$$

where

$$l(x, y)=n^{-1}\sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-x)/y]$$

and l_x is the partial derivative of l with respect to x. If $\psi(x)=x$, note that $\hat{T}_n^{(1)}=\hat{T}_n$ the L-estimator corresponding to J. The asymptotic distribution of $\hat{T}_n^{(1)}$ is now derived.

THEOREM 4. *Under the assumptions*

i) $n^{1/2}\left[\hat{T}_n^*-\mu-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\mu, X_i)\right]$ *and* $n^{-1/2}\left[\hat{V}_n^*-\gamma-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\gamma, X_i)\right]$ *are* $o_p(1)$.

ii) *The pairs* (J, ψ) *and* (J, ψ') *satisfy* A1) *and* A3) (*or* A2) *if* J *is* 0 *near* 0 *and* 1 *or* $\psi(x)$ *and* $\psi'(x)$ *are bounded*) *of Theorem* 1.

The one step L-M-estimator $\hat{T}_n^{(1)}$ *satisfies*

$$n^{1/2}\left[\hat{T}_n^{(1)}-\mu^{(1)}-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\mu^{(1)}, X_i)\right] \text{ is } o_p(1)$$

where

$$\mu^{(1)}(F)=\mu(F)-\lambda(\mu, \gamma)/\lambda_x(\mu, \gamma)$$

$$\mathrm{IC}\,(\mu^{(1)}, x)=\{-\psi_H[(x-\mu)/\gamma]-\lambda_y(\mu, \gamma)\,\mathrm{IC}\,(\gamma, x)\}/\lambda_x(\mu, \gamma)$$
$$+\lambda(\mu, \gamma)R^{(1)}(\mu, \gamma)$$

is Hampel's influence curve for $\mu^{(1)}$ *and*

$$R^{(1)}(\mu, \gamma)=\mathrm{IC}\,[\lambda_x(\mu, \gamma), x]/\lambda_x^2(\mu, \gamma)\,.$$

$\Big(\lambda_x(\mu, \gamma)$ *is considered as the functional* $-\int_0^1 J(t)\psi'[[F^{-1}(t)-\mu(F)]/\gamma(F)]dt/\gamma(F).\Big)$

PROOF. If ν_1 and ν_2 are two functionals, it is easily shown that $\mathrm{IC}\,(\nu_1+\nu_2, x)=\mathrm{IC}\,(\nu_1, x)+\mathrm{IC}\,(\nu_2, x)$ and $\mathrm{IC}\,(\nu_1/\nu_2, x)=[\mathrm{IC}\,(\nu_1, x)\nu_2-\mathrm{IC}\,(\nu_2, x)\cdot\nu_1]/\nu_2^2$ if $\nu_2\neq 0$. Therefore

$$\mathrm{IC}\,(\mu^{(1)}, x)=\mathrm{IC}\,(\mu, x)-\mathrm{IC}\,[\lambda(\mu, \gamma), x]/\lambda_x(\mu, \gamma) -\lambda(\mu, \gamma)\,\mathrm{IC}\,[\lambda_x(\mu, \gamma), x]/\lambda_x^2(\mu, \gamma)\,. \tag{3.1}$$

(Here, $\lambda(\mu, \gamma)$ is considered as a functional.) By Remark 3)

$$\mathrm{IC}\,[\lambda(\mu, \gamma), x]/\lambda_x(\mu, \gamma) =\{\psi_H[(x-\mu)/\gamma]+\lambda_y(\mu, \gamma)\,\mathrm{IC}\,(\gamma, x)\}/\lambda_x(\mu, \gamma)+\mathrm{IC}\,(\mu, x)\,.$$

Replacing $\mathrm{IC}\,[\lambda(\mu, \gamma), x]/\lambda_x(\mu, \gamma)$ by this quantity in (3.1) yields the desired expression for $\mathrm{IC}\,(\mu^{(1)}, x)$. According to Theorem 1,

$$n^{1/2}\Big[l(\hat{T}_n^*, \hat{V}_n^*)-\lambda(\mu, \gamma)-n^{-1}\sum_{i=1}^n \mathrm{IC}\,[\lambda(\mu, \gamma), X_i]\Big] \text{ is } o_p(1)\,. \tag{3.2}$$

Consider

$$l_x(\hat{T}_n^*, \hat{V}_n^*)=-n^{-1}\sum_{i=1}^n J[i/(n+1)]\psi'[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]/\hat{V}_n^*$$

since the pair (J, ψ') satisfies the assumptions of Theorem 1 and since $n^{1/2}\Big[\hat{V}_n^*-\gamma-n^{-1}\sum_{i=1}^n \mathrm{IC}\,(\gamma, X_i)\Big]$ is $o_p(1)$

$$n^{1/2}\Big[l_x(\hat{T}_n^*, \hat{V}_n^*)-\lambda_x(\mu, \gamma)-n^{-1}\sum_{i=1}^n \mathrm{IC}\,[\lambda_x(\mu, \gamma), X_i]\Big] \text{ is } o_p(1)\,. \tag{3.3}$$

Combining (3.2) and (3.3) proves the result. Q.E.D.

Remarks. 10) If μ is a solution of $\lambda(\theta, \gamma)=0$, i.e. if $\hat{T}_n^*$ and $\hat{T}_n^{(1)}$ are estimating the same parameter, $\hat{T}_n^{(1)}$ has the same asymptotic behavior as the corresponding L-M-estimator. For maximum likelihood estimators a similar conclusion has been reached by LeCam [10].

11) If μ is not a solution of $\lambda(\theta, \gamma)=0$, $\mu^{(1)}$ is the solution of $\lambda(\theta, \gamma)=0$ obtained after one iteration of the Newton Raphson procedure starting at μ. Note that $\hat{T}_n^{(1)}$ and $\hat{V}_n^*$ satisfy the assumptions of Theorem 4, therefore $\hat{T}_n^{(2)}$ the two step estimator satisfies:

$$n^{1/2}\left[\hat{T}_n^{(2)}-\mu^{(2)}-n^{-1}\sum_{i=1}^{n}\mathrm{IC}\,(\mu^{(2)}, X_i)\right] \text{ is } o_p(1)\,.$$

If the iteration procedure converges, $\mu^{(2)}$ is closer to a solution of $\lambda(\theta, \gamma)=0$ than $\mu^{(1)}$ and its influence curve is also closer to the influence curve of the corresponding L-M-estimator. Iterating this result $\hat{T}_n^{(k)}$ the k step estimator should be closer to the corresponding L-M-estimator than $\hat{T}^{(l)}$ for $l<k$.

Now the effect of a lack of robustness of $\hat{T}_n^*$ and $\hat{V}_n^*$ on $\hat{T}_n$ and $\hat{T}_n^{(1)}$ is investigated.

For instance suppose that F is t with 3 degrees of freedom, the location is to be estimated with some robust M-estimator, the scale is unknown. An a priori scale estimator, $\hat{V}_n^*$ has to be used. If $\hat{V}_n^*$ is the standard deviation then $\hat{V}_n^*$ is a consistent estimator of the population standard deviation γ. It is easily seen that $\hat{V}_n^*$ belong to the domain of attraction of a stable law with parameter 3/2. Therefore the *rate of convergence* of $\hat{V}_n^*$, $\alpha(\hat{V}_n^*)=\{\sup\beta:\ n^{1-1/\beta}(\hat{V}_n^*-\gamma)$ is $O_p(1)\}$ is 3/2. Will the slow convergence of $\hat{V}_n^*$ affect the convergence of $\hat{T}_n$? The next theorem answers this question.

So far we have assumed $\alpha(\hat{T}_n^*)=\alpha(\hat{V}_n^*)=2$, now this assumption is weakened to $\alpha(\hat{V}_n^*)$ and $\alpha(\hat{T}_n^*)\in(1, 2)$.

THEOREM 5. *Assuming*

i) A1) *and* A2) *of Theorem* 1 *hold.*

ii) $\lambda(x, y)$ *is continuously differentiable near* (μ, γ) *and* $\lambda_x(\mu, \gamma)<0$.

Then if

1) F *is symmetric with respect to* μ *and* J *and* ψ *are symmetric,* $\alpha(\hat{T}_n)=2$.

2) $\lambda_y(\mu, \gamma)\neq 0$, $\alpha(\hat{T}_n)=\alpha(\hat{V}_n^*)$.

PROOF. Assume without loss of generality $\mu=0$ and $\gamma=1$. Applying Theorem 1

$$\left[\lambda(\hat{T}_n, \hat{V}_n^*)+n^{-1}\sum_{i=1}^{n}\psi_H(X_i)\right] \text{ is } O_p(n^{-1/2})\,.$$

Applying the mean value theorem:

$$\lim_n\,[\lambda(\hat{T}_n, \hat{V}_n^*)-\lambda(0, \hat{V}_n^*)]/\hat{T}_n\lambda_x(0, 1)=1$$

in probability. If 1) holds $\lambda(0, \hat{V}_n^*)=0$ and $n^{1/2}\hat{T}_n$ has the same asymptotic distribution as $n^{-1/2}\sum_{i=1}^{n}\psi_H(X_i)$, i.e. $\alpha(\hat{T}_n)=2$. If 2) holds for any $\beta \leqq 2$,

$$\lim_n \mathrm{P}\,(n^{1-1/\beta}\hat{T}_n>g)$$
$$=\lim_n \mathrm{P}\left[n^{1-1/\beta}\Big(\lambda(0,\,\hat{V}_n^*)+n^{-1}\sum_{i=1}^{n}\psi_H(X_i)\Big)>\lambda_x(0,\,1)g\right]$$

and $\alpha(\hat{T}_n)=\alpha(\hat{V}_n^*)$. Q.E.D.

For one step estimators,

THEOREM 6. *Assuming that*

i) (J,ψ) *and* (J,ψ') *satisfy* A1) *and* A2) *of Theorem* 1.

ii) $\lambda(x,y)$ *has continuous third partial derivatives near* (μ,γ) *and* $\lambda_x(\mu,\gamma)<0$.

If

a) F *is symmetric with respect to* μ, ψ *and* J *are symmetric;*

b) $\lambda_{xy}(\mu,\gamma)$ *and* $\lambda_{(3x)}(\mu,\gamma)$ *are nonzero* $\left(\lambda_{(3x)}=\dfrac{\partial^3}{\partial x^3}\lambda(x,y)\right)$,

$$\alpha(\hat{T}_n^{(1)})=\min\{\alpha^3(\hat{T}_n^*),\,\alpha(\hat{T}_n^*)\alpha(\hat{V}_n^*),\,2\}\,.$$

If

c) $\lambda(\mu,\gamma)$, $\lambda_{xy}(\mu,\gamma)$, $\lambda_{(2x)}(\mu,\gamma)$ *are nonzero,*

$$\alpha(\hat{T}_n^{(1)})=\min\,[\alpha(\hat{T}_n^*),\,\alpha(\hat{V}_n^*)]\,.$$

PROOF. Assume without loss of generality $\mu=0$ and $\gamma=1$. As in Theorem 4, $l(\hat{T}_n^*,\hat{V}_n^*)/l_x(\hat{T}_n^*,\hat{V}_n^*)-\lambda(0,1)/\lambda_x(0,1)$ minus

$$\left[\lambda(\hat{T}_n^*,\,\hat{V}_n^*)+n^{-1}\sum_{i=1}^{n}\psi_H(X_i)-\lambda(0,\,1)\right]\Big/\lambda_x(0,\,1)$$
$$-\lambda(0,\,1)\left[\hat{V}_n^*\lambda_x(\hat{T}_n^*,\,\hat{V}_n^*)-\lambda_x(0,\,1)-n^{-1}\sum_{i=1}^{n}\psi_H^{(1)}(X_i)\right]\Big/\lambda_x^2(0,\,1)$$
$$+\lambda(0,\,1)(\hat{V}_n^*-1)/\lambda_x(0,\,1)$$

is $o_p(n^{-1/2})$ where $\psi_H^{(1)}(x)$ is the ψ_H function corresponding to J and ψ'. If a) and b) hold, $\alpha(\hat{T}_n^{(1)})$ equals

$$\alpha[\lambda_x(0,\,1)\hat{T}_n^*-\lambda(\hat{T}_n^*,\,\hat{V}_n^*)] \tag{3.4}$$

note that $\lambda(x,1)$ is odd, hence $\lambda_{(2x)}(x,1)$ is also odd, i.e. $\lambda_{(2x)}(0,1)=0$. Now using a Taylor series expansion and the fact that $\lambda(0,\hat{V}_n^*)=0$, (3.4) is equal to

$$\alpha\{\hat{T}_n^*[\lambda_x(0,\,1)-\lambda_x(0,\,\hat{V}_n^*)]-(\hat{T}_n^*)^3\lambda_{(3x)}(0,\,1)\}\,.$$

This proves the first part. If c) holds,

$$\alpha(\hat{T}_n^{(1)})=\alpha[\hat{V}_n^*\lambda_x(\hat{T}_n^*,\,\hat{V}_n^*)-\lambda_x(0,\,1)]$$
$$=\min\,[\alpha(\hat{V}_n^*),\,\alpha(\hat{T}_n^*)]\,.$$

Q.E.D.

Remark. 12) If F is symmetric, note that $\alpha(\hat{T}_n^{(1)}) \geqq \alpha(\hat{T}_n^{(2)}) \cdots$ therefore to increase the number of iterations improves the rate of convergence of the estimator.

Appendix. Sketch of the proof of Theorem 1

Without losing generality it is assumed that $J(t)$ is positive increasing bounded, $\mu=0$ and $\gamma=1$ and $\psi(x)$ is increasing.

LEMMA 1. *Under assumptions* A2)-ii) *or* A3)-iv)

$$[\lambda(\hat{T}_n^*, \hat{V}_n^*) - \lambda_n(\hat{T}_n^*, \hat{V}_n^*)] \text{ is } o_p(n^{-1/2})$$

where $\lambda_n(x, y) = n^{-1} \sum_{i=1}^{n} J[i/(n+1)]\psi\{[F^{-1}[i/(n+1)] - x]/y\}$.

PROOF. Write $\psi=\psi_1+\psi_2$ where $\psi_1(x)=\psi(x)$ if $x \geqq 0$ and $\psi_2(x)=\psi(x)$ if $x<0$. Assume $\psi(0)=0$, i.e., ψ_1 is positive increasing. For δ large enough such that $(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*>0$ if $t>\delta$,

$$\lambda_1(\hat{T}_n^*, \hat{V}_n^*) = \sum_{i=[n\delta]+1}^{n} \int_{(i-1)/n}^{i/n} J(t)\psi_1[(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*]dt .$$

Since the product of two positive increasing functions is positive increasing,

$$\lambda_1(\hat{T}_n^*, \hat{V}_n^*) < \lambda_{n1}(\hat{T}_n^*, \hat{V}_n^*) + \int_{n-1}^{n} J(t)\psi_1[(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*]dt .$$

Under A2)-ii) or A3)-iv),

$$\lim_n n^{1/2} \int_{n-1}^{n} J(t)\psi_1[(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*]dt = 0 .$$

Bounding $\lambda_1(\hat{T}_n^*, \hat{V}_n^*)$ from below yields the result for ψ_1. To prove the result for ψ_2 it can be assumed that $J(t)$ is negative increasing, hence $J(t)\psi[(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*]$ is positive decreasing as a product of negative increasing functions. The reasoning is similar to the first part.

Q.E.D.

A) *Proof under* A1) *and* A2)

Using this result,

$$n^{-1/2} \sum_{i=1}^{n} J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*] - \lambda(\hat{T}_n^*, \hat{V}_n^*)$$

can be written as $h_n(Z^{(n)}(\cdot))$ where h_n is a random function defined by

$$h_n(x(\cdot))=n^{-1/2}\sum_{i=1}^{n} J[i/(n+1)]\{\psi[[F^{-1}[i/(n+1)+n^{-1/2}x[i/(n+1)]]-\hat{T}_n^*]/\hat{V}_n^*]$$
$$-\psi[[F^{-1}[i/(n+1)]-\hat{T}_n^*]/\hat{V}_n^*]\}$$

and $Z^{(n)}(\cdot)$ is the empirical process. Heuristically for large n, $h_n(Z^{(n)}(\cdot))$ can be written as:

$$n^{-1}\sum_{i=1}^{n} J[i/(n+1)]Z^{(n)}[i/(n+1)]\left[\frac{d}{dt}\psi[(F^{-1}(t)-\hat{T}_n^*)/\hat{V}_n^*]\right]_{t=i/(n+1)}$$

this random variable should therefore converge to $\int_0^1 J(t)Z(t)d\psi[F^{-1}(t)]$, $Z(t)$ is the Brownian Bridge.

Lemma 2 of Rivest [11] contains a rigorous proof of this statement under assumption A2) (i.e. $|\psi|$ is bounded or J is 0 near 0 and 1).

Using a similar argument it is shown that $n^{-1/2}\sum_{i=1}^{n}\psi_H(X_i)$ converges to

$$\int_0^1 Z(t)d\psi_H[F^{-1}(t)]\ .$$

Now since $d\psi_H[F^{-1}(t)]=J(t)d\psi[F^{-1}(t)]$ the two random variables under consideration converge to the same limit. This proves the theorem under A1) and A2).

B) *Proof under* A1) *and* A3)

Consider

$$n^{-1/2}\sum_{i=1}^{n}\{J[i/(n+1)]\psi[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]-\lambda(\hat{T}_n^*,\hat{V}_n^*)$$
$$-\psi_H[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]+\lambda_H(\hat{T}_n^*,\hat{V}_n^*)\}\ .$$

By Lemma 1, this random variable will reach the same limit as

$$\text{(A.1)}\qquad n^{-1/2}\sum_{i=1}^{n}\int_{[F^{-1}[i/(n+1)]-\hat{T}_n^*]/\hat{V}_n^*}^{[X_{(i)}-\hat{T}_n^*]/\hat{V}_n^*}[J[i/(n+1)]-J[F(x)]]d\psi(x)\ .$$

Using assumption A3), for any $\eta>0$, it is possible to find $\delta>0$ such that

$$n^{-1/2}\left|\sum_{i=1}^{[n\delta]}(\cdots)+\sum_{i=n-[n\delta]+1}^{n}(\cdots)\right|\ \text{is}\ O_p(\eta)\ .$$

The argument used to prove the theorem under A1) and A2) serves to prove

$$n^{-1/2}\sum_{i=[n\delta]+1}^{n-[n\delta]}(\cdots)\ \text{is}\ o_p(1)\ .$$

Therefore (A.1) is $o_p(1)$.

Write $\psi_H=\psi_{1H}+\psi_{2H}$ where $\psi_{1H}=\psi_H$ when $\psi_H>0$, 0 if not. To prove the result it suffices to show that

$$\text{(A.2)}\quad n^{-1/2}\sum_{i=1}^{n}\{\psi_{jH}[(X_i-\hat{T}_n^*)/\hat{V}_n^*]-\lambda_{jH}(\hat{T}_n^*,\hat{V}_n^*)$$
$$-\psi_{jH}(X_i)+\mathrm{E}(\psi_{jH}(X_i))\}\text{ is }o_p(1)\qquad\text{for } j=1,2\,.$$

Take $j=1$. For any $\varepsilon>0$, by the assumption on $\hat{T}_n^*$ and $\hat{V}_n^*$ it is possible to find constants C_0, C_1 such that $|\hat{T}_n^*|<C_0n^{-1/2}$ and $|\hat{V}_n^*-1|<C_0n^{-1/2}$ and $|\lambda_{jH}(\hat{T}_n^*,\hat{V}_n^*)|<C_1n^{-1/2}$ for large n except on a set of probability ε. Similarly one can find C_2 such that

$$|\lambda_{jH}(-C_0n^{-1/2},1\pm C_0n^{-1/2})|<C_2n^{-1/2}$$

for large n. Now take $\delta=\varepsilon/(C_2+C_1)$, since $\psi_{1H}(x)$ is increasing, null for small x, positive for large ones,

$$n^{-1/2}\sum_{i=n-[n\delta]+1}^{n}\psi_{1H}[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]-\lambda_{1H}(\hat{T}_n^*,\hat{V}_n^*)$$
$$\leqq\varepsilon+n^{-1/2}\sum_{i=n-[n\delta]+1}^{n}\psi_{1H}[(X_{(i)}-k_n)/s_n]-\lambda_{1H}(k_n,s_n)$$

where $k_n=-C_0n^{-1/2}$, $s_n=1-k_n$. Therefore (A.2) is less than

$$\varepsilon+n^{-1/2}\sum_{i=1}^{n}\psi_{1H}[(X_i-k_n)/s_n]-\lambda_{1H}(k_n,s_n)-\psi_{1H}(X_i)+\mathrm{E}(\psi_{1H}(X_i))$$
$$+n^{-1/2}\sum_{i=1}^{n-[n\delta]}\psi_{1H}[(X_{(i)}-\hat{T}_n^*)/\hat{V}_n^*]-\lambda_{1H}(\hat{T}_n^*,\hat{V}_n^*)$$
$$-\psi_{1H}[(X_{(i)}-k_n)/s_n]+\lambda_{1H}(k_n,s_n)\,.$$

The first summation is summing independent variables with 0 expectation. It is easily seen that its variance goes to 0. The second summation is $o_p(1)$ by an argument used previously hence (A.2) is less than ε. Similarly it can be shown that (A.2) is bigger than $-\varepsilon$ for large n, therefore (A.2) is $o_p(1)$ when $j=1$. The proof when $j=2$ is similar. Q.E.D.

Acknowledgment

This work was done while I was a graduate student. I want to thank my advisers Professors Harold Ruben and Constance van Eeden for their valuable comments and helpful suggestions.

UNIVERSITÉ LAVAL

References

[1] Andrews, D. F. et al. (1972). *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, N.J.

[2] Bickel, P. J. and Lehmann, E. L. (1976). Descriptive statistics for nonparametric models, III, *Ann. Statist.*, **4**, 1139-1158.

[3] Carroll, R. J. (1978). On almost sure expansions for M-estimates, *Ann. Statist.*, **6**, 314-318.

[4] Filippova, A. A. (1962). Mises' theorem on the asymptotic behaviour of functionals of empirical distribution functions and its statistical applications, *Theory Prob. Appl.*, **7**, 24-57.

[5] Hampel, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Ass.*, **69**, 383-393.

[6] Huber, P. J. (1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73-101.

[7] Huber, P. J. (1967). The behavior of maximum likelihood estimates under non standard conditions, *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press, 221-233.

[8] Huber, P. J. (1972). Robust statistics: A review, *Ann. Math. Statist.*, **43**, 1041-1067.

[9] Jaeckel, L. A. (1971). Robust estimates of location: Symmetry and asymmetric contamination, *Ann. Math. Statist.*, **42**, 1020-1034.

[10] LeCam, L. (1956). On the asymptotic theory of estimation and testing hypothesis, *Proc. 3rd Berkeley Symp. Math. Statist. Prob.*, Univ. of California Press, 129-156.

[11] Rivest, L. P. (1979). An asymptotic theorem in the location scale model, *Tech. Rep.*, Univ. of Toronto.

[12] Stigler, M. S. (1974). Linear functions of order statistics with smooth weight functions, *Ann. Statist.*, **4**, 676-693. (Correction, **7**, 466.)