AN ASYMPTOTIC REPRESENTATION FOR M-ESTIMATORS AND LINEAR FUNCTIONS OF ORDER STATISTICS

by

Raymond J. Carroll*

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1006

May, 1975

* This work was supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-75-2840.

AN ASYMPTOTIC REPRESENTATION FOR M-ESTIMATORS AND LINEAR FUNCTIONS OF ORDER STATISTICS

by

Raymond J. Carroll*

University of North Carolina at Chapel Hill

SUMMARY

Let $X_1, X_2, \ldots$ be independent observations from a distribution function $F(x - \theta)$. Let T_n be an M-estimator or a linear function of order statistics. Conditions are given for the existence of i.i.d. mean zero random variables $Y_1, Y_2, \ldots$ such that $T_n - \theta = n^{-1} \sum_1^n Y_i + O\left(n^{-1}(\log n)^2\right)$ a.s. as $n \to \infty$; this representation is shown to hold for Huber and Hampel M-estimators as well as the trimmed mean. It can also be extended to a scale invariant version of the M-estimators; if F is symmetric there is no change in the representation, but the effect of asymmetry is to add in an extra random term. The results are applied to show (i) an almost sure equivalence between M-estimators and linear functions of order statistics (ii) that a similar representation holds for adaptive estimators in the spirit of Jaeckel (1971).

AMS 1970 Subject Classifications: Primary 62G35; Secondary 62E20

Key Words and Phrases: M-estimators, Linear function of order statistics, Robustness, Adaptive estimators

* This work was supported in part by the Air Force Office of Scientific Research under Grant AFOSR-75-2840.

INTRODUCTION

Let $X_1, X_2, \ldots$ be a sample of i.i.d. random variables with distribution function $F(x - \theta)$. Two basic types of robust estimators T_n of the location parameter θ are the M-estimators (Huber (1964)) and linear functions of order statistics (LFO's).

It is known (Huber (1972), Filippova (1962)) that in many cases the M-estimators satisfy

$$n^{\frac{1}{2}}\{T_n - \int I(x,F)dF_n(x)\} \to 0 \quad \text{in probability,}$$

where $I(x,F)$ is the influence curve (Hampel (1972)). de Wet and Venter (1974) show that the trimmed mean can be represented as the mean of i.i.d. random variables with an error term $O(n^{-1} \log_2 n)$ almost surely. Jaeckel (1971a) has shown under some conditions that for a large class of M-estimators M_n there is an LFO L_n for which $M_n - L_n = O(n^{-1})$ in probability; use of the de Wet and Venter result shows that the M-estimators corresponding to the trimmed mean can be represented as the mean of i.i.d. random variables with error term $O(n^{-1} \log_2 n)$ *in probability*.

Based on the above facts, it is reasonable to hypothesize *almost sure* linear representations for and *almost sure* equivalence between M-estimators and LFO's. In this paper we verify these hypotheses for wide classes of these estimators; the latter equivalence follows as a corollary to the representations, as does the Law of the Iterated Logarithm for these estimators.

The results are given for both the symmetric and asymmetric cases. We also study (Section 3) the scale invariant versions of M-estimators and obtain a similar linearization. However, interesting things occur if F is asymmetric and our results show precisely what is happening. The representations, besides being of interest in themselves and establishing the above equivalence, also have other applications; for example, we use them to derive in a straight-forward fashion almost sure linear representations for adaptive estimators (Jaeckel (1971b)) and certain random means (Shorack (1974)).

In Section 2, the almost sure linearization of M-estimators is given. Not only does the approach work for the monotone ψ functions (Huber (1964)), but it provides results for non-monotone ψ functions such as the important Hampel ψ function (Huber (1972), Hampel (1974)). The approach does not depend on the symmetry of F .

In Section 3, the results are extended to scale invariant M-estimators. Andrews, et. al. (1972, page 40) indicate that in this case the influence curve can become quite complicated, unless F is symmetric. Our results show quite clearly what happens; the effect of estimating the scale parameter ξ by s is to introduce an extra term into the representation of the order

$$(s - \xi)n^{-1} \sum_{i=1}^{n} X_i \psi'(X_i) \ .$$

If F is symmetric, this term will be of sufficiently small order to drop out of the representation. Interestingly enough, if F is asymmetric, this term plays a definite role in the asymptotic distribution.

In Section 4, we extend the de Wet and Venter (1974) representation to a fairly large class of LFO's. The proof relies heavily on a result of Moore (1968).

In Section 5, three applications of the representations are given. First of all, an almost sure equivalence result between M-estimators and LFO's is given. Then the representation is applied to the problem of adaptive robust estimation. Jaeckel (1971b) considers estimating the optimal trimming proportion for the trimmed mean. We use the same idea to estimate the best value of k for the Huber estimates. Under conditions analogous to those in Jaeckel (1971b), the representation theorem yields a simple proof that these adaptive M-estimators (as well as adaptive trimmed means) also satisfy the representation. Finally, the representation is applied to the problem of random means (Shorack (1974)).

The results of this paper can be extended to the one-step versions of M-estimators (see Bickel (1975), Andrews, et. al. (1972)). This will be the subject of a later paper.

Finally, it should be remarked that Carroll (1974), in connection with sequential selection procedures, has obtained in a different manner expansions of a somewhat allied nature to those given here for a specific subclass of M-estimators and LFO's. These expansions are uniform over neighborhoods of a specific distribution function F ; however, the results are not of the generality or precision given here. Although the conditions on F in the above paper are quite weak, the class of estimators is fairly restrictive; for example, the Huber and Hampel M-estimators are not covered by the results. In addition, it is not possible to obtain from Carroll (1974) the equivalence theorem and representation for adaptive estimators shown in Section 5. The uniformity mentioned above can also be obtained from our approach.

M-ESTIMATORS

Suppose $X_1, X_2, \ldots$ are i.i.d. with a distribution function $F(x - \theta)$. The M-estimators were proposed by Huber (1964) as robust estimates of the location parameter θ (see also Huber (1972), (1964)); in Andrews et. al. (1972), Hampel (see also Hampel (1974)) proposed the so-called descending three-step estimators which were shown to have very good performance. The basic version of the M-estimators T_n (see the next section for the scale-invariant version) is defined as a solution to

$$0 = n^{-1} \sum_{i=1}^{n} \psi(X_i - T_n) \ . \tag{2.1}$$

The minimax estimator proposed by Huber (1964) has a ψ function

$$\begin{aligned} \psi_0(x) = -\psi_0(-x) &= x \qquad 0 \le x \le k \\ &= k \qquad k < x \ . \end{aligned} \tag{2.2}$$

The Hampel estimates are defined by

$$\begin{aligned} \psi_1(x) = -\psi_1(-x) &= x \qquad 0 \le x < a \\ &= a \qquad a \le x < b \\ &= \frac{c-x}{c-b}\, a \qquad b \le x < c \\ &= 0 \qquad c \le x \ . \end{aligned} \tag{2.3}$$

In this section under fairly weak conditions on F (which basically involve F being smooth at points where ψ' does not exist), conditions on ψ (satisfied by ψ_0, ψ_1 above) are given for which if $a = E_F\psi'(X)$, (2.4) $T_n - \theta = (an)^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) + O\left(n^{-1}(\log n)^2\right)$ a.s. ,

where $a_n = o(b_n)$ means $a_n/b_n \to 0$, and $a_n = O(b_n)$ means $|a_n/b_n| \le M$ as $n \to \infty$.

The result (2.4) naturally yields (if $E_F\psi(X) = 0$) asymptotic normality and a Law of the Iterated Logarithm for T_n . It also shows why the M-estimators are robust: they are basically a mean of bounded random variables weighted and truncated by the ψ function. Since the first term on the right-hand-side of (2.4) is $\int I_F(x - \theta)\, dF_n(x)$ (where $I_F(x)$ is the influence function), this verifies again the value of the influence function. This also verifies again that the M-estimator formed by ψ_0 in (2.2) acts like the Winzorized mean was designed to act. It also shows that ψ_1 given by (2.3) is actually not particularly influenced by outliers.

There are two basic steps in the proof of (2.4). First, it is shown that $n^{\frac{1}{2}}(\log n)^{-1}\ (T_n - \theta) \to 0$ (a.s.) . Then, (2.1) is *essentially* expanded in a Taylor series, although the details are somewhat messy because (2.2) and (2.3) are not differentiable. Assuming (without loss of generality) that $\int \psi(x)\, dF(x) = 0$, consider the following possible conditions:

(A1) ψ satisfies a uniform Lipschitz condition or order one and there are intervals $(a_0 = -\infty, a_1), \ldots ,(a_k, +\infty = a_{k+1})$ on which ψ has two continuous, bounded derivatives.

(A2) F satisfies a Lipschitz condition of order one in neighborhoods of $\{a_1, \ldots ,a_k\}$ and $\int \psi'(x)\, dF(x) = E_F\psi'(X) \ne 0$.

(A3) If $\lambda(\theta) = \int \psi(x - \theta)\, dF(x)$, then $\lambda(\theta)$ has a unique zero at $\theta = 0$. In addition, $|\lambda(\theta)|$ increases in some neighborhood about 0 .

(A4) There is a closed interval $[-b,b]$ such that T_n is in this interval almost surely as $n \to \infty$ if $\theta = 0$.

(B1) $\lambda(\theta)$ has only a finite number of zeros and $E_F\psi'(X) \neq 0$.

(B2) The set of points of increase of F is an interval, $\lambda(\theta)$ has only a finite number of zeros on this interval, and $E_F\psi' \neq 0$.

The conditions (B1) or (B2) will be used in place of (A3) if ψ is non-monotone. The condition (A1) is slightly stronger than that of Jaeckel (1971a) but not significantly so, and (A3) is slightly weaker. The condition (A4) must be verified in each case.

Theorem 2.1. If $X_1, X_2, \ldots$ have distribution function $F(x - \theta)$, then under (A1) - (A4),

$$T_n - \theta = O\left(n^{-\frac{1}{2}}(\log n)\right) \text{ a.s.} \tag{2.5}$$

The result (2.5) still holds if, instead of (A3), the following is true:

(A3)* $\lambda(\theta)$ is continuously differentiable, and either (B1) or (B2) hold. In addition, there exists η arbitrarily close to 0 such that $\lambda(\pm\eta) \neq 0$ and $\text{sign}\left(\lambda(\eta)\right) = -\text{sign}\left(\lambda(\eta)\right)$, and the sample median is strongly consistent (i.e., almost surely converges to θ).

The proof is delayed until the end of the section. The main result is:

Theorem 2.2. If (A1), (A2) and the conclusion to Theorem 2.1 hold, then

$$T_n - \theta = \left(E_F\psi'(x)\right)^{-1} n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) + O\left(n^{-1}(\log n)^2\right) \text{ a.s. as } n \to \infty \tag{2.6}$$

The following Corollaries investigate the M-estimators formed by (2.2) and (2.3).

Corollary 2.1. Suppose (A1) holds, that ψ is increasing (strictly so in a neighborhood of zero) and skew-symmetric $(\psi(x) = -\psi(-x))$, that F is Lipschitz in neighborhoods of $\{a_1, \ldots, a_k\}$ and strictly increasing in a neighborhood of zero. Then (2.6) holds.

Corollary 2.2. Suppose F is symmetric, unimodal and sufficiently smooth at the points $0, \pm a, \pm b, \pm c$ so that (i) the sample median is strongly consistent (ii) $\lambda(\theta)$ is continuously differentiable. Define T_n as the solution to (2.1) using ψ_1 (given by (2.3)) which is closest to the sample median. Then the representation (2.6) holds if either (B1) or (B2) hold.

Remarks: First of all, note that in Corollary 2.1 it is not assumed that F is symmetric; this will be important in the equivalence theorem of Section 5. Secondly, it seems that the conditions on F in Corollary 2.2 are much stronger than necessary.

From now on, assume without loss of generality that $X_1, X_2, \ldots$ have distribution F and $E_F\psi(X) = 0$. Let $\beta(n) = n^{-\frac{1}{2}}(\log n)$. Before proving the above results, it is useful to have the following propositions.

Proposition 2.1. Let $\varepsilon > 0$ and suppose $\gamma_n \to 0$. Then

$$-\lambda(\pm\varepsilon\gamma_n) = \pm\varepsilon\gamma_n E_F\psi'(X) + o(\gamma_n) .$$

Proof: Since $\lambda(0) = 0$,

$$\lambda\big(-\varepsilon\beta(n)\big) = \lambda\big(-\varepsilon\beta(n)\big) - \lambda(0)$$

$$= \sum_{j=1}^{k+1} \int_{a_{j-1}}^{a_j - \varepsilon\beta(n)} \{\psi\big(x + \varepsilon\beta(n)\big) - \psi(x)\}\, dF(x)$$

$$+ \sum_{j=1}^{k} \int_{a_j - \varepsilon\beta(n)}^{a_j} \{\psi\big(x + \varepsilon\beta(n)\big) - \psi(x)\}\, dF(x) .$$

Now, apply Taylor's Theorem to the integrands in the first sum and apply the Lipschitz condition on ψ to the integrands in the second sum to complete the proof.

Proposition 2.2. For any $b \geq 0$,

$$\sup_{|\theta| \leq b} n^{-1} \sum_{i=1}^{n} \{\psi(X_i - \theta) - \lambda(\theta)\} = O\big(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\big) \quad \text{a.s.}$$

Proof: We may assume $b = 1$. Because ψ satisfies a Lipschitz condition, by the Borel-Cantelli Lemma it suffices to show that there is a constant $M > 0$ such that

$$(2.8) \quad \sum_{n=1}^{\infty} \sum_{j=-n}^{n} \Pr\{(\log n)^{-\frac{1}{2}} n^{\frac{1}{2}} |n^{-1} \sum_{i=1}^{n} \psi(X_i - j/n) - \lambda(j/n)| > M\} < \infty .$$

This is accomplished by using the exponential bounds given in Loéve (1963), page 254.

Proposition 2.3. Let $S_1, S_2, \ldots$ be binomial random variables (not necessarily independent) with $E\, S_n = p_n$, where $p_n = O\big(n^{-\frac{1}{2}}(\log n)\big)$. Then, if $A_1, A_2, \ldots$ are random variables with $|A_n| \leq p_n S_n$, we have

$n^{-1}A_n = O(n^{-1}(\log n)^2)$ a.s. If $p_n = o(n^{-\frac{1}{2}+\alpha})$ for all $\alpha > 0$, then $n^{-1}A_n = o(n^{-1+\alpha})$ a.s. for all $\alpha > 0$. If $p_n = o(n^{-\frac{1}{2}+\beta})$, then $n^{-1}A_n = o(n^{-1+2\beta})$ in probability.

Proof: Simple calculations show that there are constants $a > 0$, $M > 0$ such that for n large, if $p_n = O(n^{-\frac{1}{2}}(\log n))$,

$$\Pr\{n^{-1}|A_n| > aMn^{-1}(\log n)^2\} \le \Pr\{n^{-1}|S_n - np_n| > M_n^{-1}(\log n)\} .$$

The exponential bounds and the Borel-Cantelli Lemma yield the first result. From Johnson and Kotz (1969), page 54, we know that

$$E|S_n|^k = \sum_{j=0}^{k} c_j B_j(n) p_n^j ,$$

where $B_j(n)$ is a j^{th} degree polynomial in n. Thus, by the Markov inequality for k^{th} moments, the second and third results follow.

Proposition 2.4. Suppose H is continuously differentiable and has only a finite number of zeros on $C = [-b,b]$, say $\{a_1, \ldots, a_m\}$. Let $U_{ni} = (a_i - \varepsilon\beta(n), a_i + \varepsilon\beta(n))$, $i = 1, 2, \ldots, m$. Let $|H(x)|$ attain its minimum on $K_n = C - \bigcup_{i=1}^{m} U_{ni}$ at ξ_n. Then

$$\beta(n) \le \min_{1\le i\le m} |\xi_n - a_i| = O(\beta(n)) .$$

Proof: Because H is continuous and C is compact, $\min_{1\le i\le m} |\xi_n - a_i| \to 0$. We may assume $\xi_n \to a_1$. Expanding $H(\xi_n)$ and $H(a_1 + \varepsilon\beta(n))$ about a_1 and dividing by $\beta(n)$ yields

$$|\beta(n)^{-1}(\xi_n - a_1)H'(\eta_n^*)| \leq |\epsilon H'(\eta_n)| \ ,$$

where η_n^*, η_n converge to $H'(a_1)$. This yields the result.

The proof of Theorem 2.1 given below is an adaptation of the proof for maximum likelihood estimation given in Huber (1965). Basically, it involves showing something like

$$-n^{\frac{1}{2}}(\log_2 n)^{-\frac{1}{2}}\{n^{-1} \sum_{i=1}^{n} \psi\left(X_i \pm \epsilon n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\right)\} \to \pm\epsilon E_F\psi'(X) \quad \text{a.s.}$$

If ψ were monotone (see Corollary 2.1), this would indeed yield Theorem 2.1. However, ψ is not in general monotone.

Proof of Theorem 2.1. Let $C = [-b,b]$ and $U_n = \left(-\epsilon\beta(n), \epsilon\beta(n)\right)$. Then $K_n = C - U_n$ is compact, so that $|\lambda(\theta)|$ attains a positive minimum on K_n (say at a_n). Now, $|\lambda(\theta)| \to 0$ as $\theta \to 0$, so that $a_n \to 0$. Since $|\lambda(\theta)|$ is increasing in a neighborhood of 0 , $a_n = \pm\ \epsilon\beta(n)$. By Proposition 2.1, $|\lambda\left(\pm\epsilon\beta(n)\right)| \geq c\epsilon\beta(n)$ for n sufficiently large and some c . By Proposition 2.2, for small $\eta > 0$,

$$\sup_{\theta\epsilon K_n} |n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) - \lambda(\theta)| \leq (c - \eta)\epsilon\beta(n)$$

almost surely as $n \to \infty$. Hence, almost surely as $n \to \infty$,

$$\inf_{\theta\epsilon K_n} |\beta(n)^{-1}\, n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta)| \geq \eta\epsilon > 0 \ .$$

Since $\sum_{i=1}^{n} \psi(X_i - T_n) = 0$, $T_n \notin K_n$ almost surely as $n \to \infty$, completing the first part of the proof. Suppose now that (A3)* holds in place of (A3). Suppose that the zeros of $\lambda(\theta)$ allowable under (B1) or (B2) are {

$\{a_1, \ldots, a_m\}$, and form a disjoint cover of these zeros by the intervals $U_{ni} = \big(a_i - \varepsilon\beta(n),\ a_i + \varepsilon\beta(n)\big)$, $i = 1, 2, \ldots, m$. Suppose $|\lambda(\theta)|$ attains its minimum on $K_n = C - \bigcup_{i=1}^{m} U_{n,i}$ at ξ_n . By Proposition 2.4,

$$\varepsilon\beta(n) \le \min_{1\le i\le m} |\xi_n - a_i| \le c\beta(n) .$$

Thus, $|\lambda(\xi_n)| \ge c^*\varepsilon\beta(n)$ for n large and some $c^* > 0$. This, together with Proposition 2.2, means that $T_n \notin K_n$ almost surely $n \to \infty$. Now, since

$n^{-1} \sum_{i=1}^{n} \psi\big(X_i - (\pm\eta)\big) \to \lambda(\pm\eta)$ a.s. , this means by Proposition 2.2 that eventually there is a zero to (2.1) in the interval $\big(-\beta(n), \beta(n)\big)$; since the sample median is strongly consistent, this completes the proof.

<u>Proof of Theorem 2.2</u>. Assume for convenience that ψ has at most two points of non-differentiability, say $\{\pm k\}$. Define

$$A_i = \{|X_i - T_n| \le k \text{ and } |X_i| \le k\}$$
$$B_i = \{|X_i - T_n| > k \text{ and } |X_i| > k\} .$$

Now, by expanding $\psi(X_i - T_n)$ about X_i on A_i and B_i , since ψ'' is bounded and $T_n^2 = o\big(n^{-1}(\log n)^2\big)$ a.s. ,

$$\begin{aligned}(2.9)\quad 0 &= n^{-1} \sum_{i=1}^{n} \psi(X_i - T_n) \\ &= n^{-1} \sum_{i=1}^{n} \{\psi(X_i) - T_n\psi'(X_i)\} I_{A_i \cup B_i} \\ &\quad + n^{-1} \sum_{i=1}^{n} \psi(X_i - T_n)\{1 - I_{A_i\cup B_i}\} + o\big(n^{-1}(\log n)^2\big) \text{ a.s. ,}\end{aligned}$$

where $I_{A_i \cup B_i}$ is the indicator function of the event $A_i \cup B_i$. Because of Proposition 2.3, since ψ satisfies a Lipschitz condition, the middle term of equation (2.9) (call it B^n) becomes

$$(2.10)\quad B^n = n^{-1}\sum_{i=1}^{n}\{\psi(X_i) - T_n\psi'(X_i)\}(1 - I_{A_i\cup B_i}) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

Thus,

$$(2.11)\quad T_n\{n^{-1}\sum_{i=1}^{n}\psi'(X_n)\} = n^{-1}\sum_{i=1}^{n}\psi(X_i) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.} ,$$

which completes the proof since

$$n^{-1}\sum_{i=1}^{n}\psi'(X_i) = E_F\psi'(X) + O\left(n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\right) \quad \text{a.s.}$$

$$n^{-1}\sum_{i=1}^{n}\psi(X_i) = O\left(n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\right) \quad \text{a.s.}$$

Proof of Corollary 2.1. Since ψ is increasing, $\psi' \geq 0$, so that (A2) holds. Since ψ and F are strictly increasing in a neighborhood of zero, (A3) holds. Huber (1964) has shown (A4) by proving that T_n is almost surely consistent.

Proof of Corollary 2.2. Because F is symmetric and unimodal, the conditions guarantee that (A1) and (A2) hold. For the same reason, since ψ_1 is skew-symmetric and increasing at 0 , there exists $\eta > 0$ arbitrarily close to 0 such that $\lambda(\eta) \neq 0$ and sign $(\lambda(\eta))$ = -sign $(\lambda(-\eta))$. Thus, (A3)* is verified, and it suffices by Theorem 2.1 merely to show that (A4) holds. Because

$$n^{-1}\sum_{i=1}^{n}\psi_1\left(X_i - (\pm\eta)\right) \to \lambda(\pm\eta) \lessgtr 0 \quad \text{a.s.} ,$$

there is eventually a solution to (2.1) in the interval $(-\eta,\eta)$; because the sample median is strongly consistent, this gives (A4).

SCALE INVARIANT M-ESTIMATORS

An estimator $T_n = T(X_1, \ldots ,X_n)$ is called scale invariant if $T(X_1/c, \ldots ,X_n/c) = T_n/c$ for any constant $c \neq 0$. The estimators in the previous section do not satisfy this property, and it is the purpose of this section to modify the criterion (2.1) to get scale invariant versions. Andrews, et. al. (1972, page 40) indicate that even if the ψ functions are monotone as in Corollary 2.1, the influence function is complicated unless F is symmetric. The representation given here shows clearly the behavior of the estimators in both symmetric and asymmetric cases; there is basically no change in the symmetric case but if F is not symmetric, an additional term of order $O(n^{-\frac{1}{2}})$ in probability is added.

Suppose that s_n is a scale invariant estimator which, for some constant ξ , satisfies

$$(3.1) \qquad s_n(X_1 + a, \ldots ,X_n + a) = s_n(X_1, \ldots ,X_n)$$

$$s_n = \xi + O\left(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right) \quad \text{a.s.}$$

Under the conditions given in Bahadur (1966), the interquartile range is a scale invariant estimator satisfying (3.1). In the following if the rate term in (3.1) is $o(n^{-\beta})$ for some $\beta > 0$, the order term appearing in (3.3) is

$o(n^{-2\beta})$ only. In analogy with (2.1), define T_n as the solution to

$$0 = n^{-1} \sum_{i=1}^{n} \psi\left(s_n^{-1}(X_i - t)\right) . \tag{3.2}$$

Clearly, T_n is scale invariant. The result similar to Theorem 2.2 is the following:

<u>Theorem 3.1</u>. Suppose $X_1, X_2, \ldots$ have a distribution function $F\left(\xi^{-1}(x - \theta)\right)$ and that the conditions of Theorem 2.2 hold. Then,

$$\begin{aligned}(3.3)\quad \xi^{-1}(T_n - \theta) = \{E_F\psi'(X)\}^{-1}\left[n^{-1} \sum_{i=1}^{n} \psi\left(\xi^{-1}(X_i - \theta)\right)\right] + \{E_F\psi'(X)\}^{-1} \\ (1 - s_n/\xi)\left[n^{-1} \sum_{i=1}^{n} \left(\xi^{-1}(X_i - \theta)\right)\psi'\left(\xi^{-1}(X_i - \theta)\right)\right] \\ + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}\end{aligned}$$

If, in addition, $E_F X\psi'(X) = 0$,

$$\xi^{-1}(T_n - \theta) = \{E_F\psi'(X)\}^{-1}\left[n^{-1} \sum_{i=1}^{n} \psi\left(\xi^{-1}(X_i - \theta)\right)\right] + O\left(n^{-1}(\log n)^2\right) \text{ a.s.} \tag{3.4}$$

Note that the condition $E_F X\psi'(X) = 0$ is satisfied if F is symmetric about 0 and $\psi(x) = -\psi(-x)$, the latter being true for (2.2) and (2.3). Note that if $E_F X\psi'(X) \neq 0$, the middle term of (3.3) is only $O(n^{-\frac{1}{2}})$ in probability in many cases, so that there is a significant effect on the asymptotic distribution of T_n .

<u>Proof:</u> The proof is quite similar to that of Theorems 2.1 and 2.2 so only an outline will be given. Assuming that $\theta = 0$, and $\xi = 1$, by the Lipschitz

condition on ψ

(3.5) $$n^{-1} \sum_{i=1}^{n} \{\psi(X_i/s_n) - \psi(X_i)\} = o\left(n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}\right) \quad \text{a.s.} ,$$

so that $T_n = O\left(n^{-\frac{1}{2}}(\log n)\right)$ a.s. by the proof of Theorem 2.1. By the same method as in the proof of Theorem 2.2, one shows that

(3.6) $$0 = n^{-1} \sum_{i=1}^{n} \psi(X_i/s_n) - (T_n/s_n)n^{-1} \sum_{i=1}^{n} \psi'(X_i/s_n) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

$$n^{-1} \sum_{i=1}^{n} \psi(X_i/s_n) = n^{-1} \sum_{i=1}^{n} \psi(X_i) + (s_n^{-1} - 1)n^{-1} \sum_{i=1}^{n} X_i\psi'(X_i) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

$$T_n\{n^{-1} \sum_{i=1}^{n} \psi'(X_i/s_n) - n^{-1} \sum_{i=1}^{n} \psi'(X_i)\} = O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

This completes the proof, since $(s_n-1)n^{-1}\sum\psi(X_i)=O\left(n^{-1}(\log n)^2\right)$ a.s.

It is clear that if $E_F X\psi'(X) = 0$, then results similar to Corollary 2.1 and Corollary 2.2 hold.

LINEAR FUNCTIONS OF ORDER STATISTICS

Let $X_1, X_2, \ldots$ be a sample from a distribution function $F(x)$. A linear function of the order statistics $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ is

(4.1) $$T_n = n^{-1} \sum_{i=1}^{n} J(i/n)\, X_{(i)} ,$$

where J is some weighting function. The trimmed mean is given by

(4.2) $$J(t) = (1 - 2\alpha)^{-1} \qquad \alpha \le t \le 1 - \alpha$$
$$= 0 \qquad \text{otherwise.}$$

de Wet and Venter (1974) show that the trimmed mean is the sum of i.i.d. random variables plus an almost sure error term of order $O(n^{-1} \log_2 n)$. This section shows that this almost sure representation holds for a wide class of weight functions J . The method is to extend the proof of Moore (1968).

Theorem 4.1. Suppose J' exists and satisfies a uniform Lipschitz condition of order one on $[0,1]$. Let $\theta = EXJ(F(X))$. Then there are mean 0 , i.i.d. random variables $Y_1, Y_2, \ldots$ such that for $\alpha > 0$,

$$T_n - \theta = n^{-1} \sum_{i=1}^{n} Y_i + o(n^{-3/4+\alpha}) \quad \text{a.s.} \tag{4.3}$$

If, in addition, $\left| \int_0^1 d\left(F^{-1}(u)\ J(u)\right) \right| < \infty$,

$$T_n - \theta = n^{-1} \sum_{i=1}^{n} Y_i + O(n^{-1} \log_2 n) \quad \text{a.s.} \tag{4.4}$$

Proof: Following Moore (1968), let $G(x) = F^{-1}(x)$, $R_i = F(X_i)$, $U_n(u)$ be the empirical distribution function of the R_i , $W_n(u) = n^{\frac{1}{2}}\left(U_n(u) - u\right)$, and

$$J\left(U_n(u)\right) - J(u) = J'\left(V_n^*(u)\right)\left(U_n(u) - u\right) ,$$

where $V_n^*(u) = \theta U_n(u) + (1 - \theta)u$ and $|\theta| \le 1$. Then, if

$$H(R_1) = \int J(u)\ [U_1(u) - u]\ dG(u) , \tag{4.5}$$

it follows that $T_n - \theta = I_{n1} + I_{n2} + I_{n3}$, where

(4.6) $$I_{n1} = -n^{-1} \sum_{i=1}^{n} H(R_i)$$

$$I_{n2} = n^{-\frac{1}{2}} \int_0^1 G(u) \; [J'(V_n^*(u)) - J'(u)] \; W_n(u) \; dU_n(u)$$

$$I_{n3} = n^{-1} \int_0^1 G(u)J'(u)W_n(u)dW_n(u) \; .$$

Now, $|I_{n2}| \le \sup|J'(V_n^*(n)) - J'(u)| \; \sup|U_n(u) - u| \int_0^1 |G(x)| \; dU_n(x)$, and $\int_0^1 |G(x)| \; dU_n(x) \to \int |G(x)| \; dx$ a.s. By the Law of the Iterated Logarithm for the Kolmogorov-Smirnov statistics and the Lipschitz condition on J' , it follows that $|I_{n2}| = O(n^{-1} \log_2 n)$ a.s. as $n \to \infty$. As in Moore's proof, $2|I_{n3}| = n^{-1} \left| \int_0^1 G(u)J'(u)d[W_n(u)]^2 \right| + o(n^{-1} \log_2 n)$ a.s. , so that

(4.7) $$2|I_{n3}| \le \int_0^1 |U_n(u) - u|^2 \; d(G(u)J'(u)) + o(n^{-1} \log_2 n) \quad \text{a.s.}$$

If $\left| \int_0^1 d(G(u)J'(u)) \right| < \infty$, this gives $|I_{n3}| = O(n^{-1} \log_2 n)$ a.s. Otherwise, denote the first term on the right hand side of (4.7) by B_n . By the moment relation stated in Proposition 2.3, one sees that $E|n(U_n(u) - u)|^8 \le c_0 n^4 H(u)$, where $H(u) \ge 0$ is integrable with respect to $G(u)J'(u)$. Thus, two applications of Schwarz's Inequality yield

$$E \; B_n^4 \le cn^4/n^8 = cn^{-4} \; .$$

By Chebychev's Inequality and the Borel-Cantelli Lemma, $n^{-3/4+\alpha} B_n \to 0$ a.s. , completing the proof.

<u>Theorem 4.2</u>. Suppose J, J', J'' are continuous and bounded except possibly at $\{a_1, \ldots, a_k\}$, and possess left and right limits. If F satisfies the

conditions of Bahadur (1966) at $\{a_1, \ldots, a_k\}$, the conclusions (4.3) and (4.4) hold, except that the order term in (4.4) is $O\left(n^{-1}(\log n)^2\right)$.

<u>Proof</u>: Following Moore (1968) again, we have that

$$(4.8) \quad T_n = \int_0^1 G(u)J'(u)V_n(u)du + \int_0^1 G(u)J(u)dV_n(u) + \int_0^1 G(u)\{J\left(U_n(u)\right) - J(u) - J'(u)V_n(u)\}\, dU_n(n) + \int_0^1 G(u)J'(u)V_n(u)dV_n(u) ,$$

where $V_n(u) = n^{-\frac{1}{2}}W_n(u) = U_n(u) - u$. The last term is I_{n3} appearing in Theorem 4.1. The first two terms of (4.8) can be decomposed (via integration by parts) into

$$(4.9) \quad \sum_{j=1}^{k} \{V_n(a_j)G(a_j)J(a_j-) - V_n(a_{j-1})G(a_{j-1})J(a_{j-1}+)\} + \int_0^1 J(u)V_n(u)dG(u) .$$

As in Theorem 4.1, the second term in (4.9) is the sum of i.i.d. mean 0 random variables. Call the first term $\sum_{j=1}^{k} A_{j,n}$. The third term of (4.8) is

$$(4.10) \quad B_n = \sum_{j=1}^{k+1} \int_{a_{j-1}+\beta(n)}^{a_j-\beta(n)} G(u)\{J\left(U_n(u)\right) - J(u) - J'(u)V_n(u)\}\, dU_n(u) + \sum_{j=1}^{k} \int_{a_j-\beta(n)}^{a_j+\beta(n)} G(u)\{J\left(U_n(u)\right) - J(u) - J'(u)V_n(u)\}\, dU_n(u) ,$$

where $\beta(n)$ converges to zero. By choosing $\beta(n) = O\left(n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\right)$ appropriately, the first term on the right hand side of (4.10) is $O(n^{-1} \log_2 n)$ a.s. , since $U_n(u)$ and u are (a.s.) in the intervals

$\left(a_{j-1} + \beta(n),\ a_j - \beta(n)\right)$ together as $n \to \infty$ and J' is differentiable there. By choosing $\beta(n) = n^{-\frac{1}{2}}(\log n)^{\frac{1}{2}}$, since $V_n(u) = o(\beta(n))$, Proposition 2.3 can be used to show that

$$\int_{a_j-\beta(n)}^{a_j+\beta(n)} G(u)J'(u)V_n(u)dU_n(u) = O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

Thus, we need only consider

$$\text{(4.10b)} \qquad \sum_{j=1}^{k} \int_{a_j-\beta(n)}^{a_j+\beta(n)} G(u)\{J\left(U_n(u)\right) - J(u)\}\, dU_n(u) \ .$$

Assume for the moment that $R_{[na_j]} \le a_j$, where $R_{[na_j]}$ is the $[na_j]^{\underline{th}}$ order statistic in $R_1, \ldots, R_n$. Define

$$B_{jn1} = \{k\colon a_j - \beta(n) < R_{(k)} < R_{[na_j]}\}$$

$$B_{jn2} = \{k\colon R_{[na_j]} \le R_{(k)} \le a_j\}$$

$$B_{jn3} = \{k\colon a_j < R_{(k)} < a_j + \beta(n)\} \ .$$

Then, for a given j, (4.10b) becomes

$$\text{(4.11)} \qquad n^{-1}\left(\sum_{B_{jn1}} + \sum_{B_{jn2}} + \sum_{B_{jn3}} G\left(R_{(k)}\right) \{J(k/n) - J(R_{(k)})\}\right) \ .$$

Now, de Wet and Venter (1974) show that the number of terms between a_j and $R_{[na_j]}$ is $O\left(n^{\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\right)$ a.s. ; because J' is bounded, one can use Proposition 2.3 to show that the first and last sums of (4.11) are $O\left(n^{-1}(\log n)^2\right)$ a.s. By a similar argument, the middle sum of (4.11) is

$$\begin{aligned}\text{(4.12)} \quad & n^{-1} \sum_{B_{jn2}} G(R_k)\, \{J(a_j+) - J(a_j-)\} + O\left(n^{-1}(\log n)^2\right) \text{ a.s.} \\ &= n^{-1}\, \{J(a_j+) - J(a_j-)\}\, \{\text{Sum of observations between } F^{-1}(a_j) \text{ and } \\ &\quad X_{[na_j]}\} + O\left(n^{-1}(\log n)^2\right) \text{ a.s.}\end{aligned}$$

Now, the first term of (4.9) is $-\sum_{j=1}^{k} V_n(a_j)G(a_j)\big(J(a_j+) - J(a_j-)\big)$, so that this term added to the first term on the right hand side of (4.12) becomes for $c_j = J(a_j+) - J(a_j-)$,

$$(4.13) \quad c_j\{n^{-1} \sum_{B_{jn2}} G(R_{(k)}) - G(a_j)V_n(a_j)\} = c_jn^{-1} \sum_{B_{jn2}} \big(G(R_{(k)} - G(a_j)\big)$$

$$+ c_jn^{-1}\{n\big(U_n(a_j) - U_n(R_{[na_j]})\big) - nV_n(a_j)\}\ G(a_j)$$

$$= c_jn^{-1}\{na_j - [na_j]\}\ G(a_j) + O\big(n^{-1}(\log n)^2\big) \quad \text{a.s.}$$

$$= O\big(n^{-1}(\log n)^2\big) \quad \text{a.s.}$$

The second equality in (4.13) follows by using Proposition 2.3, since $G(R_{[na_j]}) - G(a_j) = O\big(n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}\big)$ a.s. This completes the proof.

APPLICATIONS

We now apply the results of Sections 2 and 4 to get the asymptotic equivalence of M-estimators and LFO's and to obtain an expansion for flexible estimates of location. We first consider the equivalence, showing that if Jaeckel's (1971a) conditions are strengthened, the result follows easily from the representations and is much stronger than any result appearing in the literature.

Theorem 5.1. Suppose the following hold:

(5.1a) $\quad J(t) = A\psi'\big(F^{-1}(t)\big) = A\psi'(x)$, $\quad A^{-1} = E_F\psi'$.

(5.1b) ψ, J satisfy the conditions of Theorems 2.1 and 4.2 respectively, and F is continuous.

(5.1c) $$J(t) = \psi'(x) = 0 \quad \text{for} \quad t \notin [a,b] \ , \quad 0 < a < b < 1 \ .$$

(5.1d) M_n is the M-estimator formed by (2.1) and L_n is the LFO formed by J.

Then,

(5.2) $$M_n - L_n = O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

Proof: Because of Theorems 2.2 and 4.2, it suffices to show that $H(R_1) = -\int J(u)\ [U_1(u) - u]\ dG(u) = A\psi(X_1)$. Now,

$$\begin{aligned} H(R_1) &= -A \int \psi'\left(G(u)\right)\ [U_1(u) - u]\ dG(u) \\ &= -A \int \psi'(u)\ [F_1(y) - F(y)]\ dy \\ &= -A \int F_1(y)\ d\psi(y) + A \int F(y)\ d\psi(y) \\ &= \ A \int \psi(y)\ dF_1(y) = A\psi(X_1) \ . \end{aligned}$$

Corollary 5.1. The conditions of Theorem 5.1 hold, for example, if

(5.3a) F', F'' exist and are bounded away from zero on an interval $\alpha_0 - \varepsilon_0 \le F(x) \le 1 - (\alpha_0 - \varepsilon_0)$, $\alpha_0, \varepsilon_0 > 0$.

(5.3b) ψ has 3 continuous bounded derivatives at all but a finite number of points.

(5.3c) $$\psi'\left(F^{-1}(t)\right) = 0 \quad \text{if} \quad t < \alpha_0 \quad \text{or} \quad t > 1 - \alpha_0 \ .$$

We now turn to flexible estimates of location. We consider the M-estimators defined by ψ_k in (2.2) and the trimmed means given by $J_\alpha(t) = (1 - 2\alpha)^{-1}$ if $\alpha \le t \le 1 - \alpha$ and zero otherwise. We will assume F is symmetric for convenience, although some similar results hold if F is asymmetric. Further, some similar results hold for the scale invariant version of the M-estimators.

The asymptotic variances of these estimators are

$$(5.4) \qquad \sigma_M^2(k) = \{F(k) - F(-k)\}^{-2} \int \psi_k^2(x) dF(x)$$

$$\sigma_L^2(\alpha) = (1 - 2\alpha)^{-2} \{ \int_{x_\alpha}^{x_{1-\alpha}} x^2 dF(x) + 2\alpha x_\alpha^2 \} ,$$

where $x_\alpha = G(\alpha) = F^{-1}(\alpha)$. Denoting the M-estimator by M_{nk} and the α-trimmed mean by $L_{n\alpha}$, we estimate these asymptotic variances by

$$(5.5) \quad s_M^2(k) = \{F_n(k - M_{nk}) - F_n(-k - M_{nk})\}^{-2} \int \psi_k^2(x - M_{nk})\, dF_n(x)$$

$$s_L^2(k) = (1 - 2\alpha)^{-2} \{n^{-1} \sum_{i=[\alpha n]+1}^{n-[\alpha n]} (X_{(i)} - L_{n\alpha})^2\} + 2\alpha(X_{([\alpha n])} - L_{n\alpha})^2$$

$$= (1 - 2\alpha)^{-2} \left\{ \int_{X_{[\alpha n]+1)}}^{X_{(n-[\alpha n])}} (x - L_{n\alpha})^2 dF_n(x) + 2\alpha(X_{[\alpha n]} - L_{n\alpha})^2 \right\} .$$

We define the optimal estimators as follows: assume a range of permissible values $k_0 \le k \le k_1$ and $0 < \alpha_0 \le \alpha \le \alpha_1 < \frac{1}{2}$ fixed in advance. Compute $s_M^2(k)$ and $s_L^2(\alpha)$ for these ranges, and choose $\hat{k}_n$, $\hat{\alpha}_n$ which minimize $s_M^2(k)$, $s_L^2(\alpha)$ on these ranges. We will assume $\hat{k}_n$, $\hat{\alpha}_n$ are uniquely defined and that the values k^*, α^* minimizing $\sigma_M^2(k)$, $\sigma_L^2(\alpha)$ on the ranges are unique. If this latter assumption is not true, modifications along the lines suggested by Jaeckel (1971b) are possible. Define

$$C_M = \{k: k_0 \le k \le k_1\} \qquad C_L = \{\alpha: \alpha_0 \le \alpha \le \alpha_1\} .$$

We first consider the M-estimators. The following Lemma is an easy consequence of an extension of Theorems 2.1 and 2.2, along with the Law of the Iterated Logarithm for the empirical distribution function.

Lemma 5.1. Let k' be any point in C_M. Suppose that F is continuously differentiable in neighborhoods C^+, C^- of $\pm k'$ and that $\inf_{k \in C_M} |F(k) - F(-k)| > 0$. Then

$$\sup_{C^+} |M_{nk}| = o\left(n^{\frac{1}{2}} (\log n)\right) \quad \text{a.s.}$$

$$\sup_{C^+} \left|M_{nk} - \left(E_F \psi_k'(X)\right)^{-1} n^{-1} \sum_{i=1}^{n} \psi_k(X_i)\right| = O\left(n^{-1} (\log n)^2\right) \quad \text{a.s.}$$

If in addition F is continuously differentiable on C_M, then

$$\sup_{C_M} |s_M^2(k) - \sigma^2(k)| = O\left(n^{-\frac{1}{2}} (\log n)\right) \quad \text{a.s.}$$

Lemma 5.2. Under the conditions of Lemma 5.1, if $\sigma_M^2(k)$ is continuously differentiable in k on C_M, then

$$\hat{k}_n - k = O\left(n^{-\frac{1}{2}} (\log n)\right) \quad \text{a.s.}$$

Proof: Define $\beta(n) = M_1 n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}$ and $I_n = \{k \in C_M: k \notin \left(k^* - \beta(n), k^* + \beta(n)\right)\}$, where M_1 will be large. Let $\xi_n = \sigma_M^2(a_n) = \inf \sigma_M^2(k)$. By

Proposition 2.4, we have $\beta(n) \le |a_n - k^*| \le M_2\beta(n)$, so that

$$|\xi_n - \sigma_M^2(k^*)| = |a_n - k^*| \; \left(\sigma_M^2(a_n^*)\right)' ,$$

where $a_n^* \to k^*$. Thus, by choosing M_1 sufficiently large, we can find a sequence of constants $D(n) = O\left(n^{-\frac{1}{2}}(\log n)\right)$ such that

$$\tfrac{1}{2}\left(\xi_n - \sigma_M^2(k^*)\right) > D(n) > 0$$

$$\sup_{C_M} |s_M^2(k) - \sigma_M^2(k)| \le \tfrac{1}{2}D(n) \quad \text{a.s.} \quad \text{as} \quad n \to \infty .$$

Thus, with probability approaching one,

$$\begin{aligned} s_n^2(k) &< \sigma_M^2(k) + D(u) & k \in C_M \\ s_n^2(k) &> \sigma_M^2(k) - D(u) > \xi_n - D(u) & k \in I_n \end{aligned}$$

so that with probability approaching one,

$$s_n^2(k) - s_n^2(k^*) > 0 \quad \text{if} \quad k \in I_n ,$$

yielding the Lemma.

Thus by bringing Lemmas 5.1 and 5.2 together and by following the proof of Theorem 2.2, we obtain the following:

<u>Theorem 5.2</u>. Under the conditions of Lemmas 5.1 and 5.2,

$$M_{n,\hat{k}_n} - \theta = \{E_F\psi_{k^*}'(X)\}^{-1} \, n^{-1} \sum_{i=1}^{n} \psi_{k^*}(X_i) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

We now show that a result similar to Theorem 5.2 holds for the trimmed mean.

Lemma 5.3. Let α' be a fixed point and suppose that for neighborhoods C^1, C^2 of α', $1 - \alpha'$, F satisfies the conditions of Bahadur (1966). Then

$$(5.6) \qquad \sup_{C^1 \cup C^2} \left| L_{n\alpha} - n^{-1} \sum_{i=1}^{n} Y_i^*(\alpha) \right| = O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.} ,$$

where

$$\begin{aligned} Y_i^*(\alpha) &= G(\alpha)(1 - 2\alpha)^{-1} & X_i &< G(\alpha) \\ &= X_i(1 - 2\alpha)^{-1} & G(\alpha) \le X_i &\le G(1 - \alpha) \\ &= G(1 - \alpha)(1 - 2\alpha)^{-1} & X_i &> G(1 - \alpha) . \end{aligned}$$

If the Bahadur conditions hold on the set $\alpha_0 - \varepsilon_0 \le F(x) \le 1 - (\alpha_0 - \varepsilon_0)$ with $\varepsilon_0 > 0$, then the sup in (5.6) is over C_L and

$$(5.7) \qquad \sup_{C_L} \left| s_L^2(\alpha) - \sigma_L^2(\alpha) \right| = O\left(n^{-\frac{1}{2}}(\log n)\right) \quad \text{a.s.}$$

<u>Proof</u>: From Sen and Ghosh (1971), if $g(n) = n^{-\frac{1}{2}}(\log n)$,

$$\sup_{0<t<1} \sup_{|a| \le g(n)} \left| U_n(t + a) - U_n(t) + a \right| = O(n^{-3/4} \log n) \quad \text{a.s.}$$

This, together with the Law of the Iterated Logarithm for the empirical distribution function, implies that

$$\sup_{\alpha_0 \le \alpha \le 1-\alpha_0} \left| F(X_{([n\alpha])}) - F\left(G(\alpha)\right) \right| = O(n^{-\frac{1}{2}} \log n) \quad \text{a.s.}$$

$$\sup_{\alpha_0 \le \alpha \le 1-\alpha_0} \left| F_n(X_{([n\alpha])}) - F_n\left(G(\alpha)\right) \right| = O(n^{-\frac{1}{2}} \log n) \quad \text{a.s.}$$

The conditions of this Lemma then show that

$$\sup_{\alpha_0 \le \alpha \le 1-\alpha_0} \left| X_{([n\alpha])} - G(\alpha) \right| = O(n^{-\frac{1}{2}} \log n) \quad \text{a.s.}$$

Now, following the proof of de Wet and Venter (1974a), we see that (5.6) holds. Also, applying the exponential bounds to $n^{-1}\sum_{i=1}^{n} Y_i^*(\alpha)$ together with (5.6) yields

$$\sup_{C_L} |L_{n\alpha}| = O\left(n^{-\frac{1}{2}}(\log n)\right) \quad \text{a.s.}$$

The proof of (5.7) now follows easily.

Theorem 5.3. Under the conditions of Lemma 5.3, if $\sigma_L^2(\alpha)$ is continuously differentiable on C_L, then

$$\hat{\alpha}_n - \alpha^* = O\left(n^{-\frac{1}{2}}(\log n)\right) \quad \text{a.s.}$$

$$L_{n\hat{\alpha}_n} = n^{-1}\sum_{i=1}^{n} Y_i^*(\alpha^*) + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

The proof is as in Lemma 5.2 and Theorem 5.2.

Finally, if F satisfies the conditions of Bahadur (1966) at $G(\alpha)$ and $-G(1-\alpha)$, one can obtain some of the results of Shorack (1974) directly from Lemma 5.3. To be specific, if we estimate α by $\hat{\alpha}_n$ and if $\hat{\alpha}_n - \alpha = o(n^{-\frac{1}{4}})$ in probability, then under the initial conditions of Lemma 5.3, equation (5.6) tells us that

$$(5.8) \quad |L_{n\hat{\alpha}_n} - L_{n\alpha}| = \left|n^{-1}\sum_{i=1}^{n} Y_i^*(\hat{\alpha}_n) - n^{-1}\sum_{i=1}^{n} Y_i^*(\alpha)\right| + O\left(n^{-1}(\log n)^2\right) \quad \text{a.s.}$$

Now, use the last part of Proposition 2.3 to show that the first term on the right-hand-side of (5.8) is $o(n^{-\frac{1}{2}})$ in probability.

BIBLIOGRAPHY

ANDREWS, D.F., BICKEL, P.J., HAMPEL, F.R., HUBER, P.J., ROGERS, W.H., and TUKEY, J.W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton University Press.

BAHADUR, R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* (37) 577-580.

BICKEL, P.J. (1975). One step estimates in the linear model. To appear in *J. Amer. Statist. Assoc.*

CARROLL, R.J. (1974). Asymptotically nonparametric sequential selection procedures II - robust estimators. Institute of Statistics Mimeo Series #953, University of North Carolina at Chapel Hill.

DE WET, T. and VENTER, J.H. (1974). An asymptotic representation of trimmed means with applications. *S. Afr. Statist. J.* (8) 127-134.

DE WET, T. (1974). Rates of convergence of linear combination of order statistics. *S. Afr. Statist. J.* (8) 35-44.

FILIPPOVA, A.A. (1962). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. *Theory of Probability and its Applications* (7) 24-57.

HAMPEL, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* (69) 383-393.

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

HUBER, P.J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* (1) 221-233.

HUBER, P.J. (1972). Robust statistics: a review. *Ann. Math. Statist.* (43) 1041-1067.

JAECKEL, L.B. (1971a). Robust estimates of location: symmetry and asymmetric contamination. *Ann. Math. Statist.* (42) 1020-1034.

JAECKEL, L.B. (1971b). Some flexible estimates of location. *Ann. Math. Statist.* (42) 1540-1552.

JOHNSON, N.L., and KOTZ, S. (1969). *Discrete Distributions.* Houghton Mifflin, Boston.

LOEVE, M. (1963). *Probability Theory.* D. Van Nostrand, Princeton.

MOORE, D.S. (1968). An elementary proof of asymptotic normality of linear functions of order statistics. *Ann. Math. Statist.* (39) 263-265.

SEN, P.K., and GHOSH, M. (1971). On bounded length sequential confidence intervals based on one-sample rank order statistics. *Ann. Math. Statist.* (42) 189-203.

SHORACK, G.R. (1974). Random means. *Ann. Statist.* (2) 661-675.