AMS 1970 Subject Classifications: Primary 62F07, 62G35; Secondary 62L99, 62E20

Key Words and Phrases: Ranking and Selection, Robust, M-estimators, Sequential Ranking, Nonparametric Selection, Linear Functions of Order Statistics

ASYMPTOTICALLY NONPARAMETRIC SEQUENTIAL SELECTION PROCEDURES II - ROBUST ESTIMATORS

by

Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #953
October, 1974

ASYMPTOTICALLY NONPARAMETRIC SEQUENTIAL SELECTION PROCEDURES II - ROBUST ESTIMATORS

by

Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill

SUMMARY

Let $\pi_1, \ldots, \pi_K$ be K independent populations with distributions $F(x; \theta_i)$ $(i = 1, \ldots, K)$ which are stochastically ordered; the basic ranking goal is to select the stochastically largest population. Sequential selection rules which are nonparametric in an asymptotic sense are given which solve the problem by using M-estimators and L-estimators (Huber (1972)). The results are also applied to selection of the largest location parameter and to construction of fixed-width confidence intervals for location parameters using the above robust estimators.

AMS 1970 Subject Classifications: Primary 62F07, 62G35; Secondary 62L99, 62E20

Key Words and Phrases: Ranking and Selection, Robust, M-estimators, Sequential Ranking, Nonparametric Selection, Linear Functions of Order Statistics.

1. INTRODUCTION

Let $\pi_1, \ldots, \pi_K$ be K independent populations with distributions $F(x; \theta_i)$ $(i = 1, \ldots, K)$ where θ_i is some unknown indexing parameter and $F(x; \theta_i)$ is unknown. The notation $\theta_i \le \theta_j$ will mean that $F(x; \theta_i) \ge F(x; \theta_j)$ for all x, i.e., $F(x; \theta_i)$ is stochastically smaller than $F(x; \theta_j)$. If $\theta_{[1]} \le \cdots \le \theta_{[K]}$ is the true (unknown) correct ordering, the ranking goal is to devise sequential procedures for selecting the stochastically largest population, which are in some sense nonparametric and robust.

In Carroll (1974), sequential rules of the form of Chow and Robbins (1965) were investigated and a general theorem (Theorem 1.1 below) was proposed for solving the problem using the indifference zone formulation of Bechhofer (1954); the sample median (under fairly weak conditions) and the sample mean (under rather restrictive conditions) were shown to satisfy this Theorem. In Section 3, the conditions on the sample mean are greatly relaxed.

The sample mean, however, is notoriously non-robust, which in this context implies that entirely too many observations are being taken in most cases. In order to reduce the number of observations and to obtain desired robustness properties, this paper shows under fairly weak conditions that the robust M-estimators of Huber (1964) and Hampel (1974) and certain linear functions of order statistics such as the trimmed mean also satisfy the conditions of Theorem 1.1; the proofs are given in Sections 4 - 6. The results, when specialized to the case of location parameters $\left(F(x;\theta) = F(x - \theta)\right)$, yield the first proofs that the above estimators may be used to select the largest location parameter and to construct fixed-width confidence intervals.

For future use, define

$$P^* = \int \Phi^{K-1}(x + b)d\Phi(x) \tag{1.1}$$

$$\Omega(\delta) = \{(\theta_1, \ldots ,\theta_K) \mid \theta_{[K]} - \theta_{[K-1]} \geq \delta\} , \tag{1.2}$$

where Φ is the distribution of the standard normal. Thus, for ease of presentation, the parameters $\theta_1, \ldots ,\theta_K$ will be assumed to be numbers on the real line; however, it is very easy to extend the notation to cope with more arbitrary indexing sets. If "CS" indicates a correct selection, the goal of the ranking procedure will be to guarantee

$$\liminf_{\delta\to 0} \; \inf_{\Omega(\delta)} P(CS) = P^* . \tag{1.3}$$

For $i = 1, \ldots ,K$, independent observations $X_{i1}, \ldots ,X_{in}$ are taken from $F(\circ ; \theta_i)$ and statistics $T_i(n)$ and $\sigma_i(n)$ are formed such that

$$\sigma^{-1}(\theta_i)n^{1/2}\left(T_i(n) - \theta_i\right) \xrightarrow{\mathcal{L}} \Phi \tag{1.4}$$

$$\sigma_i(n) \to \sigma(\theta_i) \quad \text{a.s.} \tag{1.5}$$

The ranking procedure itself is to take $N_i(\delta)$ observations from Π_i $(i = 1, \ldots ,K)$, form the statistics $T_i\left(N_i(\delta)\right)$, and make the natural decision, where

$$N_i(\delta) = \text{first integer } n \geq \left(b\sigma_i(n)/\delta\right)^2 . \tag{1.6}$$

Remark 1.1. The stopping rules (1.6) are independent of one another, which is certainly a drawback. However, improvement here must await the discovery of a nonparametric elimination rule for the location case. In the location case, if one chooses

$$\sigma^2(n) = K^{-1} \sum_{i=1}^{K} \sigma_i^2(n) \ , \tag{1.7}$$

the stopping rule would be to take $N(\delta)$ observations from each population, where

$$N(\delta) = \text{first integer } n \geq \left(b\sigma(n)/\delta\right)^2 \ . \tag{1.8}$$

Using the generic terms T_n, σ_n, θ, and $\sigma(\theta)$, Carroll (1974) proved the following:

THEOREM 1.1. If the parameter space $\Omega(0)$ is compact, then (1.3) holds if

$$\sigma(\theta) \to \sigma(\theta_0) \quad \text{as} \quad \theta \to \theta_0 \tag{A1}$$

and if for each θ_0, $\varepsilon > 0$, $\beta > 0$, and $z \varepsilon \mathbf{R}$, $\exists$ J, $\eta > 0$, $c > 0$ such that $n \geq J$ and $|\theta - \theta_0| \leq \eta$ imply

$$P_\theta\{|\sigma_n - \sigma(\theta)| > \sigma(\theta)\varepsilon \text{ for some } n \geq J\} < \varepsilon \tag{A2}$$

$$|P_\theta\{\sigma(\theta)^{-1} n^{1/2}(T_n - \theta) \leq z\} - \Phi(z)| < \varepsilon \tag{A3}$$

$$P_\theta\{n^{1/2} |T_n - T_m| > \beta \text{ for some } m \text{ with } |m - n| < cn\} < \varepsilon \ . \tag{A4}$$

Note that (A2) and (A3) mean that (1.4) and (1.5) hold continuously. (A4) is an extension of Anscombe's (1952) condition.

COROLLARY 1.1. If $F(x;\theta) = F(x - \theta)$ and (1.4) and (1.5) hold, then, using $N(\delta)$ given in (1.8), (1.3) holds if (A4) holds at $\theta = 0$ and if

(1.9) $$T_n + \theta = T_n(X_1, \ldots, X_n) + \theta = T_n(X_1 + \theta, \ldots, X_n + \theta)$$

(1.10) $$\sigma_n = \sigma_n(X_1 + \theta, \ldots, X_n + \theta) \ .$$

Remark 1.2. In Carroll (1974) there was also a condition that for all $\varepsilon > 0$, there exist η and d such that $|\theta - \theta_0| < \eta$ implies

(1.11) $$P_\theta\{\sigma_n > d \quad n = 1,2, \ldots \} \ge 1 - \varepsilon \ .$$

This condition may be removed in Theorem 1.1 simply by defining σ_n to be larger than some constant, say 100^{-1000}. The condition is unnecessary in Corollary 1.1.

The following convention will be used:

(1.12) $$X_n(\theta) \to X(\theta_o) \text{ a.s. uniformly}$$

if for all $\varepsilon > 0$, $\exists$ N, η such that $|\theta - \theta_0| < \eta$ implies

(1.13) $$P_\theta\{|X_n(\theta) - X(\theta_o)| > \varepsilon \text{ for some } n \ge N\} \le \varepsilon \ .$$

A similar meaning is attached to other types of convergence.

2. PRELIMINARY RESULTS

The Propositions of this section will be used repeatedly, and while fairly simple, should be of some interest in themselves.

Proposition 2.1. Let A be an indexing set, and let $Y_n(\alpha)$ $(\alpha \varepsilon A)$ be random variables with distributions $F(\circ;\alpha)$ with zero mean and uniformly bounded variances (say by M). Then if $0 \le \delta < 1/2$, and if $S_n(\alpha) = n^{-1} \sum_{j=1}^{n} Y_j(\alpha)$,

$$n^{\delta} S_n(\alpha) \to 0 \quad \text{a.s. uniformly .} \tag{2.1}$$

Proof: The proof follows the "method of sequences" of Chung (1968). Let $a = 1 + (1 - 2\delta)^{-1}$, where δ is chosen so that a is an integer. Then, by Chebyshev's inequality,

$$n^{a\delta} S_{n^a}(\alpha) \to 0 \quad \text{a.s. uniformly .} \tag{2.2}$$

Let $m(a) = n^a$, and

$$n^a D_n(\alpha) = \max \Big| \sum_{j=m(a)+1}^{K} Y_j(\alpha) \Big| , \tag{2.3}$$

where the max is taken over $n^a + 1 \le K < (n + 1)^a$. Kolmogorov's Inequality yields

$$n^{a\delta} D_n(\alpha) \to 0 \quad \text{a.s. uniformly ,} \tag{2.4}$$

and the proof is completed by noting that for $n^a \le K < (n + 1)^a$,

$$K^{\delta} |S_K(\alpha)| \le (n + 1)^{a\delta} \left(|S_{m(a)}(\alpha)| + |D_n(\alpha)|\right) . \tag{2.5}$$

<u>Proposition 2.2.</u> Let $X_i(\theta)$ have mean 0 and variance $\sigma^2(\theta)$ under $F(\circ;\theta)$. Then

(2.6) $$\sigma^{-1}(\theta)n^{1/2} S_n(\theta) \overset{L}{\to} \Phi \text{ uniformly as } \theta \to \theta_0$$

if for any sequence $\theta_1, \theta_2, \ldots$ converging to θ_0 , and for all $\varepsilon > 0$,

(2.7) $$\sigma^{-2}(\theta_n) \int_{A_n(\varepsilon)} x^2 dF(x;\theta_n) \to 0 ,$$

where $A_n(\varepsilon) = \{x: |x| > \sigma(\theta_n)\varepsilon n^{1/2}\}$.

<u>Proof:</u> This follows immediately from the Normal Convergence Criterion given in Loéve (1963), page 295.

<u>Corollary 2.1.</u> (2.6) holds if $X_i(\theta)$ are uniformly bounded random variables and if $\sigma^2(\theta)$ is continuous in θ .

<u>Proposition 2.3.</u> (Schuster (1969)) Let $F(x)$ be a distribution and let $F_n(x)$ be the empirical distribution. Then, there is a universal constant C such that

(2.8) $$P_F\{\sup_x |F_n(x) - F(x)| > \varepsilon\} \leq C \exp\{-2n\varepsilon^2\} .$$

<u>Corollary 2.2.</u> For $0 \leq \delta < 1/2$,

(2.9) $$n^\delta \sup_x |F_n(x) - F(x)| \to 0 \text{ a.s. uniformly} .$$

3. SAMPLE MEANS

In this section, Theorem 1.1 is proved for sample means and variances under much less restrictive conditions than those given in Carroll (1974). No symmetry assumptions are made, nor are there any conditions on the inverse of $F(x;\theta)$.

THEOREM 3.1. Suppose (2.7) holds, and that

(3.1) $$\mathrm{Var}_\theta(X) \text{ is continuous in } \theta$$

(3.2) $E_\theta\left(X - E_\theta(X)\right)^4$ is bounded in some neighborhood of every θ_0 . Then (A1) - (A4) hold.

Proof: (A2) holds because of (3.2) and Proposition (2.1). Then (3.1) will imply (A1). (A3) holds because of Proposition 2.2, and (A4) follows by extending Anscombe's (1952) proof by means of Kolmogorov's Inequality and (3.1).

4. HUBER'S M-ESTIMATORS

Let $X_1, X_2, \ldots$ have a distribution $F(x;\theta)$, and let ψ be an increasing skew-symmetric function for which $E_\theta\psi(X - \theta) = 0$. It will be assumed throughout that the θ satisfying this equation is unique. Huber (1964) defines an M-estimator T_n as the solution to the equation

(4.1) $$n^{-1} \sum_{j=1}^{n} \psi(X_j - T_n) = 0 .$$

Note that T_n , while location invariant, is not scale invariant. To get this, one needs some kind of scale invariant measure of dispersion (such as the interquartile range) S_n , and then defines T_{n2} by

$$n^{-1} \sum_{j=1}^{n} \psi\left(\frac{X_j - T_{n2}}{S_n}\right) = 0 \quad (4.2)$$

In this section, conditions are given for which the solutions to (4.1) and (4.2) satisfy (A1) - (A4). Note that one immediately learns from this that some Huber M-estimators satisfy the conditions of Geertsema (1972), and thus may be used in nonparametric robust selection and confidence intervals. The differentiability conditions on the ψ functions are needed only to invoke Taylor's Theorem but do not seem to be crucial. Although Huber's favorite ψ function $\left(\psi_0(x) = x \text{ for } |x| < K , \psi_0(x) = K \text{ sign } x \text{ for } |x| \geq K\right)$ does not satisfy the differentiability conditions, it can be uniformly approximated by "nice" functions.

It will be assumed throughout that ψ, ψ' , and ψ'' exist and are bounded respectively by $M(\psi)$, $M(\psi')$, and $M(\psi'')$. It will further be assumed that

$$F(x;\theta) \to F(x;\theta_0) \text{ as } \theta \to \theta_0 . \quad (4.3)$$

For ease of exposition, the estimators T_n formed from (4.1) will first be studied. The asymptotic variance of T_n under $F(x;\theta)$ is

$$\sigma^2(\theta) = \frac{\int \psi^2(t-\theta)dF(t;\theta)}{\left\{\int \psi'(t-\theta)dF(t;\theta)\right\}^2} \quad (4.4)$$

and will be estimated by

(4.5) $$\sigma^2_{n\theta} = \frac{\int \psi^2(t-T_n)dF_n(t;\theta)}{\left\{\int \psi'(t-T_n)dF_n(t;\theta)\right\}^2},$$

where $F_n(t;\theta)$ is the empirical distribution based on a sample of size n from $F(x;\theta)$.

The proof of the following Lemma follows easily from The Mean Value Theorem and boundedness of ψ' and ψ'', together with application of the Helly-Bray Theorem and Proposition 2.3.

LEMMA 4.1. Assume (4.3) holds. Then (A1) and (A2) hold, if $T_n - \theta \to 0$ a.s. uniformly.

The following Lemma becomes useful after it is shown that T_n is almost the sum of i.i.d. random variables. Until Proposition 4.1 is established, assume $T_n - \theta \to 0$ a.s. uniformly.

LEMMA 4.2. Let $\beta^2(\theta) = E_\theta\psi^2(X - \theta)$. Then,

(4.6) $$\frac{n^{1/2}}{\beta(\theta)}\left(\frac{1}{n}\sum_{i=1}^{n} \psi(X_i - \theta)\right) \xrightarrow{L} \Phi \quad \text{uniformly}.$$

Proof: Since ψ is bounded, it is sufficient to show by Corollary 2.1 that $\beta^2(\theta)$ is continuous in θ. This follows in a manner similar to Lemma 4.1.

It will now be shown that (A3) holds. By a Taylor's expansion,

(4.7) $$n^{1/2}(T_n - \theta) = \frac{n^{-1/2}\sum_{i=1}^{n}\psi(X_i-\theta)}{n^{-1}\sum_{i=1}^{n}\psi'(X_i-\theta)} + \{n^{1/4}(T_n - \theta)\}^2 \frac{n^{-1}\sum_{i=1}^{n}\psi''(Z_i(\theta))}{n^{-1}\sum_{i=1}^{n}\psi'(X_i-\theta)}$$

where $Z_i(\theta)$ is between $X_i - \theta$ and $T_n - \theta$.

Since

(4.8) $$n^{-1} \sum_{i=1}^{n} \psi'(X_i - \theta) \to E_\theta \psi'(X - \theta) \quad \text{a.s. uniformly}$$

by Proposition 2.1, and since ψ'' is bounded, by Lemma 4.2, (A3) follows if for $0 \le \delta < 1/2$,

(4.9) $$n^{\delta}(T_n - \theta) \to 0 \quad \text{a.s. uniformly} .$$

This requires two steps.

Proposition 4.1. $T_n - \theta \to 0$ a.s. uniformly.

Proof: Since ψ is increasing, if $|\theta - \theta_0| < \eta/2$,

(4.10) $$n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) \le n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta_0 + \eta)$$

$$n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta) \ge n^{-1} \sum_{i=1}^{n} \psi(X_i - \theta_0 - \eta) .$$

By invoking Proposition 2.1, the proof is complete.

LEMMA 4.3. For $0 \le \delta < 1/2$, if $\psi'(0) > 0$ and $F(\theta_0 + \varepsilon; \theta_0) - F(\theta_0 - \varepsilon; \theta_0) > b(\varepsilon) > 0$ for all $\varepsilon > 0$, then

(4.11) $$n^{\delta}(T_n - \theta) \to 0 \quad \text{a.s. uniformly} ,$$

and (A3) holds.

Proof: By invoking Taylor's Theorem,

$$(4.12) \qquad n^{\delta}(T_n - \theta) = \frac{n^{\delta}\left\{n^{-1}\sum_{i=1}^{n}\psi(X_i-\theta)\right\}}{n^{-1}\sum_{i=1}^{n}\psi'\big(Z_i(\theta)\big)}$$

where $Z_i(\theta)$ is between $X_i - \theta$ and $T_n - \theta$. By Proposition (2.1), it suffices to show that

$$(4.13) \qquad n^{-1}\sum_{i=1}^{n}\psi'\big(Z_i(\theta)\big) \geq c > 0 \text{ a.s. uniformly as } n \to \infty .$$

But this is true since $\psi'(x) \geq 0$, with $\psi'(0) > 0$, by using Proposition 4.1.

THEOREM 4.1. If $\psi'(0) > 0$ and $F(\theta_o + \varepsilon; \theta_o) - F(\theta_o - \varepsilon; \theta_o) \geq b(\varepsilon) > 0$ for $\varepsilon > 0$, and if (4.3) holds, then (A1) - (A4) hold for T_n defined by (4.1).

Proof: It is sufficient to check (A4), the extended version of uniform continuity in probability (Anscombe (1952)). From (4.7) and Lemma 4.3, (A4) need only be checked for the random variables

$$(4.13) \qquad H_n = \frac{n^{-1}\sum_{i=1}^{n}\psi(X_i-\theta)}{n^{-1}\sum_{i=1}^{n}\psi'(X_i-\theta)} .$$

Since, for $0 \leq \delta < 1/2$

$$(4.14) \qquad n^{\delta}\Big(n^{-1}\sum_{i=1}^{n}\{\psi'(X_i - \theta) - E_{\theta}\psi'(X - \theta)\}\Big) \to 0 \text{ a.s., uniformly,}$$

this follows from Anscombe's proof and Kolmogorov's Inequality.

Note that the above results do not require symmetry of $F(x;\theta)$ about θ, although this seems to be a common condition (compare Huber (1964), Geertsema (1972), Sen and Ghosh (1971)). However, for investigating the estimators T_{n2} formed from (4.2), symmetry will have to be imposed (because of Lemma 4.5). It will also be assumed throughout the rest of this section that

(4.15) $$\psi'(x) = \psi'(-x) .$$

The estimators S_n will be assumed to satisfy

(4.16) $$cS_n(X_1, \ldots, X_n) = S_n\big(c(X_1 + a), \ldots, c(X_n + a)\big) \qquad \text{for real } a$$

(4.17) for some $\xi(\theta)$, and for all $0 \le \delta < 1/2$,

$$n^{\delta}\big(S_n - \xi(\theta)\big) \to 0 \quad \text{a.s. uniformly} .$$

Using results of Bahadur (1966), one may find relatively mild conditions for which the interquartile range satisfies (4.17). Use of the interquartile range for S_n is studied in Andrews, et al (1972). The asymptotic variance of T_n under $F(x;\theta)$ is now

(4.18) $$\sigma^2(\theta) = \xi(\theta)^2 \frac{\int \psi^2\left(\frac{t-\theta}{\xi(\theta)}\right) dF(t;\theta)}{\left\{\int \psi'\left(\frac{t-\theta}{\xi(\theta)}\right) dF(t;\theta)\right\}^2}$$

and is estimated by

(4.19) $$\sigma^2_{n\theta} = \frac{S_n^2 \int \psi^2\left(\frac{t-T_n}{S_n}\right) dF_n(t;\theta)}{\left\{\int \psi'\left(\frac{t-T_n}{S_n}\right) dF_n(t;\theta)\right\}^2} .$$

Much of the work below follows in a manner similar to the proofs for the estimates defined by (4.1); only those proofs which require extra effort are presented.

Proposition 4.2. $T_{n2} - \theta \to 0$ a.s. uniformly.

LEMMA 4.4. Assume that (4.3) holds. Then (A1) and (A2) hold.

LEMMA 4.5. Let $\beta^2(\theta) = E\,\psi^2\left(\frac{X-\theta}{\xi(\theta)}\right)$. Then, if $E_\theta\,|X - \theta|^2$ is bounded in some neighborhood of each θ_0 ,

(4.20) $$\beta^{-1}(\theta) n^{1/2}\left\{n^{-1}\sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{S_n}\right)\right\} \overset{L}{\to} \Phi \text{ uniformly .}$$

Proof: Using Taylor's Theorem twice, it suffices to show

(4.21) $$\left\{n^{-1/2}\sum_{i=1}^{n}|X_i - \theta|^2\right\}|S_n - \xi(\theta)|^2 \to 0 \text{ a.s. uniformly}$$

(4.22) $$\{S_n - \xi(\theta)\}\, n^{-1/2}\sum_{i=1}^{n}\psi'\left(\frac{X_i-\theta}{\xi(\theta)}\right)\left(\frac{X_i-\theta}{\xi(\theta)}\right) \to 0 \text{ a.s. uniformly.}$$

(4.21) follows from (4.17) and Chebyshev's Inequality. Since $\psi'\left(\frac{X-\theta}{\xi(\theta)}\right)\left(\frac{X-\theta}{\xi(\theta)}\right)$ is symmetric about 0 , $E_\theta\psi'\left(\frac{X-\theta}{\xi(\theta)}\right)\left(\frac{X-\theta}{\xi(\theta)}\right) = 0$, so that (4.22) follows from (4.21) and Proposition 2.1.

LEMMA 4.6. For $0 \le \delta < 1/2$, under the conditions of Lemma 4.3,

(4.22) $$n^\delta(T_{n2} - \theta) \to 0 \text{ a.s. uniformly .}$$

Now for the main theorem of this section, stating explicitly all the necessary conditions.

THEOREM 4.2. Suppose the conditions of Theorem 4.1 hold, that $F(x;\theta)$ is symmetric about θ, that $\psi'(x) = \psi'(-x)$, that $E_\theta|X - \theta|^2$ is bounded in some neighborhood of each θ_0, and that S_n satisfies (A4). Then (A1) - (A4) hold for T_{n2} defined by (4.2).

Proof: By Taylor's Theorem

$$(4.23)\qquad T_{n2} - \theta = S_n\left\{\frac{n^{-1}\sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{S_n}\right)}{n^{-1}\sum_{i=1}^{n}\psi'\left(\frac{X_i-\theta}{S_n}\right)}\right\} + S_n\left(\frac{T_{n2}-\theta}{S_n}\right)^2\left\{\frac{n^{-1}\sum_{i=1}^{n}\psi''(Z_{in})}{n^{-1}\sum_{i=1}^{n}\psi'\left(\frac{X_i-\theta}{S_n}\right)}\right\}.$$

By Proposition 2.1, and Lemmas 4.5 and 4.6, (A3) is completed. Also, it suffices to prove (A4) for

$$H_n = n^{-1}\sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{S_n}\right)\left\{n^{-1}\sum_{i=1}^{n}\psi'\left(\frac{X_i-\theta}{S_n}\right)\right\}^{-1}$$

$$(4.24)\qquad = \frac{n^{-1}A_n + A_n^*\left(S_n - \xi(\theta)\right)\left(S_n\xi(\theta)\right)^{-1} + o(n^{-1/2})}{E_\theta\psi'\left(\frac{X-\theta}{\xi(\theta)}\right) + o(n^{-\delta})}$$

by expanding each term in the numerator about $\frac{X_i-\theta}{\xi(\theta)}$, where

$$(4.25)\qquad A_n = \sum_{i=1}^{n}\psi\left(\frac{X_i-\theta}{\xi(\theta)}\right)$$

$$(4.26)\qquad A_n^* = \sum_{i=1}^{n}\psi'\left(\frac{X_i-\theta}{\xi(\theta)}\right),$$

and $h_n = o(g_n)$ means $h_n/g_n \to 0$ a.s. uniformly. Let $c(\theta) = E\psi'\left(\frac{X-\theta}{\xi(\theta)}\right)$. Then, assuming (without loss of generality) that $m > n$,

$$(4.27) \qquad T_n - T_m = c(\theta)\left(\frac{1}{n} - \frac{1}{m}\right)A_n + c(\theta)m^{-1}B_m$$
$$+ c(\theta)(n^{-1}A_n^*)\left(\frac{S_n-\xi(\theta)}{\xi(\theta)S_n} - \frac{S_m-\xi(\theta)}{\xi(\theta)S_m}\right)$$
$$+ o(n^{-1/2}) \; .$$

Since S_n satisfies (A4), this completes the proof.

Remark 4.1. Carroll (1974) has shown that the interquartile range S_n satisfies (A4) if

$$(4.28) \quad C_\theta|x - y| \le |F^{-1}(F(x;\theta_0);\theta) - F^{-1}(F(y;\theta_0);\theta)| \le B_\theta|x - y|$$

(where C_θ, $B_\theta \to 1$ as $\theta \to \theta_0$) for all x,y in some neighborhood of $\xi_{1/4}$ and $\xi_{3/4}$, the relevant quartiles of $F(x;\theta_0)$.

5. M-ESTIMATORS OF HAMPEL AND OTHERS

Hampel (1974) (see also Andrews, et al (1972)) proposed estimates defined as the solution of (4.1) or (4.2) closest to the median, but generated by three-step ψ functions, the prototype of which is

$$(5.1) \qquad \psi(x) = -\psi(-x) = \begin{cases} x & 0 \le x < a \\ a & a \le x < b \\ \frac{c-x}{c-b}\,a & b \le x < c \\ 0 & x \ge c \;. \end{cases}$$

These estimators seem to have very good robustness properties. Again, as in Section 4, the conditions which guarantee (A1) - (A4) will not be strictly satisfied by (5.1), but, as before, this will not present much of a problem in the applications.

There are a number of problems in showing that estimators defined by general ψ functions (e.g., by (5.1)) satisfy (A1) - (A4). First of all, the solution to

$$\lambda(\beta;\theta) = \int \psi(x - \beta)dF(x;\theta) = 0 \tag{5.2}$$

may not be unique. It will be assumed that on any closed interval contained in the support of $F(\circ,\theta)$, the number of solutions to (5.2) is finite for each θ . A quick check will show that the increasing nature of ψ was used in Proposition 4.1, Lemma 4.3, and in the fact that

$$E_\theta \psi'(X - \theta) > 0 \ . \tag{5.3}$$

From now on, (5.3) will also be assumed.

To get results analagous to Theorems 4.1 and 4.2 for general ψ functions, one merely needs the conditions of those theorems, the two conditions mentioned above, and the conditions in the two lemmas below. For simplicity, only the estimators defined by (4.1) will be used. Lemma 5.1 is based on a proof of Huber (1967).

LEMMA 5.1. $T_n - \theta \to 0$ a.s. uniformly if $\exists\ A_o > 0$ such that for all $\varepsilon > 0$, $\exists$ an N and an η for which $|\theta - \theta_o| < \eta$ implies

$$P_\theta\{|T_n - \theta| > A_o \text{ some } n \geq N\} \leq \varepsilon \ . \tag{5.4}$$

PROOF: Note that this is obvious if ψ is like (5.1), since then if less then $n/100$ X_i satisfy $|X_i - \theta| \geq A$, $|T_n - \theta| < 2A$ for A large, and one can invoke Proposition 2.1 and (4.3).

Now, if $T_n - \theta \to 0$ a.s. uniformly does not hold, there is a sequence (N_j, θ_j) with $\theta_j \to \theta_0$, and an $\varepsilon_0 > 0$ for which

$$(5.5) \qquad P_{\theta_j}\{|T_n - \theta_j| \geq \varepsilon_0 \text{ some } n \geq N_j\} \geq \varepsilon_0 .$$

Since the number of zeros of $\lambda(\beta; \theta_0)$ is finite and λ is continuous in its arguments, for all open subsets U of $K = [-A, A]$, there is a small ε such that for n large enough,

$$(5.6) \qquad |\lambda(\beta; \theta_j)| \geq 5\varepsilon > 0 \quad j \geq n \quad , \quad \beta\varepsilon K - U$$

$$(5.7) \qquad U \text{ contains all zeros of } \lambda(\beta; \theta_j) \quad , \quad j \geq n .$$

Since

$$(5.8) \qquad |\lambda(\beta'; \theta) - \lambda(\beta;\theta)| \leq M(\psi')|\beta' - \beta| ,$$

$\forall\ \beta\varepsilon K - U$, one can choose U_β for which $\beta'\varepsilon U_\beta$ implies

$$(5.9) \qquad |\lambda(\beta'; \theta_j) - \lambda(\beta; \theta_j)| \leq \varepsilon_0 \quad j \geq n .$$

By using the proof of Huber (1967), one finds

$$(5.10) \qquad P_{\theta_j}\{T_n \varepsilon U \text{ for all } n \geq N_j\} \geq 1 - \varepsilon \quad \text{for } j \text{ large} .$$

Since T_n is the solution of (4.1) closest to the median, this completes the proof.

LEMMA 5.2. Assume that for each θ_0, there is $A > 0$ for which

(5.11) $$0 < b_0 \le \psi'(y) \quad , \quad y \varepsilon [-A, A]$$

(5.12) $$\exists\, \eta_0, \beta > 0 \text{ such that if } |\theta - \theta_0| < \eta_0 ,$$

$$b_0 P_\theta\{X - \theta \varepsilon [-A, A]\} \ge \max_{|y|>A} |\psi'(y)| P_\theta\{|X - \theta| > A\} + \beta .$$

Then, $\exists\, d > 0$ such that for all $\varepsilon > 0$, $\exists\, N$, η such that $|\theta - \theta_0| < \eta$ implies

(5.13) $$P_\theta\left\{\frac{1}{n} \left|\sum_{i=1}^{n} \psi'\big(Z_{in}(\theta)\big)\right| \ge d \text{ for all } n \ge N\right\} \ge 1 - \varepsilon ,$$

where $Z_{in}(\theta)$ is derived from the Taylor's expansion of $\sum_{i=1}^{n} \psi(X_i - T_n)$ about $X_i - \theta$.

Proof: The proof follows easily from the conditions and Lemma 5.1.

Remark 5.1. The conditions of Lemma 5.2 will be satisfied by (5.1) if $F(x;\theta)$ puts enough probability near θ. Note that (5.13) is exactly what is needed in Lemma 4.3.

6. L-ESTIMATORS

Linear functions of order statistics, especially the trimmed mean, are quite popular robust estimators. If $X_1, X_2, \ldots, X_n$ are independent with continuous distribution $F(x;\theta)$, and if $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$ denote the order statistics, then the L-estimators are

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

where J is some function. The α-trimmed mean is defined by

$$J(t) = 0 \qquad \text{if} \quad t \varepsilon \; [\alpha, 1-\alpha] \tag{6.2}$$
$$= (1-2\alpha)^{-1} \qquad t \in [\alpha, 1-\alpha] \; .$$

In this section, the α-trimmed mean is shown to satisfy (A1) - (A4) under fairly minimal conditions, and then general conditions on J are given. The method of proof here closely follows Moore(1968) who proved the following:

<u>THEOREM 6.1.</u> Let

$$\sigma^2(\theta) = 2 \int_x \int_y J\big(F(x;\theta)\big) J\big(F(y;\theta)\big) F(x;\theta) \quad \{1 - F(y;\theta)\} dx dy < \infty \tag{6.3}$$

$$E_\theta |X| < \infty \tag{6.4}$$

(6.5) J be continuous on [0,1] except for jump discontinuities at $a_1, \ldots, a_M$, and J' is continuous and of bounded variation on $[0,1] - \{a_1, \ldots, a_M\}$, and F^{-1} is continuous at $a_1, \ldots, a_M$.

Then

$$n^{1/2}\left\{T_n - \int xJ\big(F(x;\theta) \quad dF(x;\theta)\right\} \xrightarrow{L} N\big(0, \sigma^2(\theta)\big) \; . \tag{6.6}$$

<u>LEMMA 6.1.</u> Suppose J satisfies (6.2) and that the distribution of

$$Y_i = \int_0^1 J(x)[U_i(u) - u] dF^{-1}(x;\theta) \tag{6.7}$$

satisfies (2.7), where

(6.8) $$U_i(u) = 0 \quad \text{if} \quad F(X_i;\theta) > u$$
$$= 1 \quad \text{otherwise} .$$

Then (A1) - (A4) hold with $\sigma^2_{n\theta}$ being the obvious estimate of $\sigma^2(\theta)$ based on $F_n(x;\theta)$.

Proof: (A1) and (A2) obviously hold from (4.3). Let

(6.9) $$w_n(u) = n^{1/2}\left(U_n(u) - u\right)$$

where $U_n(u)$ is the empirical distribution of the uniform random variables $R_i = F(X_i;\theta)$. Then, since $F^{-1}(x;\theta) \to F^{-1}(x;\theta_0)$, using Proposition 2.3 yields with probability 1 ,

(6.10) $$n^{1/2}\left[T_n - \int xJ\left(F(x;\theta)\right)dF(x;\theta)\right] = o(n^{-1/2}) + \int_0^1 F^{-1}(x;\theta)J(x)dw_n(x) .$$

Moore's (1968) proof for his I_{1n} now suffices, together with the proof of Theorem 3.1, since Moore arranges the right-hand side of (6.10) as a sum of i.i.d. random variables.

THEOREM 6.2. Suppose that (6.5) holds and that (A1) and (A2) are true (which will be the case in translation parameter families or if J vanishes outside a compact set). Then if $E_\theta|X|$ is bounded in some neighborhood of each θ_0 , and if the random variables in (6.7) satisfy (2.7), then (A3) and (A4) hold.

Proof: The proof will be given here only in the case where J' is continuous. Following Moore (1968),

(6.11) $$n^{1/2}\left[T_n - \int xJ\left(F(x;\theta)\right)dF(x;\theta)\right] = I_{n1} + I_{n2} + I_{n3} ,$$

where

$$(6.12)\qquad I_{n1} = \int_0^1 F^{-1}(x;\theta)J'(x)w_n(x)dx + \int_0^1 F^{-1}(x;\theta)J(x)dw_n(x)$$

$$I_{n2} = \int_0^1 F^{-1}(x;\theta)\Big[J'\big(V_n(x)\big) - J'(x)\Big]w_n(x)dU_n(x)$$

$$I_{n3} = n^{-1/2}\int_0^1 F^{-1}(x;\theta)J'(x)w_n(x)dw_n(x)$$

where $V_n(x)$ is between $U_n(x)$ and x. By Proposition 2.3, $I_{n2} \to 0$ a.s. uniformly, and by Theorem 3.1 since

$$(6.13)\qquad I_{n1} = -\int_0^1 J(x)w_n(x)dF^{-1}(x;\theta) ,$$

$$(6.14)\qquad I_{n1} \xrightarrow{\mathcal{L}} \Phi \text{ uniformly} .$$

Thus, to show (A3) and (A4), it suffices to show that $K_n(\theta)$ defined by

$$(6.15)\qquad K_n(\theta) = \int_0^1 \left\{n^{1/4}\big(U_n(x) - x\big)\right\}^2 dJ'(x)F^{-1}(x;\theta)$$

satisfies

$$(6.16)\qquad K_n(\theta) \to 0 \text{ in probability uniformly}$$

$$P_\theta\{|K_m| > \varepsilon \text{ for some } m \text{ such that } \left|\frac{m}{n} - 1\right| < c\} \le \varepsilon .$$

This will be true if for $|\theta - \theta_0| < \eta$, there exists K^* for which

$$(6.17)\qquad E_\theta K_n^2(\theta) \le n^{-1} K^* .$$

By using Hölder's inequality and bringing the expectation inside, one gets

$$(6.18)\qquad nE_\theta K_n^2(\theta) \le n^{-2}\int_0^1 \left\{3\big(nx(1-x)\big)^2 + nx(1-x)\big(1-6x(1-x)\big)\right\} dJ'(x)F^{-1}(x;\theta)$$

Thus, it will be sufficient to prove that if $V(x)$ is the variation of $J'(y)F^{-1}(y;\theta)$ on $[1/2, x]$,

$$(6.19)\qquad \int_{1/2}^{1} x(1-x)dV(x) < K_0 \quad \text{if} \quad |\theta - \theta_0| < \eta .$$

Since there is a uniform constant C such that $V(x) \le CF^{-1}(x;\theta)$, integration by parts and the bound on $E|X - \theta|$ establishes (6.19) and completes the proof.

7. FIXED WIDTH CONFIDENCE INTERVALS AND SELECTION

In this section, the results in Sections 4 - 6 are discussed in relationship to the important and often discussed problems of selecting the largest location parameter (see Robbins, Sobel and Starr (1968), Geertsema (1972)) and constructing a fixed-width confidence interval for a parameter (see Chow and Robbins (1965), Sen and Ghosh (1971)).

For the selection of the largest location, suppose one has K populations $\pi_1, \dots, \pi_K$ with distributions $F(x - \theta_i)$ $(i = 1, \dots, K)$, where $\theta_1, \dots, \theta_K$ are unknown, and F is unknown. If $\theta_{[1]} \le \theta_{[2]} \le \dots \le \theta_{[K]}$ denotes the correct (unknown) ordering, then one wishes to prove that for suitably defined statistics T_n ,

$$(7.1)\qquad \liminf_{\substack{\delta\to 0 \\ \Omega(\delta)}} P(CS) = P^*$$

independently of F , where the stopping rule (following Geertsema (1972)) is to take $N(\delta)$ observations from each population, where, for
$P^* = \int \Phi^{K-1}(x + b)d\Phi(x)$,

$$N(\delta) = \text{first integer } n \geq (b\sigma_n/\delta)^2 \tag{7.2}$$

and σ_n^2 is the estimate of the variance. Since the asymptotic normality has been established for all the statistics T_n , one merely needs a strongly consistent estimate σ_n^2 of the variance and (A4) holding at $\theta = 0$. The conditions thus implied are summed up in the following theorem.

THEOREM 7.1. (7.1) holds

a) In *Theorem 4.1* if $\psi'(0) > 0$ and $F(\varepsilon) - F(-\varepsilon) > 0 \ \forall \ \varepsilon > 0$.

b) In *Theorem 4.2* if a) holds, F is symmetric about 0 and possesses a second moment, and if S_n satisfies (A4) (which it will if it is the interquartile range).

c) For *Hampel-estimators defined by (4.1) (or 4.2)* if a) (or b)) holds, $E\psi'(X) > 0$ under F , (5.2) has only a finite number of solutions at 0 , and (5.12) holds at $\theta = 0$.

d) In *Lemma 6.1* and *Theorem 6.2* if $E|X|$ exists under F and F is continuous and strictly increasing.

Remark 7.1. (7.1) holds for each F ; one can use the results of this paper and Carroll (1974) to find conditions under which (7.1) holds in the location problem uniformly over some compact subsets of the space of distributions.

The results of this paper may also be easily applied to finding fixed-width confidence intervals for a parameter. Suppose the distribution of $X_1, X_2, \ldots$ is $F(x - \theta)$, and that T_n is a statistic in this paper or in Carroll (1974), for which

(7.3) $$n^{1/2}(T_n - \theta) \xrightarrow{L} \text{Normal}\left(0, \sigma^2(\theta)\right) .$$

If α is the desired confidence level and

(7.4) $$\alpha = \Phi(b) - \Phi(-b) ,$$

one takes $N(\delta)$ observations, where

(7.5) $$N(\delta) = \text{first integer } n \geq (b\sigma_n/\delta)^2 .$$

The confidence interval formed is $I_{N(\delta)}$, where

07.6) $$I_n = (T_n - \delta , T_n + \delta) .$$

Then, under the conditions of Theorem 7.1,

(7.7) $$\lim_{\delta \to 0} P_\theta\{\theta \varepsilon I_{N(\delta)}\} = \alpha .$$

Remark 7.2. The problem of finding general conditions under which (A3) holds (say for sums of dependent or independent r.v.'s) is of interest in itself and will be the subject of a later paper.

REFERENCES

[1] 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.

[2] ANSCOMBE, F. J., (1952). Large sample theory of sequential estimation. *Proc. Camb. Phil. Soc.* (48) 600-617.

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

[4] BECHHOFER, R. E., (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* (25) 16-39.

[5] CARROLL, R. J. (1974). Asymptotically nonparametric sequential selection procedures. *Inst. of Stat. Mimeo Series* #944, Univ. of North Carolina.

[6] CHOW, Y. S., and ROBBINS, H., (1965). On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Ann. Math. Statist.* (36) 463-467.

[7] CHUNG, K. L., (1968). A Course in Probability Theory. Harcourt, Brace, and World, Inc.

[8] GEERTSEMA, J. C., (1972). Nonparametric sequential procedures for selecting the best of k populations. *J. Am. Statist. Assoc.* (67) 614-616.

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

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

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

REFERENCES (cont.)

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

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

[14] ROBBINS, H., SOBEL, M., and STARR, N., (1968). A sequential procedure for selecting the largest of k means. *Ann. Math. Statist.* (39) 88-92.

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

[16] SCHUSTER, E. F., (1969). Estimation of a probability density function and its derivatives. *Ann. Math. Statist.* (40) 1187-1195.