ON THE ASYMPTOTIC NORMALITY OF STOPPING TIMES BASED ON ROBUST ESTIMATORS*

by

Raymond J. Carroll

Department of Statistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series #1027

September, 1975

* This work was partially supported by the Office of Naval Research, contract N00014-67-A-0226-00014 at Purdue University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

SUMMARY

The asymptotic normality of certain stopping times in fixed-width interval analysis is discussed when the intervals are based on M-estimates of location. Using results of Ghosh and Mukhophadyay (1975), it suffices to consider the asymptotic normality of estimates σ_n^2 of the variance of M-estimators under random sample sizes. Two methods under differing sets of conditions are given; the first is based on finding almost sure representations for σ_n^2, while the second is based on the theory of weak convergence. The final results are also applied to one-step M-estimators (Bickel (1975)) to obtain almost sure representations and weak convergence results.

KEY WORDS AND PHRASES: Fixed-Width Confidence Intervals, Robust Estimation, M-Estimators, Sequential Analysis, Stopping Times, Almost Sure Invariance Principles

AMS 1970 SUBJECT CLASSIFICATIONS: Primary: 62L12; Secondary 62E20, 62G35.

Introduction

This paper is motivated by, but not restricted to, the problem of the asymptotic normality of stopping times in sequential analysis. The stopping times $N(d)$ are generally defined as follows: if $\{g(d)\}$ is a sequence of constants tending to zero as $d \to 0$, and if Y_n is a sequence of positive statistics, then $N(d)$ is the first time n that $n \geq Y_n/g(d)$. As long as $Y_n \to \sigma^2$ almost surely (a.s.) as $n \to \infty$, then for $n(d) = [\sigma^2/g(d)]$, $N(d)/n(d) \to 1$ (a.s.) as shown by Chow and Robbins (1965). Ghosh and Mukhophadyay (1975) show under these conditions that if there are positive constants a,b for which

$$N(d)^{\frac{1}{2}}(Y_{N(d)}-a)/b \xrightarrow{L} N(0,1) \quad \text{as} \quad d \to 0,$$

then

$$\left(ag(d)/b^2\right)^{\frac{1}{2}}\left(N(d)-a/g(d)\right) \xrightarrow{L} N(0,1) \quad \text{as} \quad d \to 0.$$

It is important to realize that $\{Y_n\}$ and σ^2 do not exist in a vacuum and are typically connected to an underlying estimation problem. The prototype is the construction of fixed-width confidence sequences for a location parameter when the observations $X_1, X_2, \ldots$ come from a distribution $F\left(\xi^{-1}(x-\theta)\right)$; one bases the interval on a sequence $\{T_n\}$ of location-scale equivariant estimators (see Bickel (1975) for definitions) and σ^2 is the asymptotic variance of the normed sequence $\{n^{\frac{1}{2}}(T_n-\theta)\}$. Thus, in this problem, $\{Y_n\}$ is simply an estimate (scale equivariant but location invariant) of σ^2. In this context, Ghosh and Mukhophadyay (1975) discussed in detail the case where $\{T_n\}$ is a sequence of U-statistics.

In light of the above discussion, it seems natural to investigate the problem of the asymptotic normality of stopping times when T_n is a robust estimator other than a U-statistic. We consider specifically the two cases where T_n is an M-estimator (Huber (1964), Andrews et. al. (1972)) or a one-step M-estimator (Bickel (1975)). Here the asymptotic variance is

$$(1.1) \qquad \sigma^2 = \int \psi^2(x)dF(x)/\{\int \psi'(x)dF(x)\}^2.$$

We show in Section 6 that it is only necessary to estimate the following functional of F:

$$(1.2) \qquad \int \rho(x)dF(x),$$

where ρ is a known function.

Thus, the main body of this paper discusses the following estimation problem, leaving until Section 6 the applications to the asymptotic normality of stopping times based on M-estimators. We are interested in estimating (1.2). There is assumed to be a sequence of constants $n(d)$ and integer-valued random variables $N(d)$ with $N(d)/n(d) \to 1$ in probability as $d \to 0$. We estimate (1.2) by

$$(1.3) \qquad \sigma_n^2 = S_n \int \rho\left(\frac{x-T_n}{S_n}\right)dF_n(x) = S_n\left\{n^{-1} \sum_{i=1}^{n} \rho\left(\frac{X_i-T_n}{S_n}\right)\right\},$$

where T_n is as above and S_n is a robust location invariant, scale equivariant estimator of scale such as the interquartile range.

The goal is to find reasonably weak conditions on ρ which guarantee that for some $A>0$,

$$(1.4) \qquad A\, N(d)^{\frac{1}{2}}\left(\sigma^2_{N(d)}-\sigma^2\right) \xrightarrow{L} N(0,1) \quad \text{as} \quad d \to 0,$$

where the convergence indicated is convergence in law.

There will be two approaches to the estimation of (1.2), both of which yield (1.4) as a corollary. In Sections 2 and 3, almost sure representations are obtained for σ_n^2 by making assumptions about (i) the almost sure behavior of T_n and S_n and (ii) the asymptotic distributions of $T_{N(d)}$ and $S_{N(d)}$; the conditions on ρ are mild and we do not require differentiability.

In Sections 4 and 5, we obtain (1.4) by means of the theory of weak convergence of stochastic processes with multidimensional time parameters (Billingsley (1968), Bickel and Wichura (1971)). In this approach we make no assumptions concerning the almost sure behavior of T_n and S_n, but the conditions on ρ are stronger (but still do not include differentiability).

In Section 6 we return to the original problem of proving (1.4) when σ^2 is given by (1.1). Mild conditions are given for (1,4) to hold and these conditions are satisfied in most cases of interest. We are also able to obtain almost sure representations and weak convergence results for the one-step M-estimates studied by Bickel (1975).

2. Preliminary (a.s.) Results

One estimator of the functional $\int \rho(x)dF(x)$ which is location invariant but not scale invariant is

$$n^{-1} \sum_{i=1}^{n} \rho(X_i - T_n), \tag{2.1}$$

where T_n is a sequence of location and scale equivariant statistics converging (in some sense) under F_θ to θ; in the rest of this paper, we assume (unless indicated) that $\theta=0$, so the distribution function is $F(x)$.

In order to find a representation for the estimator (2.1), we want to avoid global differentiability properties and instead make assumptions about the behavior of certain integrals. If ρ has two continuous bounded derivatives and $n^{\frac{1}{2}}(\log n)^{-1}T_n \to 0$ almost surely (a.s.), a Taylor's expansion shows

$$(2.2)\quad n^{-1}\sum_{i=1}^{n}\rho(X_i-T_n) = n^{-1}\sum_{i=1}^{n}\rho(X_i) - T_nE\rho'(X) + O\left(n^{-1}(\log n)^2\right) \quad \text{(a.s.)},$$

where $a_n = O(b_n)$ if $|a_n/b_n|$ is bounded as $n\to\infty$. The purpose of this section is to derive a result similar to (2.2) with the order term $O\left(n^{-3/4}(\log n)^2\right)$; the proof is elementary and makes no differentiability assumptions concerning ρ. We first study a process $V_n(t)$ given below; the major result will be obtained by looking at $V_n(a_n^{-1}T_n)$. All proofs will be delayed until the end of the section.

Definition 2.1. For some sequence of constants $\{a_n\}$ decreasing to zero,

$$V_n(t) = n^{-1}\sum_{i=1}^{n}\left\{\rho(X_i-ta_n)-\rho(x_i)-E\left(\rho(X-ta_n)-\rho(X)\right)\right\}.$$

The next Proposition and Lemma give the almost sure rates at which V_n converges to zero.

Proposition 2.1. Suppose ρ is increasing. Then there exists a constant $M>0$ such that

$$(2.3)\qquad \sup\{|V_n(t)|:\ 0\le t\le 1\} \le M(A_{1n}+A_{2n}),$$

where

$$(2.4)\qquad A_{1n} = \sup_{0\le k\le n-1}\left|E\{\rho(X-a_nk/n)-\rho\left(X-a_n(k+1)/n\right)\}\right|$$

$$A_{2n} = \sup_{0\le k\le n}|V_n(k/n)|.$$

Lemma 2.1. Suppose ρ is increasing and bounded and that for sequences $\{B_n\}$, $\{C_n\}$ converging to zero

(2.5) $$|E\{\rho(x-B_n-C_n) - \rho(X-B_n)\}^r| = O(|C_n|) \qquad (r=1,2)$$

(2.6a) $$\text{For all } c>0, \quad \sum_{n=1}^{\infty} n^2 \exp\{-cn/b_n\} < \infty$$

(2.6b) $$\text{For all } c>0, \quad \sum_{n=1}^{\infty} n^2 \exp\{-cn/(b_n^2 a_n)\} < \infty.$$

Then

$$\sup\{b_n|V_n(t)|: 0\le t\le 1\} \to 0 \quad \text{(a.s.)}.$$

The major result of this section, Theorem 2.1, finds a representation for the estimator (2.1) if T_n converges to zero (a.s.); note that the conditions on ρ are minimal. The proof is omitted since it is a consequence of Lemma 2.1 with $t = a_n^{-1}T_n$.

Theorem 2.1. Suppose $\rho = \rho^+-\rho^-$, where both ρ^+ and ρ^- are increasing and bounded. Suppose $\{a_n\}$, $\{b_n\}$, ρ^+, ρ^- satisfy the conditions of Lemma 2.1. Then

$$\sup_{|t|\le 1} b_n|V_n(t)| \to 0 \quad \text{(a.s.)}.$$

Thus if $a_n^{-1}T_n \to 0$ (a.s.)

(2.7) $$n^{-1}\sum_{i=1}^{n} \{\rho(X_i-T_n)-E\rho(X)\} = n^{-1}\sum_{i=1}^{n} \{\rho(X_i)-E\rho(X)\} + \int \{\rho(y-T_n)-\rho(y)\}dF(y) + o(b_n^{-1}) \quad \text{(a.s.)},$$

where $a_n = o(b_n)$ means $a_n/b_n \to 0$. If $a_n^{-1} = (\log n)^2$, then $b_n = n^{\frac{1}{2}}$, while if $a_n^{-1} = n^{\frac{1}{2}}(\log n)^{-1}$, $b_n = n^{3/4}(\log n)^{-2}$.

It is of interest to remove the integral in (2.6) and obtain a result similar to (2.2). The proof follows immediately from (2.7) and (2.8).

Corollary 2.1. If in addition to the conditions of Theorem 2.1, as $|h| \to 0$

$$\int \{\rho(x+h) - \rho(x)\}dF(x) = A(F)h + O(|h|^2), \tag{2.8}$$

then

$$n^{-1} \sum_{i=1}^{n} \{\rho(X_i - T_n) - E\rho(X)\} = n^{-1} \sum_{i=1}^{n} \{\rho(X_i) - E\rho(X)\} - A(F)T_n + o(b_n^{-1}) \quad \text{(a.s.)}. \tag{2.9}$$

Further, if for some function ψ $(E\psi(X)=0)$

$$T_n = n^{-1} \sum_{i=1}^{n} \psi(X_i) + o(b_n^{-1}) \quad \text{(a.s.)}, \tag{2.10}$$

we have

$$n^{-1} \sum_{i=1}^{n} \{\rho(X_i - T_n) - E\rho(X)\} = n^{-1} \sum_{i=1}^{n} \{\rho(X_i) + \psi(X_i) - E\rho(X)\} + o(b_n^{-1}) \quad \text{(a.s.)}.$$

In Corollary 2.2, the proof of which is omitted, we show that a wide class of non-differentiable functions ρ satisfy (2.5) and (2.8). This class includes the psi functions due to Huber and Hampel (see Andrews, et. al. (1972)).

Corollary 2.2. Suppose ρ is twice boundedly differentiable except at a finite number of points, and F is Lipschitz of order one in neighborhoods of these points. Then ρ satisfies (2.5) and (2.8). In particular, if $\rho(x) = I_{\{|x| \le k\}}$ or $\rho(x) = x\, I_{\{|x|\ k\}} + k\ \mathrm{sign}(x) I_{\{|x|>k\}}$, then $\rho = \rho^+ - \rho^-$, where ρ^+, ρ^- are increasing and satisfy (2.5) and (2.8).

The discussion of the asymptotic normality of the stopping rule based on (2.1) is postponed until the next section.

Proof of Proposition 2.1. Let $[\cdot]$ denote the greatest integer function. Then, since

$$|V_n(t)| \le |V_n(t)-V_n([nt]/n)| + |V_n([nt]/n)|,$$

by the monotonicity of ρ we obtain

$$\text{(2.11)}\quad \sup_{0\le t\le 1} |V_n(t)| \le \sup_{0\le k\le n-1} \left|V_n(k/n)-V_n\big((k+1)/n\big)\right| + 2 \sup_{0\le k\le n-1} \left|E\Big(\rho\big(X-a_n(k+1)/n\big)-\rho\big(X-a_n k/n\big)\Big)\right| + \sup_{0\le k\le n} |V_n(k/n)|.$$

Proof of Lemma 2.1. Using the result of Proposition 2.1, we see that by (2.5) and (2.6a),

$$b_n A_{1n} \to 0.$$

To prove that $b_n A_{2n} \to 0$ (a.s.), we make use of the Borel-Cantelli Lemma and the exponential bounds (see Loeve (1968), page 254). Specifically, letting $s_{nk}^2 = n\,\mathrm{Var}\big(\rho(X-a_n k/n)-\rho(X)\big)$, we have for $\varepsilon_0>0$,

$$\Pr\left\{\sup_{0\le k\le n} |V_n(k/n)| > \varepsilon_0/b_n\right\} \le \sum_{k=0}^{n} \Pr\left\{\frac{|nV_n(k/n)|}{s_{nk}} > \frac{\varepsilon_0 n}{b_n s_{nk}}\right\} .$$

Under (2.6a) and (2.6b), the two possible cases of the exponential bounds lead to the last sum being bounded above for some $M>0$ by one of

$$\sum_{k=0}^{n} \exp\{-Mn/(b_n^2 a_n)\} \quad \text{or} \quad \sum_{k=0}^{n} \exp\{-Mn/b_n\}.$$

This completes the proof.

3. An (a.s.) Representation

A second estimator of the functional $\int \rho(x)dF(x)$ which is both location and scale invariant is

$$(3.1) \qquad n^{-1} \sum_{i=1}^{n} \rho\left(\frac{X_i - T_n}{S_n}\right)$$

where T_n is as in Section 2 and S_n is a location invariant, scale equivariant estimator which converges to the scale parameter ξ; we assume throughout that $\xi=1$. A version which is location invariant but scale equivariant (and which would be used in practice) is

$$(3.1)^* \qquad S_n\, n^{-1} \sum_{i=1}^{n} \rho\left(\frac{X_i - T_n}{S_n}\right).$$

We find it more convenient to work with (3.1), returning to the study of (3.1)* after Corollary 2.1. Again, from a Taylor's expansion

$$(3.2) \quad n^{-1} \sum_{i=1}^{n} \rho\left(\frac{X_i - T_n}{S_n}\right)$$

$$= n^{-1} \sum_{i=1}^{n} \rho(X_i) - \{EX\rho'(X)\}(S_n - 1) - \{E\rho'(X)\}T_n + \mathcal{O}\left(n^{-1}(\log n)^2\right) \quad \text{(a.s.)}.$$

Note the rather surprising fact, also seen by Carroll (1975) in his study of M-estimators, that unless $EX\rho'(X) = 0$ the estimator (3.1) has an asymptotic distribution different from the estimator (2.1). This section gives a result on the order of (3.2) without differentiability assumptions. The process of interest is:

Definition 3.1. For a sequence of constants $\{a_n\}$ decreasing to zero,

$$V_n(t,u) = n^{-1}\sum_{i=1}^{n}\{\rho\big((1+ua_n)(X_i-ta_n)\big)-\rho(X_i-ta_n)-E\big(\rho(1+ua_n)(X-ta_n)-\rho(X-ta_n)\big)\}.$$

The outline of this section is similar to that of Section 2. Again, the proofs of all necessary results are delayed until the end of the section.

Proposition 3.1. Suppose ρ is increasing. Then there exists a constant $M>0$ such that

$$\sup\{|V_n(t,u)|:\ 0\le t,u\le 1\} \le M\{A_{1n}+A_{2n}+A_{3n}+A_{4n}+A_{5n}+A_{6n}\}, \tag{3.3}$$

where

$$A_{1n} = \sup_{0\le t,u\le 1}\left|E\{\rho\big((1+ua_n)(X-a_n\tfrac{[nt]}{n})\big)-\rho\big((1+a_n\tfrac{[nu]}{n})(X-a_n\tfrac{[nt]}{n})\big)\}\right| \tag{3.4}$$

$$A_{2n} = \sup_{0\le t\le 1}\left|n^{-1}\sum_{i=1}^{n}\{\rho(X_i-a_nt)-\rho(X_i)-E\big(\rho(X-a_nt)-\rho(X)\big)\}\right|$$

$$A_{3n} = \sup_{0\le j,k\le n}|V_n(k/n,j/n)|$$

$$A_{4n} = \sup_{0\le t,u\le 1}\left|n^{-1}\sum_{i=1}^{n}\{\rho\big((1+a_n\tfrac{[nu]}{n})(X_i-a_n\tfrac{[nt]}{n})\big)-\rho\big((1+a_n\tfrac{[nu]+1}{n})(X_i-a_n\tfrac{[nt]}{n})\big)\}I_{\{X_i>a_n\frac{[nt]}{n}\}}\right|$$

$$A_{5n} = \sup_{0\le t,u\le 1}\left|n^{-1}\sum_{i=1}^{n}\{\rho\big((1+a_n\tfrac{[nu]}{n})(X_i-a_n\tfrac{[nt]}{n})\big)-\rho\big((1+a_n\tfrac{[nu]+1}{n})(X_i-a_n\tfrac{[nt]}{n})\big)\}I_{\{X_i\le a_n\frac{[nt]}{n}\}}\right|,$$

where I_B denotes the indicator function of the set B, and

$$A_{6n} = \sup_{0\le t,u\le 1} \left| E\{\rho\left((1+ua_n)(X-a_n\frac{[nt]+1}{n})\right)-\rho\left((1+ua_n)(X-a_n\frac{[nt]}{n})\right)\}\right|.$$

Lemma 3.1. Let ρ be bounded and increasing and suppose that equations (2.6a) and (2.6b) hold and that if A_n, B_n, C_n converge to zero,

$$(3.5)\quad \left|\int \{\rho\left((1+A_n)(X-B_n-C_n)\right)-\rho\left((1+A_n)(X-B_n)\right)\}^r dF(x)\right| = O(|C_n|) \quad \text{for } r=1,2$$

$$(3.6)\quad \left|\int \{\rho\left((1+A_n+B_n)(X-C_n)\right)-\rho\left((1+A_n)(X-C_n)\right)\}^r dF(x)\right| = O(|B_n|) \quad \text{for } r=1,2,$$

where the integrals in (3.5) and (3.6) are taken over the real line on one of the sets $\{X<D_n\}$, $\{X>D_n\}$ and D_n converges to zero. Then

$$\sup\{b_n|V_n(t,u)|: 0\le t,u\le 1\} \to 0 \text{ (a.s.)}.$$

Theorem 3.1. Suppose $\rho = \rho^+-\rho^-$ where ρ^+, ρ^- satisfy the conditions of Lemma 3.1. Suppose further that $\{a_n\}$, $\{b_n\}$ satisfy (2.6a) and (2.6b). Then

$$\sup_{|u|,|t|\le 1} |b_n V_n(t,u)| \to 0 \text{ (a.s.)}.$$

Hence, if $a_n^{-1}(T_n-\theta) \to 0$ (a.s.) and $a_n^{-1}(S_n-\xi) \to 0$ (a.s.) under $F\left(\xi^{-1}(x-\theta)\right)$, then under $F(x)$,

$$(3.7)\quad n^{-1}\sum_{i=1}^{n}\left\{\rho\left(\frac{X_i-T_n}{S_n}\right)-E_F\rho(X)\right\} = n^{-1}\sum_{i=1}^{n}\{\rho(X_i)-E_F\rho(X)\} + \int\left\{\rho\left(\frac{y-T_n}{S_n}\right)-\rho(y)\right\}dF(y) + o(b_n^{-1}) \quad \text{(a.s.)}.$$

The following Corollary gives our most specific result for the estimator (3.1). It shows that

$$n^{-1}\sum_{i=1}^{n}\left\{\rho\left(\frac{X_i-T_n}{S_n}\right)-E\rho(X)\right\}$$

satisfies the Law of the Iterated Logarithm if S_n, T_n do and is asymptotically normally distributed if

$$\left(n^{-1}\sum_{i=1}^{n}\{\rho(X_i)-E\rho(X)\},\ S_n-1,\ T_n\right)$$

are jointly asymptotically normally distributed (when properly normed).

Corollary 3.1. Under the conditions of Theorem 3.1, if as $h,q \to 0$

$$(3.8)\quad \int \{\rho\big((1+h)(X+q)-\rho(X)\big)\}dF(x) = A(F)h+B(F)q+O(|h|^2)+O(|q|^2)+O(|hq|),$$

then equation (3.7) becomes

$$(3.9)\quad n^{-1}\sum_{i=1}^{n}\rho\left(\frac{X_i-T_n}{S_n}\right)-E\rho(X)$$
$$= n^{-1}\sum_{i=1}^{n}\{\rho(X_i)-E\rho(X)\}-A(F)(S_n-1)-B(F)T_n+O(|a_n|^2)+O(|b_n|)\quad \text{(a.s.)}.$$

We note that if we had chosen the scale equivariant estimator (3.1)*, then (3.9) would have become

$$(3.9)^*\quad S_n\, n^{-1}\sum_{i=1}^{n}\rho\left(\frac{X_i-T_n}{S_n}\right)-E\rho(X)$$
$$= n^{-1}\sum_{i=1}^{n}\{\rho(X_i)-E\rho(X)\}+\{E\rho(X)-A(F)\}(S_n-1)-B(F)T_n+O(|a_n|^2)+O(|b_n|)$$
$$\text{(a.s.)}.$$

Finally, the results are again applicable for a wide variety of functions ρ.

Corollary 3.2. Suppose ρ is twice continuously differentiable except at a finite number of points and that F is Lipschitz of order one in neighborhoods of these points. Then ρ satisfies (3.8) with $A(F) = EX\rho'(X)$ and $B(F) = E\rho'(X)$, so that if F is symmetric and $\rho(X) = -\rho(-X)$, $A(F) = 0$.

Corollary 3.3. Define $\rho(x)$ by

$$\rho(x) = 1 \quad \text{if} \quad |x| \le k$$
$$= 0 \quad \text{otherwise.}$$

Then if F is Lipschitz of order one in neighborhoods of $\pm k$, ρ satisfies the conditions of Theorem 3.1 and Corollary 3.1. If in addition $E|X| < \infty$, the same results hold for the Huber function

$$\rho(x) = x \qquad \text{if} \quad |X| \le k$$
$$= k \ \text{sign}(x) \quad \text{otherwise.}$$

Corollary 3.1 yields in Corollary 3.4 simple conditions under which the stopping rules described in Section 1 (and based on either (3.1) and (3.1)*) are asymptotically normally distributed. The following is easily shown because of Anscombe's (1952) Theorems 1 and 4; the additional conditions on T_n, S_n seem unavoidable.

Corollary 3.4. Let T_n, $S_n - 1$ be uniformly continuous in probability (Anscombe (1952)) and suppose

$$\left(n^{\frac{1}{2}} T_n,\ n^{\frac{1}{2}}(S_n - 1),\ n^{-\frac{1}{2}} \sum_{i=1}^{n} \{\rho(X_i) - E\rho(X)\}\right)$$

are jointly asymptotically normally distributed. Consider a sequence of integer-valued random variables $N(d)$ and constants $n(d)$ for which $N(d)/n(d) \to 1$ in probability. If (3.9) and (3.9)* hold, both

(3.10a)
$$N(d)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \left\{\rho\left(\frac{X_i - T_{N(d)}}{S_{N(d)}}\right) - E\rho(X)\right\}$$

and

(3.10b)
$$N(d)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \left\{S_{N(d)}\rho\left(\frac{X_i - T_{N(d)}}{S_{N(d)}}\right) - \xi E\rho(X)\right\}$$

are asymptotically normally distributed, under the distribution function $F\left(\xi^{-1}(x-\theta)\right)$.

Proof of Proposition 3.1. Since ρ is increasing, simple manipulations show

$$|V_n(t,u)| \le \left|n^{-1} \sum_{i=1}^{n} \left\{\rho\left((1+ua_n)(X_i - a_n t)\right) - \rho\left((1+ua_n)(X_i - a_n\frac{[nt]}{n})\right)\right\}\right|$$

$$+ \left|E\left\{\rho\left((1+ua_n)(X - a_n t)\right) - \rho\left((1+ua_n)(X - a_n\frac{[nt]}{n})\right)\right\}\right| + A_{2n} + V_n([nt]/n,u)$$

$$\le \left|n^{-1} \sum_{i=1}^{n} \left\{\rho\left((1+ua_n)(X_i - a_n\frac{[nt]+1}{n})\right) - \rho\left((1+ua_n)(X_i - a_n\frac{[nt]}{n})\right)\right\}\right|$$

$$+ \left|E\left\{\rho\left((1+ua_n)(X - a_n\frac{[nt]+1}{n})\right) - \rho\left((1+ua_n)(X - a_n\frac{[nt]}{n})\right)\right\}\right| + A_{2n}$$

$$+ \left|n^{-1} \sum_{i=1}^{n} \left\{\rho\left((1+ua_n)(X_i - a_n\frac{[nt]}{n})\right) - \rho\left((1+a_n\frac{[nu]}{n})(X_i - a_n\frac{[nt]}{n})\right)\right\}\right|$$

$$+ A_{1n} + A_{3n}.$$

Thus,

$$\sup_{0\le t,u\le 1} |V_n(t,u)| \le 3 \sup_{0\le t,u\le 1} \left|n^{-1} \sum_{i=1}^{n} \left\{\rho\left((1+ua_n)(X_i - a_n\frac{[nt]}{n})\right)\right.\right.$$

$$\left.\left. - \rho\left((1+a_n\frac{[nu]}{n})(X_i - a_n\frac{[nt]}{n})\right)\right\}\right|$$

$$+ \sup_{0\le t,u\le 1} \left|n^{-1} \sum_{i=1}^{n} \left\{\rho\left((1+a_n\frac{[nu]}{n})(X_i - a_n\frac{[nt]+1}{n})\right)\right.\right.$$

(cont

(cont.) $$-\rho\Big((1+a_n\frac{[nu]}{n})(X_i-a_n\frac{[nt]}{n})\Big)\Big\}\Big| + 2A_{1n} + A_{6n} + A_{2n} + A_{1n} + A_{3n}$$

$$\leq M(A_{1n} + A_{2n} + A_{3n} + A_{4n} + A_{5n} + A_{6n}).$$

<u>Proof of Lemma 3.1.</u> Equations (2.6a) and (2.6b) together with (3.5) and (3.6) imply that $b_nA_{1n} \to 0$, $b_nA_{6n} \to 0$. $b_nA_{2n} \to 0$ (a.s.) by Lemma 2.1. The almost sure behavior of the other terms in Proposition 3.1 follows in a manner similar to the proof of Lemma 2.1 (although A_{4n} and A_{5n} are not mean zero random variables, one may normalize them easily).

4. First Weak Convergence Results

The asymptotic normality of the stopping rule discussed in Section 1 was shown in Corollary 3.4 under (essentially) the assumptions that $a_n^{-1}T_n \to 0$ (a.s.), $a_n^{-1}(S_n-1) \to 0$ (a.s.) (where $a_n^{-1} = (\log n)^2$), and that

$$N(d)^{\frac{1}{2}}\{N(d)^{-1}\sum_{i=1}^{N(d)}\big(\rho(X_i)-E\rho(X)\big)+\big(E\rho(X)-A(F)\big)(S_{N(d)}-1)-B(F)T_{N(d)}\}$$

is asymptotically normally distributed. The conditions on ρ were minimal and no differentiability properties were needed. In this and the next section, by means of the theory of weak convergence, the assumptions $a_n^{-1}T_n \to 0$ (a.s.) and $a_n^{-1}(S_n-1) \to 0$ (a.s.) are removed; however, the price paid is strengthened restrictions on ρ. Differentiability of ρ is still unnecessary.

In this section we investigate the estimators (2.1). Theorem 4.1, discusses the asymptotic normality of

$$\big(N(d)\big)^{-\frac{1}{2}}\sum_{i=1}^{N(d)}\{\rho(X_i-T_{N(d)}) - \int\rho(y-T_{N(d)})dF(y)\},$$

the only assumption for T_n being $T_n \to 0$ (a.s.). In Lemma 4.2, the asymptotic normality of the left hand side of (3.10) is discussed, the only assumptions for T_n relating to the asymptotic behavior of $N(d)^{\frac{1}{2}}T_{N(d)}$.

Definition 4.1. Let

$$V_n(s,t) = n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{\rho(X_i-t)-E\rho(X-t)\}$$

$$V_n^*(s,t) = n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{\rho(X_i-[nt]/n)-E\rho(X-[nt]/n)\}.$$

That V_n and V_n^* are essentially the same follows from the next result.

Proposition 4.1. Suppose that ρ is increasing and satisfies as $h \to 0$,

$$\int \{\rho(x+q+h) - \rho(x+q)\}^2 dF(x) = O(|h|^2), \tag{4.1}$$

uniformly in $|q| \le 1$.

Then

$$\sup\{|V_n(s,t)-V_n^*(s,t)|: 0\le s,t\le 1\} \xrightarrow{P} 0.$$

Note that (4.1) is stronger than (2.5). It is clear from the proof that the supremum could be taken for values of t ranging over any finite interval. Indeed, this will be true of all the results.

Definition 4.2. Define

$$\Gamma(t_1,t_2) = \text{Cov}\big(\rho(X-t_1),\rho(X-t_2)\big).$$

Lemma 4.1. Assume the conditions of Proposition 4.1 hold, that $\Gamma(t_1,t_2)$ is continuous, and that there exists a constant $M>0$ with

(4.2) $$\int \{\rho(y-t)-\rho(y-s)-E(\rho(X-t)-\rho(X-s))\}^4 \le M|t-s|$$

uniformly in $|t|,|s| \le 1$.

Then there is a D_2-valued process W such that

$$V_n \xrightarrow{W} W,$$

where "$\xrightarrow{W}$" denotes weak convergence.

Note that Theorem 4.1 below assumes nothing about the asymptotic distribution of $N(d)^{\frac{1}{2}}T_{N(d)}$, but rather it assumes that T_n is strongly consistent.

<u>Theorem 4.1.</u> Suppose that the conditions of Lemma 4.1 hold and

(4.3a) $$T_n \to 0 \text{ (a.s.) under } F(x)$$

(4.3b) $N(d)$ is a sequence of integer valued random variables such that for some sequence $\{n(d)\}$, $N(d)/n(d) \xrightarrow{p} 1$ as $d \to 0$.

Then

(4.3c) $$\left(N(d)\right)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \{\rho(X_i-T_{N(d)}) - \int \rho(x-T_{N(d)})dF(x)\} \xrightarrow{W} N\left(0,A^*(F,\rho)\right),$$

where $N(\mu,\sigma^2)$ is the distribution of a normal random variable with mean 0 and variance σ^2 and

$$A^*(F,\rho) = \mathrm{Var}_F\left(\rho(X)\right).$$

<u>Corollary 4.1.</u> Suppose that in addition to the conditions of Theorem 4.1

(4.4) $$E_F\{\rho(X+h)-\rho(X)\} = hB(F,\rho) + O(|h|^2) \text{ as } |h| \to 0 \text{, and that}$$

(4.5) $$\left(N(d)\right)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \{\rho(X_i)-E\rho(X)\} + N(d)^{\frac{1}{2}}B(F,\rho)T_{N(d)}$$

has a limiting normal distribution with mean zero and variance $C(F,\rho)$.

Then,

(4.6) $$\left(N(d)\right)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \{\rho(X_i-T_{N(d)})-E_F\rho(X)\} \xrightarrow{W} N\left(0,C(F,\rho)\right).$$

The main result of this section so far, namely Theorem 4.1, is based only on the assumption that T_n is a strongly consistent estimate of the location parameter. However, to get a result like (4.6), one must assume essentially that $\left(N(d)\right)^{\frac{1}{2}}T_{N(d)}$ has a limiting distribution. If only this assumption is made (rather than the strong consistency of T_n), the conditions on ρ can be relaxed. Define now for b_n monotonically nondecreasing

$$W_n(s,t) = n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{\rho(X_i-b_n t/n^{\frac{1}{2}})-E\rho(X-b_n t/n^{\frac{1}{2}})\}.$$

<u>Lemma 4.2.</u> Suppose ρ is increasing, that $N(d)/n(d) \xrightarrow{P} \theta$ (a positive random variable) and that the following hold:

(4.7a) There is a sequence c_n with $b_n c_n/n^{\frac{1}{2}} \to 0$ and $b_n/c_n \to 0$.

(4.7b) $$\int \{\rho(x+h)-\rho(x)\}^2 dF(x) = (|h|) \quad \text{as} \quad |h| \to 0$$

(4.7c) $$\left| \int \{\rho(x+h)-\rho(x)\} dF(x) \right| = O(|h|) \quad \text{as} \quad |h| \to 0.$$

Then

(4.8) $$\sup_{0\le s,t\le 1} |W_n(s,t)-W_n(s,o)| \xrightarrow{P} 0,$$

so that if $\left(N(d)\right)^{\frac{1}{2}}T_{N(d)}$ has a limiting distribution, (4.3c) holds. If (4.4) and (4.5) hold, then (4.6) is true.

Remark 4.1. Two points are of interest here. First, note that the convergence properties of $N(d)/n(d)$ have been relaxed. Secondly, it is easy to see that (4.7b) and (4.7c) if ρ has a bounded first derivative or is twice boundedly differentiable except at a finite number of points, and F is Lipschitz in neighborhoods of these points.

Proof of Proposition 4.1. The method of proof here follows along the lines of Bickel (1975). First,

$$E|V_n(s,t)-V_n^*(s,t)| \le \int \{\rho(x-t)-\rho(x-[nt]/n)\ ^2 dF(x) = O(n^{-1}).$$

Now define $P_{jn} = [n^{\frac{1}{4}}j]/n$, $j = 1,\ldots, n^* = [n^{3/4}]$. Then, uniformly in s,

$$E \max_{1\le j\le n^*} |V_n(s,P_{jn})-V_n^*(s,P_{jn})|^2 \le \sum_{i=1}^{n^*} E|V_n(s,P_{jn})-V_n^*(s,P_{jn})|^2 = O(n^{-\frac{1}{4}}).$$

Since ρ is increasing,

$$\begin{aligned}\sup\{|V_n(s,t)-V_n(s,P_{jn})|: P_{jn}\le t\le P_{j+1,n}\} &\le |V_n(s,P_{jn})-V_n(s,P_{j+1,n})| \\ &\quad + n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} |E\{\rho(x-P_{jn})-\rho(x-P_{j+1,n})\}| \\ &\le |V_n(s,P_{jn})-V_n(s,P_{j+1,n})| + O(n^{-\frac{1}{4}}),\end{aligned}$$

the last inequality following by (4.1). A similar computation may be made for V_n^*. Thus,

$$\begin{aligned}\sup_{0\le t\le 1} |V_n(s,t)-V_n^*(s,t)| &\le \max_{0\le j\le n^*} |V_n(s,P_{jn})-V_n^*(s,P_{jn})| \\ &\quad + \max_{0\le j\le n^*} |V_n(s,P_{jn})-V_n(s,P_{j+1,n})| \\ &\quad + \max_{0\le j\le n^*} |V_n^*(s,P_{jn})-V_n^*(s,P_{j+1,n})| + O(n^{-\frac{1}{4}}).\end{aligned}$$

Since the variances of each of the terms in absolute values are $O(n^{-3/2})$, application of Kolmogorov's inequality completes the proof.

Proof of Lemma 4.1. Let $Z_n(s,t) = V_n(s,o)$ and $Z_n^*(s,t) = V_n^*(s,t)-V_n(s,o)$. Z_n converges weakly to an element w_1 in D_2. If Z_n^* also converges weakly to an element w_2 in D_2 the proof would be complete. Make the following definitions. Disjoint blocks B and C in $\mathbb{R}^2$ are neighbors if they abut and have one face in common. For any D_2 valued process X and block $B = (s_1,t_1]\times(s_2,t_2]$, we define

$$X(B) = X(s_1,s_2)-X(s_1,t_2)-X(t_1,s_2)+X(t_1,t_2).$$

Because of Theorem 6 of Bickel and Wichura (1971), since the finite dimensional distributions converge and Z_n^* vanishes along its lower boundary, it remains to show that Z_n^* is tight. In order to prove tightness, it suffices to show that if B and C are neighbors, there exists $\gamma>0$, $\beta>\frac{1}{2}$ such that

$$E|V_n^*(B)|^{\gamma}|V_n^*(C)|^{\gamma} \le \big(\mu(B)\mu(C)\big)^{\beta}, \tag{4.9}$$

where μ is a finite non-negative measure on the unit cube. Letting j,k,m,p,q,r be integers with $0\le j\le k\le m\le n$, $0\le p\le q\le r\le n$, there are two cases to deal with

$$\begin{aligned} B &= (\tfrac{j}{n},\tfrac{k}{n}] \times (\tfrac{p}{n},\tfrac{q}{n}] \\ C &= (\tfrac{k}{n},\tfrac{m}{n}] \times (\tfrac{p}{n},\tfrac{q}{n}] \end{aligned} \tag{4.10}$$

$$\begin{aligned} B &= (\tfrac{j}{n},\tfrac{k}{n}] \times (\tfrac{p}{n},\tfrac{q}{n}] \\ C &= (\tfrac{j}{n},\tfrac{k}{n}] \times (\tfrac{q}{n},\tfrac{r}{n}]. \end{aligned} \tag{4.11}$$

In the first case, under equation (4.4) $V_n^*(B)$ and $V_n^*(C)$ are independent so that there is a constant $M>0$ with

$$E|V_n^*(B)|^2|V_n^*(C)|^2 \le M\left(\frac{k-j}{n}\right)\left(\frac{m-k}{n}\right)\left(\frac{p-q}{n}\right)^2,$$

so that $\beta=1$ suffices in (4.9) with μ being Lebesgue measure on the cube. Under equation (4.11), $V_n^*(B)$ and $V_n^*(C)$ are not independent, but the Schwartz inequality may be employed. Then, letting $Z(p,q) = \rho(X-\frac{q}{n})-\rho(X-\frac{p}{n})-E\{\rho(X-\frac{q}{n})-\rho(X-\frac{p}{n})\}$,

$$\begin{aligned} E|V_n^*(B)|^4 &\le \left(\frac{k-j}{n^2}\right)E|Z(p,q)|^4 + \left(\frac{k-j}{n}\right)^2\left(E|Z(p,q)|^2\right)^2 \\ &\le M\left(\frac{k-j}{n}\right)^2\left(\frac{p-q}{n}\right)^2, \end{aligned}$$

with the last inequality following from (4.2) and the fact that $(k-j)/n\le 1/n$, so that

$$E|V_n^*(B)|^2|V_n^*(C)|^2 \le M_1\left|\frac{k-j}{n}\right|^2\left|\frac{p-q}{n}\right|\left|\frac{r-q}{n}\right|.$$

Proof of Theorem 4.1. Because the result is true for V_n (Billingsley (1968), Section 17), it will suffice to show that for all $\varepsilon,\beta > 0$, there exists η, n_0 such that if $n \ge n_0$,

$$\Pr\left\{\sup_{\substack{0\le t\le\eta \\ 0\le s\le 1}} |Z_n^*(s,t)| > \beta\right\} < \varepsilon.$$

Now, since $Z_n^*(s,o) = 0$,

$$(4.12)\qquad |Z_n^*(s,t)| \le \min\{|Z_n^*(s,t)-Z_n^*(s,o)|,\ |Z_n^*(s,\eta)-Z_n^*(s,t)|\} + |Z_n^*(s,\eta)|.$$

Thus

$$(4.13)\quad \sup_{\substack{0\le t\le \eta\\ 0\le s\le 1}} |Z_n^*(s,t)| \le \sup_{\substack{t\le u\le v\\ v-t\le\eta}} \min\left\{ \sup_{0\le s\le 1} |Z_n^*(s,u)-Z_n^*(s,t)|,\ \sup_{0\le s\le 1} |Z_n^*(s,v)-Z_n^*(s,u)| \right\} + \sup_{0\le s\le 1} |Z_n^*(s,\eta)|.$$

By Kolmogorov's inequality, the second term on the right hand side of (4.13) converges in probability to zero as $\eta \to 0$. The first term is bounded by the modulus $\omega_\eta''(Z_n^*)$ (see Bickel and Wichura (1971)) and we proved tightness in Lemma 4.1 by showing that

$$\lim_{\eta\to 0}\lim_{n\to\infty} \Pr\{\omega_\eta''(Z_n^*)>\varepsilon\} = 0.$$

Thus, since $T_n \to 0$ (a.s.), $\sup_{0\le s\le 1} |Z_{n(d)}^*(s,T_{N(d)})| \xrightarrow{P} 0$ as $d\to 0$.

Proof of Corollary 4.1. The term in (4.6) can be written as

$$\left(N(d)\right)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \{\rho(X_i-T_{N(d)}) - \int \rho(y-T_{N(d)})dF(y)\} + N(d)^{\frac{1}{2}}B(F,\rho)T_{N(d)} + O_p\left(N(d)^{-1}\right).$$

where "O_p" means bounded in probability. From the proof of Theorem 4.1, this term has the same limiting distribution as

$$\left(N(d)\right)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \{\rho(X_i)-E\rho(X)\} + N(d)^{\frac{1}{2}}B(F,\rho)T_{N(d)},$$

which completes the proof.

Proof of Lemma 4.2. The proof of (4.8) follows closely that of Proposition 4.1 with $P_{jn} = j/c_n$, $j = 0,\ldots, n^* = c_n$. Then, one sees that if $N(d)/m(d) \xrightarrow{P} c_0$ for some constant c_0, that

$$\sup_{0\le s\le 1} |W_{m(d)}(s, m(d)^{\frac{1}{2}}T_{N(d)}/b_{m(d)}) - W_{m(d)}(s,o)| \xrightarrow{P} 0.$$

Since $W_n(s,o)$ converges to a Wiener process, Billingsley's Theorem 17.2 applies.

5. Weak Convergence Results

In this section we discuss the asymptotic normality of (3.1) and (3.1)* under random sample sizes. As mentioned in Section 4, the assumptions about T_n, S_n will only include knowledge of the asymptotic behavior of $N(d)^{\frac{1}{2}}T_{N(d)}$ and $N(d)^{\frac{1}{2}}(S_{N(d)}-1)$. First consider (3.1).

Definition 5.1. For a sequence of constants $\{a_n\}$ decreasing to zero

$$V_n(s,t,u) = n^{-\frac{1}{2}} \sum_{i=1}^{[ns]} \{\rho\big((1+ua_n)(X_i-t/n^{\frac{1}{2}})\big) - E\rho\big((1+ua_n)(X-t/n^{\frac{1}{2}})\big)\}.$$

Lemma 5.1. Let $\{A_n\}$, $\{B_n\}$, $\{C_n\}$ each converge to zero. Suppose ρ is increasing and

(5.1) $$|E\{\rho\big((1+A_n)(X-B_n-C_n)\big) - \rho\big((1+A_n)(X-B_n)\big)\}| = O(|C_n|)$$

(5.2) If $H_n = E\{\rho\big((1+hA_n)(X-B_n)\big)-\rho\big((1+qA_n)(X-B_n)\big)\}$, then for $r = 1,2$ uniformly in $|h|$, $|q| \le 1$, $E\{\rho\big((1+hA_n)(X-B_n)\big)-\rho\big((1+qA_n)(X-B_n)\big)-H_n\}^{2r} \le M|h-q|^r$ for some $M\ge 0$.

Then for any $M_0>0$

$$\sup\{|V_n(s,t,u)-V_n(s,o,o)|: 0\le s,u\le 1,\ 0\le t\le M_0\} \xrightarrow{P} 0.$$

Theorem 5.1. Suppose $\rho = \rho^+-\rho^-$, where ρ^+, ρ^- satisfy the conditions of Lemma 5.1. If, in addition $N(d)/n(d) \xrightarrow{P} \theta$ (a positive random variable), $a_n^{-1}(S_n-\xi) \to 0$ (a.s.), and

$n(d)^{\frac{1}{2}}T_{N(d)}$ has a limiting distribution, then

$$(5.3)\quad \big(N(d)\big)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \left\{\rho\left(\frac{X_i-T_{N(d)}}{S_{N(d)}}\right) - \int \rho\left(\frac{y-T_{N(d)}}{S_{N(d)}}\right)dF(y)\right\} \overset{w}{\Rightarrow} N\Big(0,\ \mathrm{Var}\big(\rho(X)\big)\Big).$$

The same result is true if the almost sure behavior of S_n is unknown but $\big(N(d)\big)^{\frac{1}{2}}(S_{N(d)}-1)$ has a limiting distribution.

Corollary 5.1. Let (3.8) and the conclusion to Theorem 5.1 hold. If, in addition, T_n and S_n-1 are uniformly continuous in probability and are jointly asymptotically normally distributed with $n^{-1}\sum_{i=1}^{n}\{\rho(X_i)-E\rho(X)\}$, then both

$$N(d)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \left\{\rho\left(\frac{X_i-T_{N(d)}}{S_{N(d)}}\right) - E\rho(X)\right\}$$

and

$$N(d)^{-\frac{1}{2}} \sum_{i=1}^{N(d)} \left\{S_{N(d)}\rho\left(\frac{X_i-T_{N(d)}}{S_{N(d)}}\right) - \xi E\rho(X)\right\}$$

are asymptotically normally distributed under $F\big(\xi^{-1}(x-\theta)\big)$.

Proof of Lemma 5.1. We may assume that $0\le t\le 1$. Then

$$(5.4)\quad |V_n(s,t,u)-V_n(s,o,o)| \le |V_n(s,t,u)-V_n(s,t,o)| + |V_n(s,t,o)-V_n(s,o,o)|.$$

The second term on the right hand side of (5.4) has been handled in Section 4 with $b_n \equiv 1$. Let δ be fixed but small and define $P_j = j/m$, $j = 0,1,\ldots,$ $m = [1/\delta]$. Then, if the following suprema are taken over the set $\{0\le s,u\le 1,\ P_j\le t\le P_{j+1}\}$, we obtain

$$\begin{aligned}\sup|V_n(s,t,u)-V_n(s,t,o)| &\le \sup\{|V_n(s,t,u)-V_n(s,P_j,u)|+|V_n(s,P_j,u)-V_n(s,P_j,o)| \\ &\quad + |V_n(s,P_j,o)-V_n(s,t,o)| \\ &\le 2\sup|V_n(s,t,u)-V_n(s,P_j,u)|+\sup|V_n(s,P_j,u)-V_n(s,P_j,o)|\end{aligned}$$

By the monotonicity of ρ, this last term is bounded above by

$$2\,\sup|V_n(s,P_{j+1},u)-V_n(s,P_j,u)| + \sup|V_n(s,P_j,u)-V_n(s,P_j,o)| + 2n^{\frac{1}{2}}|E\{\rho((1+ua_n)(X-t/n^{\frac{1}{2}})) - \rho((1+ua_n)(X-P_j/n^{\frac{1}{2}}))\}|.$$

Hence, if the following suprema are taken over the set $\{0\le s,t,u\le 1\}$, then for some $M>0$,

$$(5.5)\quad \frac{1}{M}\sup|V_n(s,t,u)-V_n(s,t,o)| \le \sum_{j=0}^{m}\sup|V_n(s,P_j,u)-V_n(s,P_j,o)| + 2\sum_{j=0}^{m}\sup|V_n(s,P_j,o)-V_n(s,o,o)| + \mathcal{O}(\delta),$$

the $\mathcal{O}(\delta)$ term following by (5.1). Hence it suffices to show that for any fixed t_o,

$$(5.6)\qquad \sup_{0\le s,u\le 1}|V_n(s,t_o,u)-V_n(s,t_o,o)| \xrightarrow{P} 0.$$

The proof of (5.6) parallels that of Lemma 4.1, with the condition (5.2) being used.

Proof of Theorem 5.1. The expression on the left hand side of (5.3) is

$$(5.7)\qquad V_{n(d)}\big(N(d)/n(d),\ a_{n(d)}^{-1}(S_{N(d)}^{-1}-1),\ n(d)^{\frac{1}{2}}T_{N(d)}\big).$$

Since $|n(d)^{\frac{1}{2}}T_{N(d)}|$ is bounded by some M_o with arbitrarily high probability, Lemma 5.1 shows that (5.7) is equivalent in probability to

$$V_{n(d)}\big(N(d)/n(d),o,o\big),$$

completing the proof.

Remark 5.1. While it is possible to investigate the weak convergence of the process $W_n(s,t,u)$ given by $W_n(s,t,u) = V_n(s,n^{\frac{1}{2}}t,u)$, we have been unable to obtain results which use only global properties of ρ such as (5.1) and (5.2). Rather, our results require such local properties as differentiability.

6. Applications

We now present two applications of the results in the previous sections. It is first shown that stopping rules for fixed-width confidence intervals based on M-estimators (Huber (1964), Andrews, et al (1972)) are asymptotically normally distributed. The second application is to one step M-estimators (Bickel (1975)); almost sure representations are given for these estimators, their asymptotic normality under random sample sizes is discussed, and estimates of their variance lead to stopping rules which are asymptotically normal.

M-estimators are defined as solutions to the following equation:

$$0 = \sum_{i=1}^{n} \psi\left(\frac{X_i - T_n}{S_n}\right),$$

where S_n is a robust estimate of scale with the invariance properties discussed in Section 3. Assume $E\psi(X) = 0$ and ψ, ψ' are bounded. Then $n^{\frac{1}{2}}T_n$ is asymptotically normal with mean zero and variance

$$\int \psi^2(x)dF(x)/\{\int \psi'(x)dF(x)\}^2. \tag{6.1}$$

Hence, the natural estimator of (6.1) is

$$n^{-1} S_n \sum_{i=1}^{n} \psi^2\left(\frac{X_i - T_n}{S_n}\right) \Big/ \left\{n^{-1} \sum_{i=1}^{n} \psi'\left(\frac{X_i - T_n}{S_n}\right)\right\}^2. \tag{6.2}$$

The following Lemmas are immediate consequences of the work in Sections 3 and 5.

Lemma 6.1. Suppose ψ^2, ψ' satisfy the conclusion to Corollary 3.1 with $a_n = n^{-\frac{1}{2}} \log n$, that (3.8) holds, and $E\psi'(X) \neq 0$. Define

$$a_1 = \{E\psi'(X)\}^{-2}$$

$$a_2 = -2E\psi^2(X)/\{E\psi'(X)\}^3$$

$$a_3 = a_1 A_{\psi^2}(F) + a_2 A_{\psi'}(F) + a_1 E\psi^2(X)$$

$$a_4 = -\left(a_1 B_{\psi^2}(F) + a_2 B_{\psi'}(F)\right)$$

Then

$$(6.3)\quad \frac{S_n n^{-1} \sum_{i=1}^{n} \psi^2\left(\frac{X_i - T_n}{S_n}\right)}{\left\{n^{-1} \sum_{i=1}^{n} \psi'\left(\frac{X_i - T_n}{S_n}\right)\right\}^2} - \frac{E\psi^2(X)}{\{E\psi'(X)\}^2}$$

$$= n^{-1} \sum_{i=1}^{n} \{a_1\left(\psi^2(X_i) - E\psi^2(X)\right) + a_2\left(\psi'(X_i) - E\psi'(X)\right)\}$$

$$+ a_3(S_n^{-1} - 1) + a_4 T_n + \mathcal{O}\left(n^{-3/4}(\log n)^2\right) \quad \text{(a.s.)}.$$

Lemma 6.2. Suppose ψ^2, ψ' satisfy the conclusion to either Lemma 6.1 or Lemma 5.1, that $N(d)^{\frac{1}{2}}(S_{N(d)} - 1)$ has a limiting distribution, that (3.8) holds, and that for some constant $D(F,\psi)$,

$$(6.4)\quad N(d)^{\frac{1}{2}}\left\{N(d)^{-1} \sum_{i=1}^{N(d)} \{a_1\left(\psi^2(X_i) - E\psi^2(X)\right) + a_2\left(\psi(X_i) - E\psi(X)\right)\}\right\}$$

$$+ a_3(S_{N(d)}^{-1} - 1) + a_4 T_{N(d)}$$

$$\xrightarrow{L} N\left(0, D(F,\psi)\right).$$

Then

$$(6.5)\quad N(d)^{\frac{1}{2}}\left\{\frac{S_{N(d)} N(d)^{-1} \sum_{i=1}^{N(d)} \psi^2\left(\frac{X_i - T_{N(d)}}{S_{N(d)}}\right)}{\left\{N(d)^{-1} \sum_{i=1}^{N(d)} \psi'\left(\frac{X_i - T_{N(d)}}{S_{N(d)}}\right)\right\}^2} - \frac{E\ ^2(X)}{\left(E\psi'(X)\right)^2}\right\} \xrightarrow{L} N\left(0, D(F,\psi)\right).$$

Remark 6.1. The results (6.3) and (6.5) hold under very general conditions. One important set obtains (6.5) from (6.3). Carroll (1975) has shown that under the conditions of Corollary 3.2,

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

Bahadur (1966) has shown that for some function H, if S_n is the inter-quartile range (suitably normalized),

$$S_n-1 = n^{-1} \sum_{i=1}^{n} \{H(X_i)-EH(X)\} + O\left(n^{-3/4}(\log n)^2\right) \quad \text{(a.s.)}.$$

Under these conditions, (6.4) is immediate from Anscombe (1952), so that (6.3) and (6.5) hold.

The second appliation has to do with one step estimators. If $\hat{\beta}_n$ is a preliminary estimate of location (such as the sample median) and S_n the robust estimate of scale, we define

$$\text{(6.6)} \qquad T_n = \hat{\beta}_n + \frac{S_n n^{-1} \sum_{i=1}^{n} \psi\left(\frac{X_i-\hat{\beta}_n}{S_n}\right)}{n^{-1} \sum_{i=1}^{n} \psi'\left(\frac{X_i-\hat{\beta}_n}{S_n}\right)} .$$

Then the following Lemmas are also immediate from the results of Sections 3 and 5.

Lemma 6.3. Suppose that ψ, ψ' satisfy the conclusion to Corollary 3.1. Suppose also that $E_F\psi(X) = 0$ and

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

Then

$$\text{(6.7)} \quad T_n = \left(E\psi'(X)\right)^{-1}\{n^{-1} \sum_{i=1}^{n} \psi(X_i)-A_\psi(F)(S_n-1)+\hat{\beta}_n\left(E\psi'(X)-B_\psi(F)\right)\} + O\left(n^{-3/4}(\log n)^2\right) \quad \text{(a.s.)}.$$

If, in addition, $\psi(x) = \psi(-x)$ and the conclusion of Corollary 3.2 holds, then

$$T_n = \frac{n^{-1} \sum_{i=1}^{n} \psi(X_i)}{E\psi'(X)} + O\left(n^{-1}(\log n)^2\right) \quad \text{(a.s.)}. \tag{6.8}$$

Proof of Lemma 6.3. We have

$$n^{-1} \sum_{i=1}^{n} \psi\left(\frac{X_i - \hat{\beta}_n}{S_n}\right) = n^{-1} \sum_{i=1}^{n} \psi(X_i) + A_\psi(F)(S_n - 1) - B_\psi(F)\hat{\beta}_n + O\left(n^{-1}(\log n)^2\right) \quad \text{(a.s.)}$$

and

$$n^{-1} \sum_{i=1}^{n} \psi'\left(\frac{X_i - \hat{\beta}_n}{S_n}\right) = n^{-1} \sum_{i=1}^{n} \psi'(X_i) + O\left(n^{-\frac{1}{2}}(\log n)\right) \quad \text{(a.s.)}$$

$$= E\psi'(X) + O\left(n^{-\frac{1}{2}}(\log n)\right) \quad \text{(a.s.)}.$$

The result now follows since $n^{-1} \sum_{i=1}^{n} \psi(X_i) = O\left(n^{-\frac{1}{2}}(\log n)\right)$ (a.s.) (remember, $E\psi(X) = 0$).

Lemma 6.4. Suppose ψ, ψ' satisfy the conclusion to Lemma 5.1, that (3.8) holds, and that both $N(d)^{\frac{1}{2}}\hat{\beta}_{N(d)}$ and $N(d)^{\frac{1}{2}}(S_{N(d)}-1)$ have limiting distributions. Then $N(d)^{\frac{1}{2}}T_{N(d)}$ has the same limit distribution as

$$\frac{N(d)^{\frac{1}{2}}\left\{N(d)^{-1} \sum_{i=1}^{N(d)} \psi(X_i) - A_\psi(F)(S_{N(d)}-1) - B_\psi(F)\hat{\beta}_{N(d)}\right.}{E\psi'(X)} . \tag{6.9}$$

Since the median satisfies the Bahadur representation, (6.8) and (6.9) hold under a set of conditions similar to those of Remark 6.1. Note that the one-steps have the asymptotic variance given by (6.1), so that stopping times based on (6.2) are also asymptotically normal when T_n is a one-step.

It should be noted that results similar to those given here can be obtained by embedding the process

$$Z_n^*(t,u) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{\rho\big((1+u)(X_i-t)\big) - E\big((1+u)(X-t)\big)\right\}$$

in a Brownian motion in a manner similar to that of Bickel and Rosenblatt (1973). This approach requires from the outset that F be continuous; in contrast, the results given here make virtually no assumptions about F if ρ is continuous, while discontinuities in ρ are handled by making F behave nicely in neighborhoods of the discontinuities. To be fair, the embedding approach obtains (3.7) under the assumption that T_n, S_n are almost surely convergent , but in working with M-estimators one generally can find rate results once strong consistency is assured. Thus, the methods of this paper are not only of interest in themselves but also yield results which compare quite favorably with those obtainable from embedding.

ACKNOWLEDGEMENT: I wish to thank Malay Ghosh for sending a preprint of his 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*. Cambridge: 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] BICKEL, P.J. (1975). One step Huber estimates in the linear model. *J. Amer. Statist. Assoc. 70*, 428-434.

[5] BICKEL, P.J. and WICHURA, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist. 12*, 1656-1670.

[6] BICKEL, P.J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist. 1*, 1071-1095.

[7] BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. New York: John Wiley & Sons, Inc.

[8] CARROLL, R.J. (1975). An asymptotic representation for M-estimators and linear functions of order statistics. *Institute of Statistics Mimeo Series #1006*, University of North Carolina at Chapel Hill.

[9] 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.

[10] GHOSH, M. and MUKHOPHADYAY, H. (1975). Asymptotic normality of stopping times in sequential analysis. Unpublished paper.

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

[12] LOEVE, M. (1968). *Probability Theory, 3rd ed.*, Princeton: Van Nostrand.