The Annals of Statistics
1978, Vol. 6, No. 2, 314–318

ON ALMOST SURE EXPANSIONS FOR M-ESTIMATES

By Raymond J. Carroll

University of North Carolina at Chapel Hill

Let T_n be an M-estimator with defining function ψ and preliminary estimate of scale s_n. Without loss of generality, let $s_n \to 1$ and take $E\psi(X/\xi) = 0$. Under various conditions, it is shown that any consistent version of T_n is almost surely to order $O(n^{-1}\log_2 n)$ a linear combination of $n^{-1}\sum_1^n \psi(X_i)$ and s_n. Only in the case $EX_1\psi'(X_1) = 0$ does the contribution of s_n vanish; it is shown how this affects the estimation of the asymptotic variance of T_n.

1. Introduction. The object of this note is to obtain an almost sure invariance principle for M-estimators (Huber (1964)). Specifically, let $X_1, X_2, \cdots$ be i.i.d. with common distribution function $F(x)$, and let s_n be a robust estimate of scale which is location invariant and converges almost surely to a constant ξ. For a given bounded function ψ, define $\theta = \theta(F)$ as a suitable solution to

$$0 = E_F\psi(\xi^{-1}(X_1 - t)), \tag{1.1}$$

where E_F denotes expectation under F. We propose to estimate θ by a suitable solution T_n to

$$0 = \textstyle\sum_1^n \psi(s_n^{-1}(X_i - t)). \tag{1.2}$$

Examples of s_n include the sample interquartile range (which Bahadur (1966) shows satisfies the law of the iterated logarithm (LIL)) and the trivial case $s_n \equiv 1$. In Theorem 1, general conditions are found for the existence of a constant C with

$$\limsup_n n(\log_2 n)^{-1}|\{E_F\psi'(\xi^{-1}(X_1 - \theta))\}(T_n - \theta)/\xi - (H_n + G_n)| \leq C \quad \text{(a.s.)}, \tag{1.3}$$

where

$$H_n = n^{-1}\textstyle\sum_1^n \psi(\xi^{-1}(X_i - \theta))$$
$$G_n = (1 - s_n/\xi)E_F(\xi^{-1}(X_1 - \theta))\psi'(\xi^{-1}(X_1 - \theta)).$$

Phrased loosely, (1.3) shows that, except for a remainder term of almost sure order $O(n^{-1}\log_2 n)$, $T_n - \theta$ is a linear combination of s_n and the average of n bounded random variables. Some consequences of (1.3) are sketched in Section 3.

2. Main results. Throughout this section, unless otherwise stated, s_n is taken to be location invariant and scale equivariant, so that we may set $\theta = 0$, $\xi = 1$ without loss of generality. Consider the following assumptions:

Received June 1975; revised May 1977.

[1] This research was supported in part by the Air Force Office of Scientific Research under Grants AFOSR-75-2840 and AFOSR-75-2796.

AMS 1970 *subject classifications*. Primary 62G35; Secondary 62E20.

Key words and phrases. M-estimators, robustness, Studentization, invariance principles.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

(A1) ψ is bounded and there are intervals $(a_0 = -\infty, a_1), \cdots, (a_k, a_{k+1} = +\infty)$ on which ψ has two continuous bounded derivatives.

(A2) F satisfies a Lipschitz condition of order one in neighborhoods of $\{a_1, \cdots, a_k\}$ and $E_F\psi'(X_1) \neq 0$.

(A3) We have $\sup x^2|\psi''(x)| < \infty$ and $E_F X_1^2(\psi'(X_1))^2 < \infty$.

(A4) The sequence of statistics $u_n = s_n^{-1}$ satisfies

$$\limsup_n n^{\frac{1}{2}}(\log_2 n)^{-\frac{1}{2}}|u_n - 1| < C_u \quad \text{(a.s.)}, \quad \text{for some constant} \quad C_u .$$

In the sequel we take $C_u = 1$ without loss of generality.

DEFINITION. If for almost all ω a solution T_n of (1.2) exists for $n \geqq n(\omega)$ such that $T_n \to \theta$ almost surely under F, $\{T_n\}$ is called a strongly consistent sequence of solutions.

THEOREM 1. *Suppose* (A1)—(A4) *hold and* $\{T_n\}$ *is a strongly consistent sequence of solutions. Then* (1.3) *holds.*

Before proving Theorem 1, two preliminary results are necessary.

LEMMA 1. *For* $n = 1, 2, \cdots$, *let* $\{X_{kn}, k = 1, 2, \cdots, n\}$ *be independent Bernoulli random variables with success probabilities* $\{p_{kn}\}$. *Define* $S_n = \sum_1^n X_{kn}$, $\mu_n = \sum_1^n p_{kn}$ *and* $\alpha_n = \max(\mu_n, \log n) = (\mu_n \vee \log n)$. *Then*

$$\limsup_n S_n/\alpha_n \leqq 4 \quad \text{(a.s.)} .$$

PROOF. By Bernstein's inequality,

$$\Pr\{S_n - \mu_n \geqq 3\alpha_n\} \leqq \exp\{-9\alpha_n/4\} \leqq n^{-\frac{9}{4}} ,$$

so the result follows from the Borel–Cantelli lemma.

LEMMA 2. *Suppose* (A1)—(A4) *holds and define* $\gamma_n = n^{-\frac{1}{2}}(\log_2 n)^{\frac{1}{2}}$. *Then there exist positive numbers C and* ε *such that for any sequence* $\varepsilon_1, \varepsilon_2, \cdots$ *in* $[0, \varepsilon]$ *satisfying for n sufficiently large*

$$\gamma_n \max_j |a_j| < (1 - \gamma_n)\varepsilon_n , \tag{2.1}$$

the following holds under F: for almost all ω, *there exists* $n(\omega)$ *such that* $n \geqq n(\omega)$ *and* $|t| \leqq \varepsilon_n$ *imply*

$$|n^{-1}\textstyle\sum_1^n \psi(u_n(X_i - t)) - n^{-1}\sum_1^n \psi(X_i) - (u_n - 1)E_F X_1\psi'(X_1) + tE_F\psi'(X_1)| \tag{2.2}$$
$$\leqq C\left\{(|t| + \gamma_n)^2 + (\gamma_n + |t|)\left(\varepsilon_n \vee \frac{\log n}{n}\right)\right\} ,$$

$$|n^{-1}\textstyle\sum_1^n \psi(u_n(X_i - t)) + tE_F\psi'(X_1)| \leqq C\{|t|\varepsilon_n + \gamma_n\} . \tag{2.3}$$

PROOF. By (A4), choose $n(\omega)$ so that $n \geqq n(\omega)$ implies $|u_n - 1| \leqq \gamma_n$. Choose ε so small that F is Lipschitz on $\bigcup_{j=1}^k [a_j - 3\varepsilon, a_j + 3\varepsilon]$. Define $B_n = \bigcup_{j=1}^k [a_j - 2\varepsilon_n, a_j + 2\varepsilon_n]$, so that $\Pr\{X_1 \in B_n\} \leqq C_0\varepsilon_n$ for some constant C_0. Now if $x \in B_n$, then since ψ is Lipschitz and x is bounded, for $x \neq a_j$ $(j = 1, \cdots, k)$ and $n \geqq n(\omega)$,

$$|\psi(u_n(x - t)) - \psi(x) - (u_n - 1)x\psi'(x) + u_n t\psi'(x)| \leqq C_1 I_{B_n}(x)\{|u_n - 1| + |t|\}$$
$$\leqq C_1 I_{B_n}(x)\{\gamma_n + |t|\} ,$$

where $I_A(x)$ is the indicator function for the set A evaluated at x. If $x \notin B_n$, say $a_j + 2\varepsilon_n < x < a_{j+1} - 2\varepsilon_n$, then (2.1) guarantees that both $u_n x$ and $u_n(x - t)$ are in the interval (a_j, a_{j+1}) (here we consider $n \geqq n(\omega)$). On this interval, ψ'' exists and is continuous so that

$$|\psi(u_n(x - t)) - \psi(x) - (u_n - 1)x\psi'(x) + u_n t\psi'(x)| \leqq (|u_n - 1|x + u_n|t|)^2|\psi''(\eta_n(x))|/2 , \tag{2.4}$$

where $\eta_n(x)$ is between $u_n(x - t)$ and x. By (A3),

$$\sup_x \sup \{x^2\psi''(\alpha x + \beta) : \tfrac{1}{2} \leqq \alpha \leqq \tfrac{3}{2}, |\beta| \leqq \tfrac{1}{2}\}$$
$$\leqq 4 \sup_x \{(1 + |x| + x^2)|\psi''(x)|\} < \infty ,$$

so that for $n \geqq n(\omega)$, if $x \neq a_j$ $(j = 1, \cdots, k)$, there is a constant C_1 for which

$$|\psi(u_n(x - t)) - \psi(x) - (u_n - 1)x\psi'(x) + u_n t\psi'(x)|$$
$$\leqq C_1\{(\gamma_n + |t|)^2 + (\gamma_n + |t|)I_{B_n}(x)\} .$$

This means that for $n \geqq n(\omega)$, the left-hand side of (2.2) is bounded by

$$\begin{aligned} &C_2\{(\gamma_n + |t|)^2 + (\gamma_n + |t|)n^{-1} \textstyle\sum_1^n I_{B_n}(X_i)\} \\ &\quad + |t|n^{-1}|\textstyle\sum_1^n (\psi'(X_i) - E_F\psi'(X_1))| \\ &\quad + \gamma_n n^{-1}|\textstyle\sum_1^n (X_i\psi'(X_i) - E_F X_1\psi'(X_1))| . \end{aligned} \tag{2.5}$$

Now, the boundedness of ψ' and the second part of (A3) guarantee the LIL for the sequences

$$n^{-1} \textstyle\sum_1^n (\psi'(X_i) - E_F\psi'(X_1))$$
$$n^{-1} \textstyle\sum_1^n (X_i\psi'(X_i) - E_F X_1\psi'(X_1)) ,$$

and this, together with Lemma 1, show that for $n \geqq n'(\omega)$, the left-hand side of (2.2) is bounded by

$$C_3\left\{(\gamma_n + |t|)^2 + |t|\gamma_n + \gamma_n^2 + (\gamma_n + |t|)\left(\varepsilon_n \vee \frac{\log n}{n}\right)\right\} ,$$

completing the first part of the lemma. To prove (2.3), note that by (1.1) and (A3) $n^{-1} \sum_1^n \psi(X_i)$ satisfies the LIL.

Proof of Theorem 1. Take $u_n = s_n^{-1}$ and $\varepsilon_n = \varepsilon' = \min(\varepsilon, |E_F\psi'|(X_1)/2C|)$ for all n. It is clear that (2.1) holds. Then, applying (2.3), for almost all ω there exists $n(\omega)$ such that if $n \geqq n(\omega)$, $\sum_1^n \psi(u_n(X_i - t))$ cannot vanish in the intervals

$$[-\varepsilon', -C_4\gamma_n] , \qquad [C_4\gamma_n, \varepsilon'] ,$$

and it has opposite signs on these intervals. Here C_4 is a large positive constant such that for t in these intervals,

$$\begin{aligned} C\{|t|\varepsilon_n + \gamma_n\} &\leqq C_4\gamma_n(C/C_4) + |t|\,|E_F\psi'(X_1)|/2 \\ &\leqq |t|\{(C/C_4) + |E_F\psi'(X_1)|/2\} \\ &< |t|\,|E_F\psi_1'(X_1)| \,. \end{aligned}$$

Since $\{T_n\}$ is a strongly consistent sequence of solutions and $\theta = 0$, we must have $|T_n| \leqq C_4\gamma_n$ if $n \geqq n(\omega)$. Now take $\varepsilon_n = \min(\varepsilon, C_4\gamma_n)$. By choosing C_4 sufficiently large, (2.1) again holds. By choosing $t = T_n$ in (2.2) and making use of (1.2) we obtain (1.3) with

$$\begin{aligned} G_n^{(1)} &= (u_n - 1)E_F X_1\psi'(X_1) \\ &= G_n + (1 - s_n)^2 E_F X_1\psi'(X_1)/s_n \,. \end{aligned}$$

Application of (A4) completes the proof.

The following corollary establishes the invariance principle (1.3) for the estimates which are not required to be scale equivariant. The proof is the same as that of Theorem 1.

COROLLARY 1. *If* $s_n \equiv 1$ *and* (A1), (A2) *hold except* ψ' *is only required to be Lipschitz on each subinterval, then* (1.2) *holds.*

The proof of Theorem 1 makes it clear that a strongly consistent sequence of solutions to (1.2) always exists. Two examples where it is easy to identify $\theta(F)$ and the suitable sequence of solutions are:

(i) ψ is nondecreasing, $\lambda(t) = E_F\psi(\xi^{-1}(X_1 - t))$ has a unique zero at $t = 0$.

(ii) F is symmetric about and continuous at θ, and T_n is the solution to (1.2) closest to the sample median.

3. Further considerations. The result (1.3) is interesting for at least two reasons. First, taking $\theta = 0$, $\xi = 1$, if F is symmetric and ψ is skew-symmetric then $E_F X_1\psi'(X_1) = 0$, so that $G_n \equiv 0$; hence the effect of asymmetry is to add on a "biasing" term to the distribution of T_n. This effect is not clear from the influence curve (Andrews et al. (1972)). Second, it shows that the stochastic process

$$V_n(t) = n^{\frac{1}{2}}(T_{[nt]} - \theta)$$

is tight in $D[\frac{1}{2}, 1]$ if a corresponding process for s_n is tight (which is true under Bahadur's conditions); this means that, as in Chow and Robbins (1966), sequential confidence intervals for θ of fixed length can be constructed using M-estimates.

Theorem 1 can be generalized to admit the cases $X_1, X_2, \cdots$ neither independent nor identically distributed. The changes in proof basically require a weakening of Lemma 1 (Hoeffding (1963)), finding an appropriate sequence s_n (Sen (1972)), and checking that the LIL holds for appropriate sequences (Stout

(1974)). The basic smoothness conditions on ψ given in Theorem 1 remain unchanged for stationary, m-dependent sequences. For independent, not identically distributed sequences with distribution functions F_i, we require that ψ' vanish outside a finite interval and that $n^{-1} \sum_1^n F_i(x)$ have certain convergence properties. Similar but somewhat stronger conditions on ψ permit the sequence $X_1, X_2, \cdots$ to be stationary, ϕ-mixing.

A different problem is that of estimating the asymptotic variance $A(F, \psi)$ of $n^{\frac{1}{2}}T_n$. Huber (1970) and Gross (1976) suggest using

$$\sigma_n^2 = s_n^2 \int \psi^2(s_n^{-1}(x - T_n))\, dF_n(x)/\{\int \psi'(s_n^{-1}(x - T_n))\, dF_n(x)\}^2$$

when $s_n \equiv 1$ or when F is symmetric and ψ skew-symmetric; either circumstance results in $C_n = 0$. However, if s_n is random and F asymmetric, Theorem 1 shows that σ_n^2 will not be a consistent estimate of $A(F, \psi)$. Thus, consistent estimation of the variability $A(F, \psi)$ will either require transformations to symmetry or the use of a technique such as the jackknife.

Acknowledgment. The author expresses his thanks to the associate editor, whose useful comments were directly responsible for the present formulation of this paper.

REFERENCES

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

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

CHOW, Y. S. and ROBBINS, H. (1965). On the asymptotic theory of fixed width sequential confidence intervals for the mean. *Ann. Math. Statist.* **36** 457–462.

GHOSH, M. (1972). On the representation of linear functions of order statistics. *Sankyā Ser. A* **34** 349–356.

GROSS, A. M. (1976). Confidence interval robustness with long-tailed symmetric distributions. *J. Amer. Statist. Assoc.* **71** 409–416.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

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

HUBER, P. J. (1970). Studentizing robust estimates. *Nonparametric Techniques in Statistical Inference* (M. L. Puri, ed.). Cambridge Univ. Press.

SEN, P. K. (1972). On the Bahadur representation of sample quantiles for sequences of ϕ-mixing random variables. *J. Multivariate Analysis* **2** 77–95.

STOUT, W. F. (1974). *Almost Sure Convergence*. Academic Press, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514