PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 123, Number 2, February 1995

LAW OF THE ITERATED LOGARITHM AND INVARIANCE PRINCIPLE FOR M-ESTIMATORS

XUMING HE AND GANG WANG

(Communicated by Wei Y. Loh)

ABSTRACT. We prove the law of the iterated logarithm for a general class of M-estimators which covers in particular robust M-estimators and S-estimators of multivariate location-scatter. We also obtain an almost sure invariance principle (Bahadur-type representation) for these estimators.

1. INTRODUCTION

The law of the iterated logarithm (LIL) for a sum $S_n = \sum_{i=1}^n X_i$ of independent and identically distributed (i.i.d.) random variables dates back to Khintchine and Kolmogorov in the 1920s. It is important in that it characterizes the asymptotic behavior of S_n by its exact rate of convergence. Since then, there has been a tremendous amount of work on the LIL for various kinds of dependent structures and for stochastic processes. See Stout [26] and Bingham [1] for excellent surveys.

Work in laws of the iterated logarithms has also been carried out by many authors in connection with statistics. Robbins [22] applied LIL-type results to sequential tests of powers. Heyde [13] gave an LIL result for martingales and found its applications to regression and time series. Hall [10] established LIL's in density estimation. Recently, Dabrowska [5] considered it for Kaplan-Meier estimates for censored data. The same law for U-statistics was obtained by Dehling [7]. Iverson and Randles [16] studied the effect on LIL convergence by substituting parameters into U-statistics and applying their theorems to adaptive M-estimators and other statistics.

Stronger asymptotic representations with more precise error rates are usually obtained by invariance principles. Initiated by Strassen [27], the invariance principle has been established mainly for partial sums under a variety of assumptions. In statistics, classical work centers around the Bahadur representations

Received by the editors May 19, 1993.

1991 *Mathematics Subject Classification*. Primary 62F12, 60F17; Secondary 60F15, 62F35.

Key words and phrases. Bahadur representation, invariance principle, law of the iterated logarithm, location-scatter, M-estimator, robustness, S-estimator.

The research of the first author was partially supported by a National University of Singapore Research Project Grant. The research of the second author was partially supported by a DePaul University Research Grant. The authors are grateful to a referee for his very helpful comments and suggestions.

©1994 American Mathematical Society
0002-9939/94 \$1.00 + \$.25 per page

for sample quantiles; see Kiefer [18] for a bibliography. For one-dimensional M-estimators of location (with a preliminary estimate of the scale), a strong invariance principle was obtained by Carroll [3]. Related results also appear in Jurečková and Sen [17] and Sen [24].

The purpose of the present paper is to establish the law of the iterated logarithm and a strong invariance principle for a large class of multivariate M-estimators. Such estimators frequently show up in many areas of modern statistics. In univariate cases LIL for M-estimators were discussed by Serfling [25] and Boos and Serfling [2]. Direct generalizations of their results are possible, but the conditions would become too strong for many interesting and commonly-used M-estimators. In this area, we find Huber's [14] approach useful. In the spirit of Huber [14], we give reasonable conditions under which the law of the iterated logarithm holds for a rather general class of M-estimators which covers in particular robust M-estimators of multivariate location and scatter. An almost sure invariance principle with an error term of $o(n^{-1+\epsilon})$ for any $\epsilon > 0$ is also obtained.

The rest of the paper is organized as follows. In §2, our main theorems of the LIL and invariance principle are given for general M-estimators. We also consider the minimum L_p distance estimators of location and M-estimators of direction as examples. In §3, our results are applied to M-estimators and S-estimators of multivariate location and scatter, two classes of estimators that frequently come up in robust statistics.

2. Main results

Suppose that $x_1, x_2, \dots, x_n$ are independent observations from an underlying distribution F_θ ($\theta \in \Theta \subset R^m$). An M-estimator θ_n of θ is a solution of the equation $\sum_{i=1}^{n} \psi(x_i, \theta_n) = 0$ where $\psi(x, \theta)$ is a score function which maps from Θ to R^m for each x. The maximum likelihood estimators and the least squares estimators are special cases of M-estimators. Many robust estimators take the same form. In fact, discussions regarding M-estimators in location, scale, regression, and time series among others are abundant in recent years, mainly due to their desirable robustness properties (see [11]). Strong consistency and asymptotic normality for general M-estimators were studied by Huber [14]. Our theorem below is stated in the same framework. We fix notation and conditions first.

Let $\lambda(\theta) = E\psi(x, \theta)$ be the expected score under a given distribution F and

$$u(x, \theta, d) = \sup_{|\tau-\theta|\le d} |\psi(x, \tau) - \psi(x, \theta)|$$

where $|\cdot|$ is taken to be the sup-norm: $|\theta| = \max(|\theta_1|, \dots, |\theta_m|)$. We assume that $\theta_n \equiv \theta_n(x_1, \dots, x_n)$ satisfies

$$\frac{1}{\sqrt{n \log\log n}} \sum_{i=1}^{n} \psi(x_i, \theta_n) \to 0 \quad \text{a.s.} \tag{2.1}$$

with the following conditions

(M1) For each fixed $\theta \in \Theta$, $|\psi(x, \theta)|^2$ is integrable and $\psi(x, \theta)$ is separable in the sense of Doob: there is a F-null set N and a countable subset

$\Theta' \subset \Theta$ such that for every open set $U \subset \Theta$ and every closed interval A, the sets $\{x : \psi(x, \theta) \in A, \ \forall\theta \in U\}$, $\{x : \psi(x, \theta) \in A, \ \forall\theta \in U \cap \Theta'\}$ differ by a subset of N.

(M2) There is a $\theta_0 \in \Theta$ such that $\lambda(\theta_0) = 0$ and λ has a non-singular derivative Λ at θ_0.

(M3) There exist positive numbers $a, b, c, d, \alpha, \beta$, and d_0 such that $\alpha \geq \beta > 2$, and

(i) $|\lambda(\theta)| \geq a|\theta - \theta_0|$ for $|\theta - \theta_0| \leq d_0$,
(ii) $Eu(x, \theta, d) \leq bd$ for $|\theta - \theta_0| + d \leq d_0$,
(iii) $Eu^\alpha(x, \theta, d) \leq cd^\beta$ for $|\theta - \theta_0| + d \leq d_0$, or
(iii)′ $Eu^2(x, \theta, d) \leq cd$ for $|\theta - \theta_0| + d \leq d_0$ and $\sup_{x \in \mathscr{X}} u(x, \theta_0, d_0) < \infty$.

(M4) $|\theta_n - \theta_0| \leq d_0$ almost surely as $n \to \infty$.

For any random vector $Y = (Y_1, Y_2, \ldots, Y_n)$, we define $\sigma(Y) = (\sigma(Y_1), \sigma(Y_2), \ldots, \sigma(Y_n))$ where $\sigma^2(Y_i)$ is the variance of Y_i.

Theorem 2.1. *Under conditions* (M1)–(M4), *any sequence* θ_n *satisfying* (2.1) *obeys the law of the iterated logarithm*:

$$\limsup_{n\to\infty} \frac{\sqrt{n}(\theta_n - \theta_0)}{\sqrt{2\log\log n}} = \sigma(\Lambda^{-1}\psi(x, \theta_0)),$$
$$\liminf_{n\to\infty} \frac{\sqrt{n}(\theta_n - \theta_0)}{\sqrt{2\log\log n}} = -\sigma(\Lambda^{-1}\psi(x, \theta_0)).$$

Note that condition (M2) implies (M3)(i). We spell it out explicitly for convenience of the proof given later. Conditions (M3)(ii), (iii) imply $Eu^2(x, \theta, d) \leq ed$ for some positive number e, which is condition (N3)(ii) of [14]. Our condition (M4) holds automatically if the estimator is asymptotically consistent. If for almost every x and for some positive numbers r and s,

$$u(x, \theta, d) \leq rd^s \quad \text{for} \quad |\theta - \theta_0| + d \leq d_0,$$

then (M3)(iii) is satisfied. In addition, if $s \geq 1/2$, then both (M3)(iii) and (M3)(iii)′ hold. If the score function is Lipschitz, we have the following:

Corollary 2.1. *Suppose that for almost every* x, $\psi(x, \theta)$ *is a continuous function of* θ *and there exists an open neighborhood* Θ_0 *around* θ_0 *such that*

$$|\psi(x, \theta) - \psi(x, \tau)| \leq L|\theta - \tau|, \qquad \theta, \tau \in \Theta_0,$$

for some constant $L < \infty$. *Then any sequence of strongly consistent M-estimators obeys the LIL provided that* $\lambda(\theta_0) = 0$ *and* $\lambda(\theta)$ *has a nonsingular derivative at* θ_0.

The proof of Theorem 2.1 relies on the following lemma, which can be viewed as an analogue to Lemma 3 of [14]. Let

$$Z_n(\tau, \theta) = \frac{|\sum_{i=1}^n [\psi(x_i, \tau) - \psi(x_i, \theta) - \lambda(\tau) + \lambda(\theta)]|}{\sqrt{n \log\log n} + n|\lambda(\tau)|}.$$

Lemma 2.1. *Assumptions* (M1)–(M3) *imply* $\lim_{n\to\infty} n^\gamma \sup_{|\tau - \theta_0| \leq d_0} Z_n(\tau, \theta_0) \to 0$ *a.s. for any* $\gamma < (\beta - 2)/(2(\alpha + m + 1))$ *under* (M3)(iii) *or* $\gamma < 1/4$ *under* (M3)(iii)′.

Lemma 2.1 also enables us to strengthen the almost sure representation for θ_n if (2.1) is replaced by

$$\frac{1}{\sqrt{n \log \log n}} \sum_{i=1}^{n} \psi(x_i, \theta_n) = o(n^{-\gamma}). \tag{2.2}$$

Theorem 2.2. *Under the same conditions of Theorem* 2.1, *if* $\frac{\partial}{\partial \theta}\lambda(\theta)$ *is Lipschitz in a neighborhood of* θ_0, *then any sequence* θ_n *satisfying* (2.2) *has the following almost sure expansion*:

$$\theta_n - \theta_0 = -\frac{1}{n} \sum_{i=1}^{n} \Lambda^{-1} \psi(x_i, \theta_0) + o(n^{-(\frac{1}{2}+\gamma)} \sqrt{\log \log n}) \tag{2.3}$$

where γ *is given in Lemma* 2.1. *Under the conditions of Corollary* 2.1, *the remainder term is* $o(n^{-1+\epsilon})$ *for any* $\epsilon > 0$.

The last statement follows from the fact that (M3)(iii) holds for any $\alpha = \beta > 2$ and, consequently, γ can be arbitrarily close to $1/2$. Although our arguments lead to an error rate of $o(n^{-1+\epsilon})$, we believe that some improvement is possible to replace n^ϵ by some order of $\log n$.

As an application of Theorem 2.1, we consider the minimum L_p distance estimates of location (in R^m) defined by $\sum_{i=1}^{n} \|x_i - \theta_n\|^p = \min$. When $p > 1$ or $p = 1$ but $m \geq 3$, it is easy to show (cf. Huber [14, p. 232]) that the estimate satisfies our conditions with (M3)(iii). When $p = 1$ and $m \leq 2$, condition (M3)(iii)$'$ is satisfied instead. Therefore, we have

Corollary 2.2. *For* $1 \leq p \leq 2$, *the minimum* L_p *distance estimators of multivariate location satisfy the law of the iterated logarithm.*

In the special case of $m = 1$ and $p = 1$, the estimator corresponds to the sample median. Theorem 2.2 provided a Bahadur representation with a remainder of $O(n^{-3/4+\epsilon})$ for any $\epsilon > 0$, which is slightly behind the optimal rate of $n^{-3/4}(\log \log n)^{3/4}$.

In cases where the sample space and the parameter space are both compact, our conditions usually become easier to check. For instance, for directional data modeled by, say, a von Mises distribution, an M-estimator of location solves

$$\sum_{i=1}^{n} \psi(\kappa^{1/2} \|x_i - \theta\|)(x_i - (x_i'\theta)\theta) = 0$$

for some score function ψ; see He and Simpson [12, p. 362]. Since both x and θ are on unit spheres, if ψ has a bounded derivative, then it is easy to verify our conditions with (M3)(iii)$'$.

Proofs of our main results are given in the rest of the section. Our proof of Lemma 2.1 uses a variant of Huber's method helped by the Marcinkiewicz-Zygmund inequality (see, for example, Chow and Teicher [4]) and Freedman's inequality to obtain sharper estimates of probabilities. First we list these two classical inequalities for reference.

Marcinkiewicz and Zygmund Inequality ([20]). *If $X_1, \ldots, X_n$ are independent and identically distributed random variables with mean 0, then for all $m \geq 2$,*

$$E|X_1 + \cdots + X_n|^m \leq c_m n^{\frac{m}{2}} E|X_1|^m$$

for some constants c_m depending only on m.

Freedman Inequality ([8]). *Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with mean 0. If $|X_i| \leq M$ for $1 \leq i \leq n$, then*

$$P\{|X_1 + \cdots + X_n| \geq a\} \leq \exp\left(-\frac{a^2}{2(Ma + nEX^2)}\right),$$

where X has the same distribution of X_i.

Proof of Lemma 2.1. In the following proof, we use C to denote a positive constant independent of n and ε which may vary from line to line. Without loss of generality, let $\theta_0 = 0$ and $d_0 = 1$ in our assumptions.

Put $q = \frac{1}{M}$, where $M \geq 2$ is an integer to be chosen later. For any $K_0 \geq 0$, divide the hypercube $C_0 = \{\theta : |\theta| \leq 1\}$ into $N + 1$ disjoint hypercubes $C_{(1)}, \ldots, C_{(N)}, C_{K_0}$ the same way as in Huber's proof of his Lemma 3. Then $N \leq K_0(2M)^m + 1$, as there are at most $(2M)^m$ hypercubes having size $q(1-q)^k$ and centered $(1 - q/2)(1 - q)^k$ away from 0 for each $k = 0, 1, \ldots, K_0 - 1$. They will be called group k hypercubes and denoted by $C_1^k, C_2^k, \ldots$. Given n and small $\varepsilon > 0$, by choosing $M = M(\varepsilon)$ and $K_0 = K_0(n, \varepsilon)$ properly, we can estimate the right hand-side of

$$\begin{aligned} &P\{\sup_{\tau\in C_0} Z_n(\tau, 0) \geq 2\varepsilon\} \\ &\quad \leq P\{\sup_{\tau\in C_{K_0}} Z_n(\tau, 0) \geq 2\varepsilon\} + \sum_{j=1}^{N} P\{\sup_{\tau\in C_{(j)}} Z_n(\tau, 0) \geq 2\varepsilon\} \\ &\quad = P\{\sup_{\tau\in C_{K_0}} Z_n(\tau, 0) \geq 2\varepsilon\} + \sum_{k=0}^{K_0-1}\sum_{j} P\{\sup_{\tau\in C_j^k} Z_n(\tau, 0) \geq 2\varepsilon\}. \end{aligned} \tag{2.4}$$

In fact, we shall choose M to be the smallest integer no less than $3b/\epsilon a$ and $K_0 \equiv K_0(n)$ the largest integer less than $\log n/\{2|\log(1-q)|\}$. As a result,

$$N = O(\varepsilon^{-(m+1)} \log n) \quad \text{and} \quad (1-q)^{K_0} = n^{-1/2} + o(n^{-1/2}). \tag{2.5}$$

We first estimate $P\{\sup_{\tau\in C_j^k} Z_n(\tau, 0) \geq 2\varepsilon\}$ for $1 \leq j$ and $0 \leq k \leq K_0 - 1$.

Let ξ_j and $2d_j$ be the center and the length of C_j^k, respectively. Then $|\xi_j| = (1-q/2)(1-q)^k$ and $d_j = q/2(1-q)^k$. For $\tau \in C_j^k$, by condition (M3), we have

$$|\lambda(\tau)| \geq a|\tau| \geq a(|\xi_j| - d_j) = a(1-q)^k, \tag{2.6}$$

$$Eu(x, \xi_j, d_j) \leq bd_j \leq b(1-q)^k q. \tag{2.7}$$

By the triangle inequality, we obtain

$$\sup_{\tau\in C_j^k} Z_n(\tau, 0) \leq U_n + V_n, \tag{2.8}$$

where the functions $U_n = U_n^{(k)}$ and $V_n = V_n^{(k)}$ are defined as follows:

$$U_n = \frac{1}{na(1-q)^k} \sum_{i=1}^{n} [u(x_i, \xi_j, d_j) + Eu(x, \xi_j, d_j)],$$

$$V_n = \frac{1}{na(1-q)^k} \left| \sum_{i=1}^{n} [\psi(x_i, \xi_j) - \psi(x_i, 0) - \lambda(\xi_j)] \right|.$$

Let $S_n^j = \sum_{i=1}^{n} [u(x_i, \xi_j, d_j) - Eu(x, \xi_j, d_j)]$. Following the argument of Huber [14, p. 229], we have by using (2.6), (2.7), and the definition of $M = 1/q$,

$$P\{U_n \geq \varepsilon\} \leq P\{S_n^j \geq nbq(1-q)^k\}.$$

If the condition (M3)(iii) holds, by the Marcinkiewicz and Zygmund inequality and the fact that $(1-q)^k \geq (1-q)^{K_0} \sim n^{-1/2}$, we have for some constant C and for sufficiently large n,

$$P\{U_n \geq \varepsilon\} \leq \frac{cc_\alpha n^{\alpha/2} (q(1-q)^k)^\beta}{(nbq(1-q)^k)^\alpha} \leq C\varepsilon^{\beta-\alpha} n^{-\beta/2}.$$

Under the condition (M3)(iii)$'$, Freedman's inequality yields

$$P\{U_n \geq \varepsilon\} \leq \exp\left(\frac{-n^2 b^2 q^2 (1-q)^{2k}}{2(nBbq(1-q)^k + ncq(1-q)^k)} \right) \leq \exp(-C\varepsilon n^{1/2}),$$

instead, where $B = \sup_x u(x, 0, 1)$. Similarly, $\sum_{i=1}^{n} [\psi(x_i, \xi_j) - \psi(x_i, 0) - \lambda(\xi_j)]$ is a sum of i.i.d. mean 0 random variables and $|\psi(x_i, \xi_j) - \psi(x_i, 0) - \lambda(\xi_j)| \leq u(x_i, 0, \xi_j) + Eu(x_i, 0, \xi_j)$; thus,

$$P\{V_n \geq \varepsilon\} \leq \frac{cc_\alpha n^{\alpha/2} 2^\alpha (1-q)^{k\beta}}{(na\varepsilon(1-q)^k)^\alpha} \leq C\varepsilon^{-\alpha} n^{-\beta/2}$$

or

$$P\{V_n \geq \varepsilon\} \leq \exp\left(\frac{-n^2 a^2 \varepsilon^2 (1-q)^{2k}}{2(2nBa\varepsilon(1-q)^k + 4nc(1-q)^k)} \right) \leq \exp(-C\varepsilon^2 n^{1/2})$$

under condition (M3)(iii) or (M3)(iii)$'$. Hence it follows from (2.8) that

$$P\{\sup_{\tau \in C_j^k} Z_n(\tau, 0) \geq 2\varepsilon\} \leq Cn^{-\beta/2} \quad \text{or} \quad \exp(-Cn^{1/2}) \tag{2.9}$$

for sufficiently large n.

Next we estimate $P\{\sup_{\tau \in C_{K_0}} Z_n(\tau, 0) \geq 2\varepsilon\}$. By the triangle inequality again, we have

$$\sup_{\tau \in C_{k_0}} Z_n(\tau, 0) \leq \frac{\sum_{i=1}^{n} [u(x_i, 0, d) + Eu(x, 0, d)]}{\sqrt{n \log \log n}}$$

where $d = (1-q)^{K_0} \leq n^{-\frac{1}{2}}$. Let $S_n = \sum_{i=1}^{n} [u(x_i, 0, d) - Eu(x, 0, d)]$. Since $nEu(x, 0, d) \leq nbd \leq b\sqrt{n}$, it follows that $\varepsilon\sqrt{n \log \log n} - 2nEu(x, 0, d) \geq \varepsilon\sqrt{n}$ when n is large enough. Thus, under condition (M3) we have

$$P\{\sup_{\tau \in C_{K_0}} Z_n(\tau, 0) \geq 2\varepsilon\} \leq C\varepsilon^{-\alpha} n^{-\beta/2} \quad \text{or} \quad \exp(-C\varepsilon^2 n^{1/2}). \tag{2.10}$$

Combining (2.4), (2.5), (2.9), and (2.10), we see that

$$P\{\sup_{\tau\in C_0} Z_n(\tau, 0) \geq 2\varepsilon\} = C\varepsilon^{-(\alpha+m+1)} n^{-\beta/2} \log n$$

or

$$C\varepsilon^{-(m+1)} \exp(-C\varepsilon^2 n^{1/2}) \log n$$

if condition (M3)(iii) or (M3)(iii)$'$ holds. For any $\delta > 0$, let $\varepsilon = \delta n^{-\gamma}$, we have shown that $P\{n^\gamma \sup_{\tau\in C_0} Z_n(\tau, 0) \geq 2\delta\} = P\{\sup_{\tau\in C_0} Z_n(\tau, 0) \geq 2\varepsilon\}$ is summable over n. This proves the lemma. As a consequence, we have

Lemma 2.2. *Under the conditions* (M1)–(M4), *if* θ_n *satisfies* (2.1), *then*

$$\frac{1}{\sqrt{n \log\log n}} \sum_{i=1}^{n} \psi(x_i, \theta_0) + \sqrt{\frac{n}{\log\log n}}\lambda(\theta_n) \to 0 \quad a.s. \tag{2.11}$$

In particular, $\lim_{n\to\infty} \theta_n = \theta_0$ *a.s.*

Lemma 2.2 implies that condition (M4) is equivalent to the consistency of θ_n under conditions (M1)–(M3).

Proof of Lemma 2.2. Assume again $\theta_0 = 0$, $d_0 = 1$. Let Ω' be the set of ω such that

$$\varlimsup_{n\to\infty} |\theta_n| \leq 1,$$

$$\varlimsup_{n\to\infty} \frac{1}{\sqrt{2n \log\log n}} \sum_{i=1}^{n} |\psi(x_i, 0)| = K' < K < \infty,$$

$$\sup_{|\tau|\leq 1} Z_n(\tau, 0) \to 0$$

where K is as given in the proof of Lemma 2.1. By the classical law of the iterated logarithm and Lemma 2.1, $P(\Omega') = 1$. We will show that (2.11) holds on Ω'.

Let $\frac{1}{2} > \varepsilon > 0$. For each $\omega \in \Omega'$, there exists an $N = N(\varepsilon, K, \omega)$ such that for $n \geq N$,

$$\sup_{|\tau|\leq 1} Z_n(\tau, 0) \leq \varepsilon \quad \text{and} \quad |\theta_n| \leq 1, \tag{2.12}$$

$$\frac{1}{\sqrt{n \log\log n}} \sum_{i=1}^{n} |\psi(x_i, \theta_n)| \leq \varepsilon, \tag{2.13}$$

and

$$\frac{1}{\sqrt{n \log\log n}} \left| \sum_{i=1}^{n} \psi(x_i, 0) \right| \leq 2K. \tag{2.14}$$

Therefore (2.12)–(2.14) together with the inequality

$$\frac{\left|\sum_{i=1}^{n} [\psi(x_i, 0) + \lambda(\theta_n)]\right|}{\sqrt{n \log\log n} + n\lambda(\theta_n)} \leq \sup_{|\tau|\leq 1} Z_n(\tau, 0) + \frac{1}{\sqrt{n \log\log n}} \left| \sum_{i=1}^{n} \psi(x_i, \theta_n) \right|,$$

which is an application of the triangle inequality, imply

$$\left|\sum_{i=1}^{n}(\psi(x_i, 0) + \lambda(\theta_n))\right| \le 2\varepsilon(\sqrt{n \log\log n} + n\lambda(\theta_n)). \tag{2.15}$$

Then by (2.14),

$$\sqrt{\frac{n}{\log\log n}}\lambda(\theta_n)(1-2\varepsilon) \le 2\varepsilon + \frac{1}{\sqrt{n \log\log n}}\left|\sum_{i=1}^{n}\psi(x_i, 0)\right| \le K + 2\varepsilon.$$

Dividing (2.15) by $\sqrt{n \log\log n}$, we have

$$\left|\frac{1}{\sqrt{n \log\log n}}\sum_{i=1}^{n}\psi(x_i, 0) + \sqrt{\frac{n}{\log\log n}}\lambda(\theta_n)\right| \le 2\varepsilon\frac{K+1}{1-2\varepsilon}. \tag{2.16}$$

The first part of the lemma follows by letting $\varepsilon \to 0$. The second part of the lemma follows from (2.14), (2.16), and condition (M3)(i). This completes the proof for Lemma 2.2.

Now we go back to the proof of Theorem 2.1. Since λ has a nonsingular derivative Λ at θ_0, we have

$$\theta_n - \theta_0 = \Lambda^{-1}(\lambda(\theta_n) - \lambda(\theta_0)) + o(|\theta_n - \theta_0|).$$

By Lemma 2.2, this leads to

$$\frac{\sqrt{n}(\theta_n - \theta_0)}{\sqrt{2 \log\log n}} + o(\frac{\sqrt{n}|\theta_n - \theta_0|}{\sqrt{2 \log\log n}}) = -\frac{1}{\sqrt{2n \log\log n}}\sum_{i=1}^{n}\Lambda^{-1}\psi(x_i, \theta_0) + o(1).$$

Using the LIL for the sum of i.i.d. random variables, we first obtain

$$\limsup_{n\to\infty}\frac{\sqrt{n}|\theta - \theta_0|}{\sqrt{2\log\log n}} < \infty$$

and then

$$\frac{\sqrt{n}(\theta_n - \theta_0)}{\sqrt{2 \log\log n}} = -\frac{1}{\sqrt{2n \log\log n}}\sum_{i=1}^{n}\Lambda^{-1}\psi(x_i, \theta_0) + o(1).$$

Theorem 2.1 is an immediate consequence of the above.

If (2.1) is replaced by (2.2), the same proof shows that (2.11) can be strengthened to

$$\frac{1}{\sqrt{n \log\log n}}\sum_{i=1}^{n}\psi(x_i, \theta_0) + \sqrt{\frac{n}{\log\log n}}\lambda(\theta_n) = o(n^{-\gamma}). \tag{2.17}$$

If $\frac{\partial}{\partial\theta}\lambda(\theta)$ is Lipschitz in a neighborhood of θ_0, we then have

$$\theta_n - \theta_0 = \Lambda^{-1}(\lambda(\theta_n) - \lambda(\theta_0)) + O(|\theta_n - \theta_0|^2). \tag{2.18}$$

Theorem 2.1 implies that

$$O(|\theta_n - \theta_0|^2) = O(\log\log n/n) = o(n^{-(1/2+\gamma)}\sqrt{\log\log n})$$

for any $\gamma < 1/2$. Combining (1.17) and (2.18), we obtain Theorem 2.2.

3. Applications to M- and S-estimators of multivariate location and scatter

Consider a family of elliptically symmetric distributions $\{F_{\mu,\Sigma}\}$ with F having the probability density function of the form

$$f(x, \mu, \Sigma) = |\Sigma|^{-1/2} g((x-\mu)'\Sigma^{-1}(x-\mu))$$

where $\theta = (\mu, \Sigma)$, $\mu \in R^m$, and $\Sigma \in R^{m\times m}$ is positive definite. In what follows, we denote by $d^2 \equiv d^2(x, \mu, \Sigma) = (x-\mu)'\Sigma^{-1}(x-\mu)$ the Σ-standardized distance between x and μ. Also, we use $|\Sigma|$ and $||\Sigma||$ to denote the determinant and L_2 norm of the matrix Σ, respectively. Σ_{ij} will be the usual (i, j)-th element of the matrix.

Given n independent observations $x_1, \dots, x_n$, an M-estimator of location-scatter θ is the solution $\hat{\theta}_n = (t, S)$ of

$$\left\{ \begin{array}{l} \sum_{i=1}^n u_1(d_i)(x_i - t) = 0, \\ \sum_{i=1}^n \{u_2(d_i)(x_i-t)(x_i-t)' - u_3(d_i)S\} = 0 \end{array} \right. \tag{3.1}$$

where u_1, u_2, and u_3 are properly chosen real-valued functions on $[0, \infty)$ and $d_i = d(x_i, t, S)$. The existence, uniqueness, asymptotic distributions and robustness of such M-estimators at any $F_{\mu,\Sigma}$ were discussed by Maronna [21]. Note that our u_2 function differs slightly from that of Maronna [21] simply for our notational convenience.

In the development of high breakdown point estimators, Davies [6], following the idea of Rousseeuw and Yohai [23], considered S-estimators of θ defined as the solution to the problem of minimizing $|S|$ subject to

$$n^{-1}\sum_{i=1}^n \rho(d_i) = b_0 \tag{3.2}$$

for properly chosen $\rho : [0, \infty) \to [0, \infty)$ and $b_0 > 0$. The relationship between M- and S-estimators is explored by Lopuhaä [19]. Under mild conditions, both the M- and S-estimators $\hat{\theta}_n$ can be consistent and asymptotically normal.

We establish the Bahadur-type representation for the M-estimators first by verifying the conditions of Corollary 2.1. Let

$$\begin{aligned} \Psi_1(x, \tau) &= u_1(d)(x-t), \\ \Psi_2(x, \tau) &= u_2(d)(x-t)(x-t)' - u_3(d)S, \end{aligned} \tag{3.3}$$

and $\Psi = (\Psi_1, \Psi_2)$, where $\tau = (t, S)$, $d^2 = d^2(x, t, S) = (x-t)'S^{-1}(x-t)$.

Theorem 3.1. *Suppose the functions u_i $(i = 1, 2, 3)$ are piecewise differentiable and satisfy the following conditions:*

(U1) *$u_1(d)$, $u_2(d)d$, and $u_3(d)$ are bounded.*
(U2) *$u_1'(d)d$, $u_2'(d)d^2$, and $u_3'(d)$ are bounded.*
(U3) *$u_1'(d)d^2$, $u_2'(d)d^3$, and $u_3'(d)d$ are bounded.*

If $\lambda(\tau) = E_{F_\theta}\Psi(X, \tau)$ has a nonsingular derivative Λ at $\tau = \theta$ and $\lambda(\theta) = 0$, then any sequence of strongly consistent M-estimators θ_n obeys the LIL. If, in addition, $\frac{\partial}{\partial\tau}\lambda(\tau)$ is Lipschitz in a neighborhood of θ, then θ_n satisfies the following invariance principle for any $\epsilon > 0$:

$$\theta_n - \theta = -\frac{1}{n}\sum_{i=1}^n \Lambda^{-1}\Psi(x_i, \theta) + o(n^{-1+\epsilon}). \tag{3.4}$$

Proof. Let $d^2 = (x-t)'S^{-1}(x-t)$ and λ be the smallest eigenvalue of S. By Corollary 2.1 and Theorem 2.2, it suffices to show that

$$\sup_{\|\theta-\theta_0\|\leq\frac{\lambda}{2p}} \sup_{x\in\mathscr{X}} \left|\frac{\partial}{\partial\theta}\psi(x,\theta)\right|$$

is bounded. This follows from $\frac{1}{\lambda}d^2 \leq \|x-t\|^2 \leq \|S\|d^2$, $(x_i-t_i)(x_j-t_j) \leq \frac{1}{2}\|x-t\|^2$, and

$$\frac{\partial\psi_1}{\partial t} = -\left(\frac{u_1'(d)}{d}S^{-1}(x-t)(x-t)' + u_1(d)I\right),$$

$$\frac{\partial\psi_{1,j}}{\partial S} = -\frac{u_1'(d)}{2d}(x_j-t_j)(2V-D_V),$$

$$\frac{\partial\psi_{2,ij}}{\partial t} = -\frac{u_2'(d)}{d}(x_i-t_i)(x_j-t_j)S^{-1}(x-t) + u_2(d)\frac{\partial(x_i-t_i)(x_j-t_j)}{\partial t} + \frac{u_3'(d)}{d}S_{ij}S^{-1}(x-t),$$

$$\frac{\partial\psi_2}{\partial S_{ij}} = -\frac{u_2'(d)}{2d}(2V-D_V)_{ij}(x-t)(x-t)' + \frac{u_3'(d)}{2d}(2V-D_V)_{ij}S - u_3(d)\frac{\partial S}{\partial S_{ij}}$$

where $V = S^{-1}(x-t)(x-t)'S'^{-1}$ and $D_V = \operatorname{diag}\{V_{11},\ldots,V_{mm}\}$. We refer to [9] for differentiation with respect to matrices.

Remark. If u_i's are piecewise twice differentiable and satisfy additional conditions (U4)–(U6) below, then one can show by calculus that $\lambda(\tau)$ is twice differentiable as long as g is continuous.

(U4) $u_2(d)$, $u_1'(d)$, $u_2'(d)d$, and $u_3'(d)/d$ are bounded.
(U5) $u_1''(d)d$, $u_2''(d)d^2$, and $u_3''(d)$ are bounded.
(U6) $u_1''(d)d^3$, $u_2''(d)d^4$, and $u_3''(d)d^2$ are bounded.

These conditions are satisfied by common robust M-estimators including Huber's proposal 2 where $u_1(r) = \psi_k(r)/r$, $u_2(r) = \psi_{k^2}(r^2)/r^2$, $u_3(r) = 1$, and $\psi_k(r) = \min\{k, \max\{r, -k\}\}$ for any $k > 0$.

Let $\psi(x) = \rho'(x)$, and $u(x) = \psi(x)/x$. Then the S-estimators given by (3.2) satisfy the first-order condition of M-estimators

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(x_i,\theta_n) = 0, \tag{3.5}$$

where Ψ is defined as in (3.3) with $u_1(d) = u(d)$, $u_2(d) = mu(d)$, and $u_3(d) = \psi(d)d - \rho(d) + b_0$ (see Lopuhaä [19, p. 1666]). Since $\psi(x)$ is zero for all $x > c_0$, it is then straightforward to verify the conditions of Theorem 3.1. We omit the details.

References

1. N. H. Bingham, *Variants on the law of the iterated logarithm*, Bull. London Math. Soc. **18** (1986), 433–467.

2. D. D. Boos and R. J. Serfling, *A note on differentials and the CLT and LIL for statistical functions, with applications to M-estimates*, Ann. Statist. **8** (1980), 618–624.

3. R. J. Carroll, *On almost sure expansions for M-estimates*, Ann. Statist. **6** (1978), 314–318.
4. Y. S. Chow and T. Teicher, *Probability theory: Independence, interchangeability, martingale*, Springer-Verlag, New York, 1978.
5. D. M. Dabrowska, *Kaplan-Meier estimate on the plane: Weak convergence, LIL, and the bootstrap*, J. Multivariate Anal. **29** (1989), 308–325.
6. P. L. Davies, *Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices*, Ann. Statist. **14** (1987), 1269–1292.
7. H. Dehling, *Complete convergence of triangular arrays and the law of the iterated logarithm for U-statistics*, Stat. Probab. Lett. **7** (1989), 319–321.
8. D. Freedman, *On tail probability for martingales*, Ann. Probab. **3** (1975), 100–118.
9. F. A. Greybill, *Matrices with application in statistics*, 2nd ed., Wadsworth, Belmont, CA, 1983.
10. P. Hall, *Laws of the iterated logarithm for non-parametric density estimators*, Z. Wahrsch. Verw. Gebiete **56** (1981), 47–61.
11. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust statistics: The approach based on influence functions*, Wiley, New York, 1986.
12. X. He and D. G. Simpson, *Robust direction estimation*, Ann. Statist. **20** (1992), 351–369.
13. C. C. Heyde, *An iterated logarithm result for martingales and its application in estimation theory for autoregressive processes*, J. Appl. Probab. **10** (1973), 146–157.
14. P. J. Huber, *The behavior of maximum likelihood estimates under nonstandard conditions*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, 1967, pp. 221–233.
15. ——, *Robust statistics*, Wiley, New York, 1981.
16. H. K. Iverson and R. H. Randles, *The effects on convergence of substituting parameter estimates into U-statistics and other families of statistics*, Probab. Theory Related Fields **81** (1989), 453–471.
17. J. Jurečková and P. K. Sen, *Invariance principles for some stochastic processes relating to M-estimators and their role in sequential statistical inference*, Sankhyā Ser. A **43** (198)), 190–210.
18. J. Kiefer, *Old and new methods for studying order statistics and sample quantiles*, Proc. Conference on Nonparametric Techniques in Statistical Inference (M. L. Puri, ed.), Cambridge Univ. Press, New York, 1970, pp. 349–357.
19. H. P. Lopuhaä, *On the relation between S-estimators and M-estimators of multivariate location and covariance*, Ann. Statist. **17** (1989), 1662–1683.
20. J. Marcinkiewicz and A. Zygmund, *Sur les fonctions independantes*, Fund. Math. **29** (1937), 60–90.
21. R. A. Maronna, *Robust M-estimators of multivariate location and scatter*, Ann. Statist. **4** (1976), 51–67.
22. H. Robbins, *Statistical methods related to the law of the iterated logarithm*, Ann. Math. Statist. **41** (1970), 1397–1409.
23. P. R. Rousseeuw and V. Yohai, *Robust regression by means of S-estimators*, Robust and Nonlinear Time Series Analysis (J. Franke, W. Hardle, and R.D. Martin, eds.), Lecture Notes in Statistics, vol. 26, Springer-Verlag, New York, 1984, pp. 256–272.
24. P. K. Sen, *Sequential nonparametrics: Invariance principle and statistical inference*, Wiley, New York, 1981.
25. R. J. Serfling, *Approximation theorems of mathematical statistics*, Wiley, New York, 1980.
26. W. F. Stout, *Almost sure convergence*, Academic Press, New York, 1974.
27. V. Strassen, *An invariance principle for the law of the iterated logarithm*, Z. Wahrsch. Verw. Gebiete **3** (1964), 211–226.

DEPARTMENT OF STATISTICS, UNIVERSITY OF ILLINOIS, CHAMPAIGN, ILLINOIS 61820

DEPARTMENT OF MATHEMATICS, DEPAUL UNIVERSITY, CHICAGO, ILLINOIS 60614