SECOND ORDER ASYMPTOTICS FOR M-ESTIMATORS UNDER NON-STANDARD CONDITIONS

NICOLAS W. HENGARTNER AND MARTEN H. WEGKAMP

ABSTRACT. This paper establishes, under non-standard conditions, an explicit stochastic approximation of studentized M-estimators $\widehat{\theta}_n$, implicitly defined as solutions to $\sum_{j=1}^n \psi(X_j, \widehat{\theta}) = o(n^{-1})$, by a U-statistic U_n that is probably concentrated about $\hat{\theta}_n$ in the sense that $\mathbb{P}[|\hat{\theta}_n - U_n| > (n \log n)^{-1}] = o(n^{-1/2})$. The expansion and concentration hold under weaker smoothness conditions on ψ than those assumed by Lahiri (1994). This approximation is key in rigorously establishing a second order expansion for the sampling distribution of the studentized estimator. Under stronger smoothness assumptions on ψ, a similar expansion relates the bootstrap approximation to the true distribution of the studentized M-estimator.

1. INTRODUCTION

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables from a common probability measure P, of which we want to estimate the population parameter $\theta = \theta_P$ implicitly defined as the unique solution of

$$\int \psi(x, \theta_P) dP(x) = 0, \tag{1.1}$$

for some measurable function $\psi(x, \theta)$. For ease of exposition, we shall restrict our attention to univariate parameters. Generalization to the multidimensional parameter case is straightforward, and therefore will not be pursued here.

Denote by P_n the empirical probability measure based on the sample $X_1, \ldots, X_n$ which assigns mass $P_n(A) = n^{-1}\#(X_j \in A)$ to each set A. This paper studies M-estimators $\hat{\theta}$ that solve the empirical counterpart of (1.1) within $o(n^{-1})$, e.g.,

$$\int \psi(x, \hat{\theta})\, dP_n(x) = o(n^{-1}). \tag{1.2}$$

Key words and key phrases: Bootstrap, Edgeworth expansion, M-estimation, Second order asymptotics, Stochastic approximation.

AMS 1991 Classification: Primary 62E17, 62P11; Secondary 65U05.

There is a considerable amount of literature about M-estimators. Huber (1967) proved the asymptotic normality of M-estimators under "non standard" conditions that do not require explicit smoothness assumptions on $\psi(x,\theta)$. Asymptotic normality of $\hat{\theta}$ entails that

$$\sup_x \left| \mathbb{P}\left[\frac{\hat{\theta}-\theta_P}{\sigma(\hat{\theta})} \leq x \right] - \Phi(x) \right| = o(1), \tag{1.3}$$

where Φ denotes the standard normal distribution function and $\sigma^2(\hat{\theta})$ the variance of $\hat{\theta}$. When $\psi(X_1,\theta)$ is continuously differentiable and its third moment is finite, a Berry-Esséen type of theorem strengthens (1.3) by bounding the error of the normal approximation to

$$\sup_x \left| \mathbb{P}\left[\frac{\hat{\theta}-\theta_P}{\sigma(\hat{\theta})} \leq x \right] - \Phi(x) \right| = O(n^{-\frac{1}{2}}). \tag{1.4}$$

The bootstrap provides an alternative approximation to the sampling distribution of $\hat{\theta}$. Arcones & Giné (1992) have shown consistency of the bootstrap approximation for the limiting distribution of the standardized M-estimator under essentially the same conditions as those needed for proving its asymptotic normality. They rely upon empirical process theory: stochastic equicontinuity of the process $\sqrt{n}\int \psi(x,\theta)\,d(P_n-P)(x)$ and continuous differentiability (in θ) of $\int \psi(x,\theta)\,dP(x)$. However, it does not follow from their work that the bootstrap outperforms the classical normal approximation, as this requires a careful analysis of the second order properties of the estimator.

Lahiri (1992a) and Hall & Horowitz (1996) investigate the asymptotic refinements of the bootstrap distribution to the normal approximation of the standardized and studentized M-estimators, respectively. It is well known that without studentization, the improved accuracy of the bootstrap distribution to order $o(n^{-\frac{1}{2}})$ does not hold true [see Helmers (1991) and Hall (1995)]. In both cases, they assume that $\psi(x,\theta)$ has three continuous derivatives in θ to conduct their proof and use the implicit function theorem to establish the existence of an Edgeworth expansion for both $\hat{\theta}$ and its bootstrapped counterpart θ^*. Their results are useful to establish the second order correctness of the bootstrap.

The bootstrap distribution is typically approximated by Monte Carlo simulations. In large parameter spaces Θ this becomes computationally intensive, as large systems of equations have to be solved. Empirical Edgeworth expansions provide an attractive alternative to the bootstrap in these instances. But this requires an explicit expression for the Edgeworth expansion. Early explicit Edgeworth expansions for normalized M-estimators were derived by Pfanzagl (1973), under the assumptions that $\psi(x, \theta)$ is twice differentiable in θ and P belongs to a somewhat restricted parametric family. For the special case of Edgeworth expansions of location parameters, we refer to Bickel (1974) and references therein. The case of general order Edgeworth expansion for normalized M-estimators of a regression parameter is explored in Lahiri (1992b). The latter paper is particularly interesting in that the derivation of the expansion relies on assumed montonicity properties of ψ, instead of smoothness assumptions.

Lahiri (1994) rigorously establishes an explicit second order expansion for multivariate M-estimators that holds generally, provided that $\psi(x, \theta)$ is three times continuously differentiable in θ. He proceeds by first approximating the M-estimator $\hat{\theta}_n$ by a U-statistic U_n for which the tail of the distribution of the difference $n(\hat{\theta}_n - U_n)$ is of order $o(n^{-1/2})$. The second order properties for the estimator then follow from an Edgeworth expansion for U-statistics originated in Götze (1987), see also Bickel, Götze & van Zwet (1986). Armed with these results, Lahiri concludes that both the bootstrap and the empirical Edgeworth approximation provide second order corrections to the sampling distribution of M-estimators. The main contribution of this paper is a relaxation of the smoothness conditions on ψ required to establish the second order properties of M-estimators. Essentially, we require two continuous derivatives on ψ instead of three. Our approach refines Lahiri's argument and derives under "non standard" conditions a different approximating U-statistic U_n to the M-estimator. The technical aspects of this paper focus on proving rigourously that tails of the distribution of our remainder is appropriately small, i.e.,

$$\mathcal{P}[n|\hat{\theta}_n - U_n| > (\log n)^{-1}] = o(n^{-1/2}).$$

Our results rely implicitly on the smoothness of the parent distribution of X which does not hold for the empirical distribution. Hence our refinements can not be applied to derive an analogous expansion for the bootstrap estimate. But Lahiri's (1994) expansion, which requires more smoothness on ψ, still holds. The fact that more smoothness on ψ is needed for the bootstrap is not an artifact from our method of proof since it is known that the naive bootstrap performs worse in certain non-smooth cases. For example, Arcones (1995) shows a similar phenomenon for U-quantiles. From this we conclude that the empirical Edgeworth expansion improves upon the normal approximation more generally than the bootstrap.

Our paper is organized as follows. Section 2 contains the main assumptions and derives an approximation of the studentized M-estimator by a U-statistic. Section 3 presents our main results: an Edgeworth expansion for the studentized M-estimator, an empirical Edgeworth expansion and second order correctness of the bootstrap.

2. Preliminaries.

We begin this section by stating the assumptions used throughout this paper.

(A1) Assume that θ_P is an interior point of the parameter space Θ. Let $\widehat{\theta}$ be a consistent estimator of θ_P such that

$$\mathbb{P}\left\{\left|\sqrt{n}(\widehat{\theta}-\theta_P)\right| \geq \log(n)\right\} = o(n^{-1/2}). \tag{2.1}$$

(A2) The function $\psi(x,\theta)$ is twice differentiable in θ. The partial derivatives are denoted by $\Delta_k(x,\theta) = \partial^k\psi(x,\theta)/\partial\theta^k$, $k=1,2$. In addition, assume that $\Delta_2(x,\theta)$ is Hölder continuous at θ_P, *i.e.*,

$$|\Delta_2(x,\theta) - \Delta_2(x,\theta_P)| \leq r(x)|\theta-\theta_P|^\alpha, \tag{2.2}$$

for some $\alpha > 0$, and function $r(x)$ with

$$\mathbb{E}|r(X)|^k < \infty, \qquad \text{for some } k > 1/\alpha.$$

(A3) The expectations $\mathbb{E}|\Delta_k(X,\theta_P)|^2$ are finite, $k=1,2$.

(A4) The expectation $\mathbb{E}\psi(X_1,\theta_P)^6 < \infty$, and $\mathbb{E}\Delta_1(X,\theta_P) \neq 0$.

(A5) $|\mathbb{E}e^{it\psi(X,\theta_P)}| < 1-\chi(t)$ for all t, for some positive continuous function $\chi(t)$ that satisfies $\lim_{t\longrightarrow\infty}\chi(t) > 0$.

We briefly comment on the above assumptions.

(1) Assumption A1, while a little more stringent than assuming the consistency of $\hat{\theta}$ as in Arcones and Giné (1992), Giné (1997), Pollard (1984), and, Van der Vaart and Wellner (1996), makes our proofs much more transparent. It is also a key requirement in Lahiri (1994), who establishes (2.1) under suitable smoothness conditions on ψ, cf. Lahiri (1994, Proposition 2.1, p. 205).

(2) Assumptions A2 and A3 are required for the Edgeworth expansions of the bootstrap estimator but can otherwise be relaxed (see Theorem 2.2 below) to requiring slightly more than one derivative of $\psi(X,\theta)$, together with an entropy condition.

(3) Assumption A4 requires four moments for ψ instead of the customary three moments required for the Edgeworth expansions. The reason is that we need to control the size of tail probability of $\mathbb{P}[n^{-1}\sum_{i=1}^n(\psi(X_i)^3 - \mathbb{E}\psi^3(X)) > (\log n)^{-1}]$. Assumption A3 and A4 also imply bounds for the tail probability of $\mathbb{P}[n^{-1}\sum_{i=1}^n(\psi(X_i,\theta_P)\Delta_1(X_i,\theta_P) - \mathbb{E}\psi(X,\theta_P)\Delta(X,\theta_P)) > (\log n)^{-1}]$.

(4) Assumption A5 is a standard condition needed for the Edgeworth expansion of the U-statistics.

2.1. **A Stochastic Approximation for M-estimators.** Before stating our results, we introduce some additional notation. Define

$$\begin{aligned}
\varphi(\theta) &\equiv \varphi(\theta;P) = \int \psi(x,\theta)dP(x), \qquad \varphi_n(\theta) = \varphi(\theta;P_n) = \frac{1}{n}\sum_{i=1}^n \psi(X_i,\theta),\\
\varphi'(\theta) &\equiv \varphi'(\theta;P) = \int \Delta_1(x,\theta)dP(x), \qquad \varphi_n'(\theta) = \varphi'(\theta;P_n) = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}\psi(X_i,\theta),\\
\gamma(\theta) &\equiv \gamma(\theta;P) = \int \psi^2(x,\theta)\,dP(x), \qquad \gamma_n(\theta) = \gamma(\theta,P_n) = \frac{1}{n}\sum_{i=1}^n \psi^2(X_i,\theta).
\end{aligned}$$

The main difficulty in the following results is to guarantee that the tail probability of the remainder term in the stochastic expansions is of order $o(n^{-1/2})$.

Theorem 2.1. *Let conditions A1-A4 be satisfied. Then the expansion*

$$\hat\theta - \theta_P = -\frac{\varphi_n(\theta_P)}{\varphi'(\theta_P)}\left(1 + \frac{\varphi_n(\theta_P)\varphi''(\theta_P)}{2[\varphi'(\theta_P)]^2} - \frac{\varphi_n'(\theta_P) - \varphi'(\theta_P)}{\varphi'(\theta_P)}\right) + \xi_n, \tag{2.3}$$

holds true, where the remainder satisfies $\mathbb{P}\{|\xi_n| > (n\log n)^{-1}\} = o(n^{-\frac{1}{2}})$.

Proof. Assume without loss of generality that $\theta_P = 0$ and remark that by definition,

$$\varphi(0) = 0 \quad \text{and} \quad \varphi_n(\hat\theta) = o(n^{-1}).$$

Denote the centered random variables $\varphi_n'(\theta) - \varphi'(\theta)$ by $D_n(\theta)$. From Assumption A3,

$$\varphi(\hat\theta) = \hat\theta\varphi'(0) + \frac{1}{2}\hat\theta^2\varphi''(0) + \frac{1}{2}\hat\theta^2\left[\varphi''(\tilde\theta) - \varphi''(0)\right], \tag{2.4}$$

where $\tilde\theta$ is between 0 and $\hat\theta$. The left hand side of (2.4) equals

$$\varphi(\hat\theta) = -(\varphi_n - \varphi)(\hat\theta) = -(\varphi_n - \varphi)(0) - \hat\theta D_n(0) - \hat\theta\left(D_n(\overline{\theta}) - D_n(0)\right), \tag{2.5}$$

with $\overline{\theta}$ between 0 and $\widehat{\theta}$. Combining these two expressions yields

$$\hat{\theta} = -\frac{\varphi_n(0)}{\varphi'(0)} + \frac{R_n}{\varphi'(0)},$$

where $R_n = R_{n1} + R_{n2} + R_{n3} + R_{n4}$ with

$$R_{n1} = -\frac{1}{2}\hat{\theta}^2\varphi''(0), \qquad R_{n2} = -\frac{1}{2}\hat{\theta}^2\left[\varphi''(\tilde{\theta}) - \varphi''(0)\right],$$
$$R_{n3} = -\hat{\theta}D_n(0), \qquad R_{n4} = -\hat{\theta}\left(D_n(\overline{\theta}) - D_n(0)\right).$$

We begin by showing that

$$\mathbb{P}\left\{|R_n| > n^{-11/16}\right\} = o(n^{-1/2}). \tag{2.6}$$

First, condition A1 implies that

$$\mathbb{P}\left\{\hat{\theta}^2 \geq \frac{\log^2 n}{n}\right\} \leq \mathbb{P}\left\{(\sqrt{n}\hat{\theta})^2 \geq \log^2 n\right\} = o(n^{-1/2}). \tag{2.7}$$

Second, $|R_{n2}| \leq C|\hat{\theta}|^{2+\alpha}/2$ and

$$\mathbb{P}\left\{|\hat{\theta}|^{2+\alpha} > \frac{1}{n\log n}\right\} \leq \mathbb{P}\left\{|\sqrt{n}\hat{\theta}|^{2+\alpha} > \frac{n^{\frac{\alpha}{2}}}{\log n}\right\} = o(n^{-\frac{1}{2}}) \tag{2.8}$$

entail that $\mathbb{P}\{|R_{n,2}| > (n\log n)^{-1}\} = o(n^{-\frac{1}{2}})$. Third, we split

$$\begin{aligned}
\mathbb{P}\left\{|R_{n3}| > n^{-11/16}\right\} &\leq \mathbb{P}\left\{|\sqrt{n}\hat{\theta}| > \log n\right\} + \mathbb{P}\left\{|\sqrt{n}D_n(0)| > \frac{n^{5/16}}{\log n}\right\} \\
&\leq o(n^{-\frac{1}{2}}) + \frac{(\log n)^2}{n^{10/16}}\mathbb{E}|\sqrt{n}D_n(0)|^2 \\
&= o(n^{-\frac{1}{2}})
\end{aligned}$$

by A1 and the fact that $\mathbb{E}|\sqrt{n}D_n(0)|^2 < \infty$ by A3. Fourth, using a Taylor expansion and the Lipschitz condition A2, we bound

$$|R_{n4}| \leq |\widehat{\theta}|^2|\sqrt{n}(P_n - P)\Delta_2(\cdot,0)| + |\widehat{\theta}|^{2+\alpha}(P_n r + Pr),$$

and it follows that

$$\begin{aligned}
&\mathbb{P}\left\{|R_{n4}| > (n\log n)^{-1}\right\} \qquad (2.9)\\
&\le \mathbb{P}\left\{|\sqrt{n}\hat{\theta}|^2 > \frac{1}{2}(\log n)^2\right\} + \mathbb{P}\left\{|\sqrt{n}(P_n - P)\Delta_2(\cdot,0)| > \frac{n^{\frac{1}{2}}}{(\log n)^3}\right\}\\
&\quad +\mathbb{P}\left\{|\sqrt{n}\hat{\theta}|^{2+\alpha} > \frac{1}{2}(\log n)^{2+\alpha}\right\} + \mathbb{P}\left\{P_n r + Pr \ge \frac{n^{\alpha/2}}{(\log n)^{3+\alpha}}\right\}\\
&\le o(n^{-\frac{1}{2}}) + \frac{\log^6 n}{n}\mathbb{E}|\sqrt{n}(P_n - P)\Delta_2(\cdot,0)|^2 + \frac{(\log n)^{3p+p\alpha}}{n^{p\alpha/2}}\mathbb{E}|P_n r + Pr|^p.\\
&= o(n^{-\frac{1}{2}})
\end{aligned}$$

by assumptions A1, A2 and A3.

In view of the preceding calculations, only R_{n1} and R_{n3} need further refinements. For this purpose, we write

$$R_{n1} = -\frac{1}{2}\hat{\theta}^2\varphi''(0) = -\frac{1}{2}\varphi''(0)\left[\frac{-\varphi_n(0)+R_n}{\varphi'(0)}\right]^2 = -\frac{1}{2}\frac{\varphi''(0)}{[\varphi'(0)]^2}\varphi_n^2(0) + \tilde{R}_n,$$

where

$$\tilde{R}_n = \frac{\varphi''(0)}{2[\varphi'(0)]^2}\left(2\varphi_n(0)R_n - R_n^2\right).$$

By virtue of (2.6), it is easy to show that

$$\begin{aligned}
\mathbb{P}\{|\tilde{R}_n| > (n\log n)^{-1}\} &\le \sum_{i=1}^{4}\mathbb{P}\{|R_{ni}\varphi_n(0)| > (8n\log n)^{-1}\} + \mathbb{P}\{R_n^2 > (2n\log n)^{-1}\}\\
&= o(n^{-1/2})
\end{aligned}$$

as $\mathbb{P}\{|R_n| > (2n\log n)^{-1/2}\} = o(n^{-1/2})$ by the reasoning above, and since, for $1 \le i \le 4$, $\mathbb{P}\{|R_{ni}\varphi_n(0)| > (8n\log n)^{-1}\} = o(n^{-1/2})$. For instance,

$$\begin{aligned}
&\mathbb{P}\left\{\left|\hat{\theta}D_n(0)\varphi_n(0)\right| > \frac{1}{n\log n}\right\} \le \mathbb{P}\left\{\left|\sqrt{n}\hat{\theta}\right| > \log^2 n\right\} + \qquad (2.10)\\
&\quad +\mathbb{P}\left\{|\sqrt{n}D_n(0)| > n^{\frac{5}{16}}\right\} + \mathbb{P}\left\{|\sqrt{n}\varphi_n(0)| > \frac{n^{\frac{3}{16}}}{(\log n)^3}\right\} = o(n^{-\frac{1}{2}}).
\end{aligned}$$

Here we used the fact that

$$\mathbb{E}|\sqrt{n}\varphi_n(0)|^3 \lesssim n^{-3/2}\left\{\left(\sum_{i=1}^{n}\mathbb{E}\psi(X_i)^2\right)^{3/2} + \sum_{i=1}^{n}\mathbb{E}|\psi(X_i)|^3\right\}$$

by Rosenthal's inequality [cf. Ibragimov and Sharakhmetov (1998)] and A3. Similarly one can show that

$$R_{n3} = -D_n(0)\left[-\frac{\varphi_n(0)}{\varphi'(0)} + R_n\right] = D_n(0)\frac{\varphi_n(0)}{\varphi'(0)} + \overline{R}_n,$$

with $\mathbb{P}\{|\overline{R}_n| > (n\log n)^{-1}\} = o(n^{-\frac{1}{2}})$. Hence

$$\hat{\theta} = -\frac{\varphi_n(0)}{\varphi'(0)}\left(1 + \frac{\varphi_n(0)\varphi''(0)}{2[\varphi'(0)]^2} - \frac{D_n(0)}{\varphi'(0)}\right) + \xi_n, \tag{2.11}$$

with $\mathbb{P}\{|\xi_n| > (n\log n)^{-1}\} = o(n^{-\frac{1}{2}})$ as asserted. □

We may relax assumptions A2 and A3 as we do not need a second partial derivative of ψ if we require instead some regularity conditions on the first partial derivative. Define the class

$$\mathcal{F}_n = \left\{\Delta_1(\cdot,\theta) - \Delta_1(\cdot,\theta_P) : |\theta - \theta_P| \le \frac{\log n}{\sqrt{n}}\right\}$$

and, for each $\delta > 0$, let $H_n(\delta)$ be the δ-entropy with bracketing number of $\mathcal{F}_n$ in $L^2(P)$ (cf. Van der Vaart and Wellner 1996, p. 83, for its definition). We replace Assumptions (A2) and (A3) by

(A2') The function $\psi(x,\theta)$ is differentiable in θ with partial derivative $\Delta_1(x,\theta) = \partial\psi(x,\theta)/\partial\theta$.
In addition, assume that $H_n(\delta) \lesssim \delta^{-V}$ for some $V < 2$ and

$$|\Delta_1(x,\theta) - \Delta_1(x,\theta_P)| \le |\theta - \theta_P|^\alpha r(x)$$

with $\mathbb{E}|r(X)|^p < \infty$ for $p > \max(2/\{\alpha(2-V)\}, 3)$.
The function $\varphi(\theta) = \mathbb{E}\psi(X,\theta)$ is twice continuously differentiable in θ.
(A3') $\mathbb{E}|\Delta_1(X,\theta_P)|^2 < \infty$.

Theorem 2.2. *Let conditions A1, A4, A2' and A3' be satisfied. Then the expansion*

$$\hat{\theta} - \theta_P = -\frac{\varphi_n(\theta_P)}{\varphi'(\theta_P)}\left(1 + \frac{\varphi_n(\theta_P)\varphi''(\theta_P)}{2[\varphi'(\theta_P)]^2} - \frac{\varphi_n'(\theta_P) - \varphi'(\theta_P)}{\varphi'(\theta_P)}\right) + \xi_n, \tag{2.12}$$

holds true, where the remainder satisfies $\mathbb{P}\{|\xi_n| > (n\log n)^{-1}\} = o(n^{-\frac{1}{2}})$.

Proof. Recall that $D_n(\theta) = \varphi_n'(\theta) - \varphi'(\theta)$. The proof of the theorem is essentially the same as that of Theorem 2.1, except for the bound of the term

$$R_{n4} = -\hat{\theta}\left(D_n(\bar{\theta}) - D_n(0)\right).$$

We have

$$\begin{aligned}
\mathbb{P}\{|R_{n4}| &> (n\log n)^{-1}\} \\
&\le \mathbb{P}\left\{|\sqrt{n}\hat{\theta}| > \log n\right\} + \mathbb{P}\left\{\sqrt{n}|D_n(\bar{\theta}) - D_n(0)| > (\log n)^{-2}\right\} \\
&\le 2\mathbb{P}\left\{|\sqrt{n}\hat{\theta}| > \log n\right\} + \mathbb{P}\left\{\sup_{f\in\mathcal{F}_n} |\sqrt{n}(P_n - P)f| > (\log n)^{-2}\right\} \\
&\le o(n^{-1/2}) + (\log n)^{2p}\, \mathbb{E}\left(\sup_{f\in\mathcal{F}_n} |\sqrt{n}(P_n - P)f|\right)^p
\end{aligned}$$

by A1. Notice that the class $\mathcal{F}_n(x)$ has envelope $F_n = (n^{-1/2}\log n)^{\alpha} r(x)$. The second term can be handled as follows: Let $a_n \lesssim b_n$ mean that there exists a positive constant c such that $a_n \le c b_n$ for all n. Then

$$\begin{aligned}
&\mathbb{E}\left(\sup_{f\in\mathcal{F}_n} |\sqrt{n}(P_n - P)f|\right)^p \\
&\lesssim \left(\mathbb{E}\sup_{f\in\mathcal{F}_n} |\sqrt{n}(P_n - P)f|\right)^p + n^{-\frac{1}{2}+\frac{1}{p}}\mathbb{E}|F_n(X)|^p \\
&\qquad \text{by Theorem 2.14.5 in Van der Vaart and Wellner (1996, p. 244)} \\
&\lesssim \left(\mathbb{E}|F_n(X)|^2\right)^{p/2}\left(\int_0^1 H_n^{1/2}(\varepsilon(\mathbb{E}|F_n(X)|^2)^{1/2})\,d\varepsilon\right)^p + n^{-\frac{p}{2}+1}\mathbb{E}|F_n(X)|^p \\
&\qquad \text{by Theorem 2.14.2 in Van der Vaart and Wellner (1996, p. 240)} \\
&\lesssim \left(\frac{\log n}{\sqrt{n}}\right)^{\frac{p\alpha}{2}(2-V)} + n^{-\frac{p}{2}+1}\left(\frac{\log n}{\sqrt{n}}\right)^{\alpha p}
\end{aligned}$$

for p satisfying A2'. Observe that this choice yields

$$\mathbb{P}\{|R_{n4}| > (n\log n)^{-1}\} = o(n^{-1/2}).$$

This concludes the proof. □

In applications, the variance of the estimator $\hat{\theta}$ is typically unknown, and thus theorems 2.1 and 2.2 are of theoretical interest only. In practice, one usually considers the studentized M-estimator

$$T_n \equiv \sqrt{n\gamma_n^{-1}(\hat{\theta})}\,\varphi_n'(\hat{\theta})\left(\hat{\theta}-\theta_P\right). \tag{2.13}$$

Our next theorem develops a stochastic approximation for the studentized M-estimator.

Theorem 2.3. *Define the constants*

$$c_0 = 2\ \mathbb{E}\Delta_1(X_1,\theta_P)\psi(X_1,\theta_P), \tag{2.14}$$

$$c_1 = \frac{c_0}{2\varphi'(\theta_P)\gamma(\theta_P)} - \frac{\varphi''(\theta_P)}{2[\varphi'(\theta_P)]^2},\ \textit{and}\ c_2 = -\frac{1}{2\gamma(\theta_P)}.$$

Assume that A1-A4 are satisfied. Then

$$T_n = -\sqrt{n\gamma^{-1}(\theta_P)}\varphi_n(\theta_P)\left[1+c_1\varphi_n(\theta_P)+c_2(\gamma_n(\theta_P)-\gamma(\theta_P))\right]+\xi_n, \tag{2.15}$$

where $\mathbb{P}\left\{|\xi_n| > (n^{1/2}\log n)^{-1}\right\} = o(n^{-\frac{1}{2}})$.

Proof: Without loss of generality, $\theta_P = 0$. Recall the definition $D_n(\theta) = \varphi_n'(\theta) - \varphi'(\theta)$, and expand

$$\varphi_n'(\hat{\theta}) = \varphi'(0) + D_n(0) + \hat{\theta}\cdot\varphi''(0) + R_{n5}, \tag{2.16}$$

with remainder

$$R_{n5} = D_n(\hat{\theta}) - D_n(0) + \hat{\theta}\left(\varphi''(\overline{\theta}) - \varphi''(0)\right)$$

for some $\overline{\theta}$ between 0 and $\hat{\theta}$. The same argument used for the remainder R_{n2} in Theorem 2.1 yields $\mathbb{P}[\hat{\theta}\cdot R_{n5} > (n\log n)^{-1}] = o(n^{-1/2})$. Assumption A2 entails us to write

$$\begin{aligned}\gamma_n(\hat{\theta}) - \gamma(0) &= \gamma_n(\hat{\theta}) - \gamma_n(0) + (\gamma_n-\gamma)(0)\\ &= 2\hat{\theta}\frac{1}{n}\sum_{i=1}^{n}\psi(X_i,0)\Delta_1(X_i,0) + (\gamma_n-\gamma)(0) + R_{n6},\end{aligned}$$

in which, for some $\overline{\theta}$ between 0 and $\hat{\theta}$,

$$R_{n6} = 2\hat{\theta}\cdot\frac{1}{n}\sum_{j=1}^{n}[\psi(X_i,\overline{\theta})\Delta(X_i,\overline{\theta}) - \psi(X_i,0)\Delta_1(X_i,0)]. \tag{2.17}$$

Similarly to R_{n5}, we also have that $\mathbb{P}\left[\hat{\theta}R_{n6} > (n\log n)^{-1}\right] = o(n^{-1/2})$. Substitution in (2.17) of the average $n^{-1}\sum_{j=1}^{n}\psi(X_i,0)\Delta_1(X_i,0)$ by its expected value $c_0 = \mathbb{E}[\psi(X_i,0)\Delta_1(X_i,0)] = \gamma'(0)$, results in the stochastic approximation

$$\gamma_n(\hat{\theta}) - \gamma(0) = 2c_0\hat{\theta} + (\gamma_n-\gamma)(0) + R_{n7}. \tag{2.18}$$

Using the techniques used for R_{n5} together with an application of Chebychev's inequality and Assumption A4 yields

$$\mathbb{P}\left[\hat{\theta} \cdot R_{n7} > \frac{1}{n \log n}\right] = o(n^{-1/2}).$$

In light of (2.17), (2.18), and the Taylor expansion for $x^{-1/2}$, it follows that

$$\frac{\varphi_n'(\hat{\theta})}{\sqrt{\gamma_n(\hat{\theta})}} = \frac{1}{\sqrt{\gamma(0)}}\left(\varphi'(0) + D_n(0) + \hat{\theta}\left(\varphi''(0) - \frac{c_0\varphi'(0)}{2\gamma(0)}\right) - \frac{\varphi'(0)(\gamma_n - \gamma)(0)}{2\gamma(0)}\right) + R_{n8}$$

with $\mathbb{P}[\hat{\theta} R_{n8} > (n \log n)^{-1}] = o(n^{-1/2})$. Using the above expression together with the expansion (2.12) derived in Theorem 2.1, leads after algebraic manipulations to

$$\begin{aligned} T_n &= \sqrt{n} \cdot \frac{\varphi_n'(\hat{\theta})}{\sqrt{\gamma_n(\hat{\theta})}} \cdot \hat{\theta} \\ &= -\sqrt{n}\frac{\varphi_n(0)}{\sqrt{\gamma(0)}}\left(1 + \varphi_n(0)\left(\frac{c_0}{2\gamma(0)\varphi'(0)} - \frac{\varphi''(0)}{2\varphi'(0)^2}\right)\right. \\ &\quad \left. - \frac{(\gamma_n - \gamma)(0)}{2\gamma(0)}\right) + R_{n9}, \end{aligned} \tag{2.19}$$

in which the remainder satisfies $\mathbb{P}[R_{n9} > (n^{1/2} \log n)^{-1}] = o(n^{-1/2})$. Setting c_1 and c_2 as in (2.14) leads to the desired conclusion. □

Remark 2.4. *Again, we could weaken the assumptions slightly, and proceed to show that the remainder is small using the techniques used in Theorem 2.2.*

From Expansion (2.15) it is easily verified that the studentized M-estimator T_n can be decomposed into a stochastic and deterministic part, the random part being a U-statistic of degree two with zero mean as alluded to in the introduction.

Corollary 2.5. *Under the conditions A1-A4, the studentized M-estimator admits the following decomposition:*

$$T_n = \sqrt{n}\binom{n}{2}^{-1} \sum\sum_{1 \le i < j \le n} K(X_i, X_j) + \frac{B}{\sqrt{n}} + \xi_n, \tag{2.20}$$

where $\mathbb{P}\{|\xi_n| > (\sqrt{n}\log n)^{-1}\} = o(n^{-1/2})$, *where the kernel* $K(x,y)$ *is defined by*

$$K(x,y) = \left(\frac{-\psi(x,\theta_P)}{\sqrt{\gamma(\theta_P)}}\left\{1 + c_1\psi(y,\theta_P) + c_2[\psi^2(y,\theta_P) - \gamma(\theta_P)]\right\}\right)^{sym}, \qquad (2.21)$$

in which $h(x,y)^{sym} = 2^{-1}h(x,y) + 2^{-1}h(y,x)$ *is the symmetrization of the function* $h(x,y)$, *and*

$$B = \mathbb{E}K(X_1, X_1). \qquad (2.22)$$

Proof: Set

$$h(x,y) = \frac{-\psi(X_i,\theta_P)}{\sqrt{\gamma(0)}}\left[1 + c_1\psi(X_j,\theta_P) + c_2(\psi^2(X_j,\theta_P) - \gamma(\theta_P))\right],$$

and notice that $\mathbb{E}h(X_i,X_j) = 0$ $(i \neq j)$ and $\mathbb{E}h(X,X) = B$. Apply Theorem 2.3 and expand to get

$$\begin{aligned}
T_n &= \frac{-\sqrt{n}\varphi_n(\theta_P)}{\sqrt{\gamma(\theta_P)}}\left[1 + c_1\varphi_n(\theta_P) + c_2(\gamma_n(\theta_P) - \gamma(\theta_P))\right] + \xi_n, \\
&= \frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{-\psi(X_i,\theta_P)}{\sqrt{\gamma(\theta_P)}}\left[1 + c_1\psi(X_j,\theta_P) + c_2(\psi^2(X_j,\theta_P) - \gamma(\theta_P))\right] + \xi_n \\
&= \frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i,X_j) + \xi_n \\
&= \frac{\sqrt{n}}{n(n-1)}\sum\sum_{i\neq j} h(X_i,X_j) + \frac{\sqrt{n}}{n}B \\
&\quad + \frac{\sqrt{n}}{n^2}\sum_{i=1}^{n}\{h(X_i,X_i) - \mathbb{E}h(X_i,X_i)\} + \frac{\sqrt{n}}{n^2(n-1)}\sum\sum_{i\neq j} h(X_i,X_j) + \xi_n
\end{aligned}$$

where ξ_n is given in (2.15). By virtue of Chebychev's inequality and Assumption A4, one concludes that both

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^{n}(h(X_i,X_i) - \mathbb{E}h(X_i,X_i)\right| > \frac{1}{\log n}\right] &= o(n^{-1/2}) \\
\mathbb{P}\left[\frac{1}{n(n-1)}\left|\sum_{i\neq j} h(X_i,X_j)\right| > \frac{1}{\log n}\right] &= o(n^{-1/2}),
\end{aligned}$$

from which the conclusion follows. □

3. Main Results

Let $G_n(x) = \mathbb{P}(T_n \le x)$ be the distribution function of the studentized statistic T_n. With B defined in (2.22), let

$$\kappa_3 = -\frac{\mathbb{E}\psi^3(X_1, \theta_P) + 6\gamma(\theta_P)B}{\gamma(\theta_P)^{3/2}}. \tag{3.1}$$

Denote by Φ and ϕ the distribution function and density of a standard normal respectively. While $|G_n(x) - \Phi(x)| = O(n^{-\frac{1}{2}})$, Theorem 3.1 below shows that

$$\tilde{G}_n(x) = \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left\{\frac{\kappa_3}{6}(x^2 - 1) + B\right\} \tag{3.2}$$

approximates $G_n(x)$ to order $o(n^{-\frac{1}{2}})$.

Theorem 3.1 (Edgeworth expansion). *Under A1-A5*

$$\sup_x \left|G_n(x) - \tilde{G}_n(x)\right| = o(n^{-\frac{1}{2}}). \tag{3.3}$$

Proof: From Corollary 2.5, the studentized M-estimator can be written as

$$T_n = \sqrt{n}U_n + \frac{B}{\sqrt{n}} + \xi_n,$$

with

$$U_n = \binom{n}{2}^{-1} \sum\sum_{1\le i<j\le n} K(X_i, X_j).$$

and $B = \mathbb{E}K(X, X)$, to which we shall apply the Edgeworth expansion derived in Bickel, Götze & van Zwet (1986). Since $\mathbb{E}\psi(X, \theta_P) = 0$ and $\mathbb{E}\psi^2(X, \theta_P) - \gamma(\theta_P) = 0$, it follows that in the notation of Bickel et al. (1986),

$$g(x) = \mathbb{E}[K(X, Y)|X = x] = \frac{1}{2}\psi(x, \theta_P).$$

Our assumptions A4 and A5 guarantee that the conditions of their Theorem 1.2 are satisfied, which entails

$$\begin{aligned} G_n(x) &= \mathbb{P}\{T_n \le x\} \\ &\le \mathbb{P}\left\{\sqrt{n}U_n \le x - n^{-\frac{1}{2}}B - (n\log^2 n)^{-1/2}\right\} + \mathbb{P}\left\{|\xi_n| > (\sqrt{n}\log n)^{-1}\right\} \\ &= \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left\{\frac{\kappa_3}{6}(x^2 - 1) + B\right\} + o(n^{-\frac{1}{2}}), \end{aligned} \tag{3.4}$$

where κ_3 is given in (3.1) and the remainder ξ_n is defined in Corollary 2.5. In a similar fashion, one establishes

$$G_n(x) \geq \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left\{\frac{\kappa_3}{6}(x^2-1)+B\right\}+o(n^{-\frac{1}{2}}),$$

and the result follows. □

In practice B and κ_3 appearing in (3.2) are unknown and have to to be estimated from the data. We estimate them by

$$-\sqrt{\gamma_n(\hat{\theta})}\widehat{B} = \widehat{c}_1\gamma_n(\widehat{\theta}) + \frac{\widehat{c}_2}{n}\sum_{j=1}^{n}\psi^3(X_j,\widehat{\theta}) \tag{3.5}$$

and

$$\widehat{\kappa_3} = -\frac{n^{-1}\sum_{j=1}^{n}\psi^3(X_j,\widehat{\theta})+6\gamma_n(\widehat{\theta})\widehat{B}}{\gamma_n(\widehat{\theta})^{3/2}}, \tag{3.6}$$

where

$$\widehat{c}_0 = \gamma_n'(\hat{\theta}), \quad \widehat{c}_1 = \frac{-\varphi_n''(\hat{\theta})}{2[\varphi_n'(\hat{\theta})]^2} + \frac{\widehat{c}_0}{2\varphi_n'(\hat{\theta})\gamma_n(\hat{\theta})}, \quad \widehat{c}_2 = -\frac{1}{2\gamma_n(\hat{\theta})} \tag{3.7}$$

estimate the constants c_0, c_1 and c_2. The following theorem states that

$$\tilde{E}_n(x) = \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left\{\frac{\widehat{\kappa_3}}{6}(x^2-1)+\widehat{B}\right\}, \tag{3.8}$$

also approximates $G_n(x)$ to order $o(n^{-\frac{1}{2}})$.

Theorem 3.2 (Empirical Edgeworth expansion). *Suppose A1-A5 hold. Then we have with probability one*

$$\sup_x\left|G_n(x)-\tilde{E}_n(x)\right| = o(n^{-\frac{1}{2}}). \tag{3.9}$$

Proof: The theorem rests on showing that both $\widehat{\kappa_3}$ and $\widehat{B}$ converge with probability one to κ_3 and B, respectively. By the mean value theorem, there exists a $\tilde{\theta}$ between $\widehat{\theta}$ and θ_P such that

$$\gamma_n(\widehat{\theta}) - \gamma(\theta_P) = \gamma_n'(\tilde{\theta})\left(\widehat{\theta}-\theta_P\right) + (\gamma_n(\theta_P)-\gamma(\theta_P)) \xrightarrow{a.s.} 0,$$

and similarly,

$$\frac{1}{n}\sum_{j=1}^{n}\psi^3(X_j,\widehat{\theta}) \xrightarrow{a.s.} \int \psi^3(x,\theta_P)\,dP(x), \qquad \widehat{c_i} \xrightarrow{a.s.} c_i,\ i=1,\ldots,3.$$

Therefore $\widehat{B} \xrightarrow{a.s.} B$ and $\widehat{\kappa_3} \xrightarrow{a.s.} \kappa_3$. The conclusion follows. □

Given $X_1,\ldots,X_n$, let $X_1^*,\ldots,X_n^*$ be the bootstrap sample drawn with replacement from P_n. Define P_n^* as the empirical probability measure based on $X_1^*,\ldots,X_n^*$, conditionally given the sample $X_1,\ldots,X_n$, and let $G_n^*(x)$ be the bootstrap distribution function of

$$T_n^* \equiv \sqrt{n(\gamma_n^*)^{-1}(\theta^*)}\,(\varphi_n^*)'(\theta^*)\left(\theta^* - \hat{\theta}\right). \tag{3.10}$$

The quantities γ_n^* and φ_n^* are the bootstrap equivalents of γ_n and φ_n, that is,

$$\gamma_n^*(\theta) = \int \psi^2(x,\theta)\,dP_n^*(x) \ \text{ and } \ \varphi_n^*(\theta) = \int \psi(x,\theta)\,dP_n^*(x).$$

We set out with a bootstrap version of Theorem 2.1. We replace Assumption A1 by its bootstrap counterpart.

(A1*) Let θ^* be a consistent estimator of $\widehat{\theta}$, and suppose that

$$\mathbb{P}^*\left[\left|\sqrt{n}(\theta^* - \widehat{\theta})\right| \geq \log(n)\right] = o(n^{-1/2}) \ \text{ a.s.} \tag{3.11}$$

Proposition 3.3. *Suppose* A1*, *A2 – A5 hold true. Then*

$$T_n^* = -\sqrt{n\gamma_n^{-1}(\hat{\theta})}\left(\varphi_n^*(\hat{\theta}) + \widehat{c}_1(\varphi_n^*)^2(\hat{\theta}) + \widehat{c}_2\varphi_n^*(\hat{\theta})(\gamma_n^* - \gamma_n)(\hat{\theta})\right) + \xi_n^*,$$

where $\mathbb{P}^*\left\{|\xi_n^*| > (\sqrt{n}\log n)^{-1}\right\} = o(n^{-\frac{1}{2}})$.

Proof: The proof is as the one given in full detail of Proposition 2.1. This is possible because ψ is smooth. □

Theorem 3.4 (Bootstrap). *Under conditions A1*, A1 – A5 we have with probability one that*

$$\sup_x |G_n(x) - G_n^*(x)| = o(n^{-\frac{1}{2}}). \tag{3.12}$$

Proof: The proof of Theorem 3.4 parallels the one of Theorem 3.1. Express

$$T_n^* = -\sqrt{n\gamma_n^{-1}(\hat{\theta})}\binom{n}{2}^{-1}\sum\sum_{1\le i<j\le n} K_n(X_i, X_j) + n^{-\frac{1}{2}}\widehat{B} + \xi_n^*, \tag{3.13}$$

where $\widehat{B}$ is defined in (3.5), and the kernel

$$K_n(x,y) = \left(-\psi(x,\hat{\theta})\left\{1 + \widehat{c}_1\psi(y,\hat{\theta}) + \widehat{c}_2[\psi^2(y,\hat{\theta}) - \gamma_n(\hat{\theta})]\right\}\right)^{sym} / \sqrt{\gamma_n(\hat{\theta})} \tag{3.14}$$

is an estimate of (2.21). Also, in perfect analogy with the previous result,

$$\mathbb{E}^* K_n(X_1^*, X_2^*) = 0, \quad \mathbb{E}^*(K_n(X_1^*, X_2^*)|X_1^* = x) = \psi(x,\hat{\theta})/\sqrt{\gamma_n(\hat{\theta})}.$$

The only non trivial fact yet to be established is that the distribution function of $\psi^*(X_1^*, \hat{\theta})$ is nonlattice. In view of the nonlatticeness of $\psi(X, \theta_P)$, it suffices to show that for each $M > 0$ [cf. Helmers (1991)],

$$\sup_{|t|\le M}\left|\mathbb{E}^* e^{it\psi(X_1^*,\hat{\theta})} - \mathbb{E}e^{it\psi(X_1,\theta_P)}\right| \to 0. \tag{3.15}$$

The latter follows from

$$\sup_{|t|\le M}\left|\mathbb{E}^* e^{it\psi(X_1^*,\hat{\theta})} - \mathbb{E}^* e^{it\psi(X_1^*,\theta_P)}\right| \le M\frac{1}{n}\sum_{i=1}^{n}\left|\psi(X_i,\hat{\theta}) - \psi(X_i,\theta_P)\right| \xrightarrow{a.s.} 0,$$

the smoothness of ψ, the strong consistency of $\hat{\theta}$, and Theorem 2.1 of Feuerberger and Mureika (1977) stating that

$$\sup_{|t|\le M}\left|\frac{1}{n}\sum_{i=1}^{n} e^{it\psi(X_i,\theta_P)} - \mathbb{E}e^{it\psi(X_1,\theta_P)}\right| \xrightarrow{a.s.} 0.$$

Arguing as in Theorem 3.1 we conclude that

$$G_n^*(x) = \Phi(x) - \frac{\phi(x)}{\sqrt{n}}\left(\frac{\widehat{\kappa}_3}{6}(x^2-1) + \widehat{B}\right) + o(n^{-\frac{1}{2}}), \tag{3.16}$$

whence $\sup_x |G_n(x) - G_n^*(x)| = o(n^{-\frac{1}{2}})$. □

Acknowledgements. We want to thank the referees for their constructive comments and bringing to our attention the paper Lahiri (1994) and the editor for his encouragments.

REFERENCES

[1] ARCONES M. (1995). The asymptotic accuracy of the bootstrap of U-quantiles. *Annals of Statistics* **23** 1802-1822.

[2] ARCONES, M. AND GINÉ, E. (1992). On the bootstrap of M-estimators and other statistical functionals. In *Exploring the limits of the bootstrap* (R. LePage and L. Billard eds.), pp. 13-48, Wiley.

[3] BICKEL, P.J. (1974). Edgeworth expansions in nonparametric statistics. *Annals of Statistics* **2** 1-20.

[4] BICKEL, P.J., GÖTZE, F. AND VAN ZWET, W.R. (1986). The Edgeworth expansion for U-statistics of degree two. *Annals of Statistics* **14** 1463-1484.

[5] CHOW, Y AND TEICHER, H. (1978). Probability theory, independence, interchangeability, Martingales. Springer-Verlag. New York. FEUERBERGER, A. AND MUREIKA R.A. (1977). The empirical characteristic function and its applications. *Annals of Statistics* **5** 88-97.

[6] GINÉ, E. (1997). *Lectures on some aspects of the bootstrap.* Lectures on probability and statistics (Saint-Flour, 1996), pages 37 – 151, Lecture Notes in Mathematics, **1665**, Springer.

[7] GÖTZE, F. (1987). Approximations for multivariate U-statistics. *Journal of Multivariate Analysis* **22** 212-229.

[8] HALL, P. (1995). The bootstrap and Edgeworth expansion. *Springer Verlag.* New-York.

[9] HALL, P. AND HOROWITZ J.L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica* **64** 891-916.

[10] HELMERS, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a studentized U-statistic. *Annals of Statistics* **19** 470-484.

[11] HUBER, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium.* Vol 1, pp221-234.

[12] IBRAGIMOV, R. AND SHARAKHMETOV, SH. (1998). *On an exact constant for the Rosenthal inequality.* Theory of Probability and its Applications **42**, 294–302.

[13] LAHIRI, S.N. (1992A). On bootstrapping M-estimators. *Sankhyã*, Series A **54** 157–170.

[14] LAHIRI, S.N. (1992B). Edgeworth expansions for M-estimators of a regression parameter. *Journal of Multivariate Analysis*, **43** 125-132.

[15] LAHIRI, S.N. (1994). On two-term Edgeworth expansions and bootstrap approximations for studentized multivariate M-estimators. *Sankhã*, Series A **56** 201–226.

[16] PFANZAGL, J. (1973). Asymptotic Expansions Related to Minimum Contrast Estimators. *Annals of Statistics* **1** 993-1026. Correction **2** 1357-1358.

[17] POLLARD, D. (1984). Convergence of empirical processes. *Springer Verlag.* New-York.

[18] VAN DER VAART, A. AND WELLNER, J. (1996). Weak convergence and empirical processes. Springer Verlag. Heidelberg.

NICOLAS HENGARTNER
D-1 STATISTICAL SCIENCE GROUP
LOS ALAMOS NATIONAL LABORATORY
LOS ALAMOS, NEW MEXICO
NM 87545
U.S.A.
nickckh@lanl.gov

MARTEN WEGKAMP
DEPARTMENT OF STATISTICS
FLORIDA STATE UNIVERSITY
TALLAHASSEE, FLORIDA
FL 32306–4330
U.S.A.
wegkamp@stat.fsu.edu