A NOTE ON THE UNIFORM ASYMPTOTIC NORMALITY OF LOCATION M-ESTIMATES

José R. Berrendero [1]

Departamento de Matemáticas

Universidad Autónoma de Madrid

28049 Madrid, Spain.

joser.berrendero@uam.es

Ruben H. Zamar [2]

Department of Statistics

University of British Columbia

V6T 1Z2 Vancouver B.C., Canada.

ruben@stat.ubc.ca

Running title: Uniform asymptotic normality

Second revision, September 2004

[1]Corresponding author. Research partially supported by Spanish Grant BFM2001-0169.

[2]Research partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Abstract

In the robustness framework, the parametric model underlying the data is usually embedded in a neighborhood of other plausible distributions. Accordingly, the asymptotic properties of robust estimates should be uniform over the whole set of possible models. In this paper, we study location M-estimates calculated with a previous generalized S-scale and show that, under some regularity conditions, they are uniformly asymptotically normal over contamination neighborhoods of known size. There is a trade off between the size of the neighborhood and the breakdown point of the GS-scale, but it is possible to adjust the estimates so that they have 50% breakdown point whereas the uniform asymptotic normality is ensured over neighborhoods that contain up to 25% of contamination. Alternatively, both the breakdown point and the size of the neighborhood could be chosen to be 38%. These results represent an improvement over those obtained recently by Salibian-Barrera and Zamar (2003).

Key words: M-estimates, U-statistics, GS-estimates, robust estimation.
MSC: 62F35.

1 Introduction

In the robustness framework, it is considered that a parametric model may not be totally adequate to characterize the distribution F underlying the data. Therefore, parametric models are usually embedded in a larger family of plausible distributions $\mathcal{F}$.

In this context, it would be highly desirable that most of the properties of robust estimates were uniform over $\mathcal{F}$. In particular, the asymptotic normality should be uniform, meaning that the distance between the normal distribution and the distribution of the estimate, for each sample size n, does not depend on the particular model $F \in \mathcal{F}$ but rather on general properties of the family $\mathcal{F}$. This requirement of uniformity was acknowledged in the beginnings of the theory [see, for instance, Hampel (1971) or Huber (1981), p. 51] but apparently it has not received in the literature as much attention as other questions. Still, there are a number of papers that deal with uniform asymptotics in different ways, differing in the class of estimates and the neighborhood $\mathcal{F}$ that they consider.

For instance Hampel (1971) studied location M-estimates over Prokhorov neighborhoods whereas Huber (1981) showed that when the scale is known, location M-estimates are uniformly asymptotically normal over the set of symmetric distributions that give probability zero to the points at which the estimating function is not differentiable. Clarke (1986) and Davies (1998) considered simultaneous estimation of location and scale. More complex models have been studied by Bednarski and Zontek (1996), and Bednarski and Müller (2001). Zielinski (1998) gave necessary and sufficient conditions on $\mathcal{F}$ for the uniform strong consistency of sample quantiles, and Berrendero (2003) studied the uniform strong consistency of the estimates that solve a generalized estimating equation. Most of these results establish the existence of a neighborhood over which the uniform asymptotic normality holds but are unable to assess its size. Recently, Salibian-Barrera and Zamar (2004) have studied the uniform asymptotic properties over a contamination neighborhood of a location M-estimate combined, to obtain scale-equivariance, with an S-scale (see Rousseeuw and Yohai, 1984). A nice feature of their results is that they are able to determine the size of the neighborhood over which the uniformity holds. They find a trade-off between this size and the breakdown point of the estimates. When the breakdown point is 50%, the maximum fraction of contamination under which the uniform asymptotic normality can be ensured is 11%. However, it is possible to tune the estimates so that both the breakdown point and the size of the maximum fraction of contamination are 25%.

In this article we obtain a result on uniform asymptotic normality (see Theorem 1 below) that is general in two directions: firstly, it is valid for general neighborhoods of distributions satisfying some regularity conditions and, secondly, it is valid for estimates defined as solutions of generalized estimating equations based on U-statistics. This second feature paves the way for improvements over the results obtained by Salibian-Barrera and Zamar (2004).

A difficulty with the approach of Salibian-Barrera and Zamar (2004) is that, when the underlying distribution is not symmetric, the asymptotic properties of the M-location depend on those of the S-scale which in turn depend on the corresponding S-location. This fact produce quite complicated formulae for the asymptotic variances and reduces the size of the neighborhoods over which the uniformity holds. To overcome these problems we propose here to use a generalized S-scale [hereafter GS-scale, see Croux, Rousseeuw and Hössjer (1994)] instead of an S-scale. For our purposes, the key feature of GS-scales is that they do not depend on a specific location since they are based exclusively on pairwise differences of data. This fact simplifies the relationship between the asymptotic distribution of the scale and the location estimates and also allows to obtain estimates with 50% breakdown point that are uniformly asymptotically normal over a neighborhood containing up to 25% of contamination. The estimates can alternatively be adjusted so that both the breakdown point and the size of the neighborhood are 38%.

The rest of this paper is organized as follows: in Section 2 we give precise definitions of the estimates included in our proposal and the uniform asymptotic properties we are going to establish. In Section 3 we give a general result on the uniform asymptotic normality of estimates that are solutions of a U-statistics equation. We particularize this result to GS-scales and contamination neighborhoods in Section 4. In Section 5 we derive the uniform asymptotic normality of an M-location combined with a GS-scale. Section 6 contains some conclusions and further remarks. A final appendix contains the proofs.

2 Definitions and technical preliminaries

Consider the location-scale model on the real line, that is, we observe a sample $x_1, \ldots, x_n$ satisfying

$$x_i = \mu_0 + \sigma_0 \epsilon_i, \quad i = 1, \ldots, n,$$

where $\epsilon_1, \ldots, \epsilon_n$ are i.i.d. random variables drawn from a distribution F. The main interest resides in estimating the location parameter μ_0. In this context, Huber (1964) defined a location M-estimate as the solution, $\hat{\mu}_n$, of the equation

$$\frac{1}{n} \sum_{i=1}^{n} \psi \left(\frac{x_i - \hat{\mu}_n}{\hat{\sigma}_n} \right) = 0, \tag{1}$$

for a score function ψ and a scale equivariant dispersion estimate $\hat{\sigma}_n$. For most of the score functions, using a scale equivariant estimate $\hat{\sigma}_n$ is essential to obtain a scale-equivariant location estimate $\hat{\mu}_n$.

When the underlying distribution of the errors F is symmetric, it can be shown under mild conditions that the asymptotic distribution of $\hat{\mu}_n$ depends only on the choice of $\hat{\sigma}_n$ through its asymptotic value, denoted by $\sigma(F)$. Contaminated real data are often generated by complex mechanisms which cannot be well described by a simple symmetric probability model. Models with different levels of complexity have been proposed to analyze the properties of robust estimates in the presence of contamination. Contamination neighborhoods, introduced by Tukey (1960), have the virtue of providing an approximate model that is mathematically tractable and easy to explain to non-statisticians. To allow for a fraction ϵ of outliers in the sample, we assume that F is any of the distributions lying in the ϵ-contamination neighborhood

$$\mathcal{F}(\epsilon, F_0) = \{F: \ F = (1-\epsilon)F_0 + \epsilon H, \ H \text{ arbitrary distribution}\}, \tag{2}$$

where $0 < \epsilon < 1/2$ and F_0 is a specified distribution with mean equal to zero and variance equal to one. Under this assumption, the distribution F can be asymmetric. In this case, the asymptotic distribution of $\hat{\mu}_n$ is more involved and depends heavily on the choice of $\hat{\sigma}_n$ [see Huber (1964, 1981)].

As we have said in the introduction, we propose to select $\hat{\sigma}_n$ in the class of GS-scales. These estimates were defined by Croux, Rousseeuw and Hössjer (1994) for regression estimation. A GS-scale is defined as the solution $\hat{\sigma}_n$ of the equation

$$\binom{n}{2}^{-1} \sum_{i<j} \chi\left(\frac{x_i - x_j}{\hat{\sigma}_n}\right) = b, \tag{3}$$

for a score function χ and a constant b. Hössjer, Croux and Rousseeuw (1994) studied some non-uniform asymptotic properties of regression and scale GS-estimates.

To conclude this section, we define uniform versions of the usual $o_p(1)$, $O_p(1)$ symbols and of the uniform asymptotic normality, following Salibian-Barrera and Zamar (2004). Given a family of distributions $\mathcal{F}$ and a sequence of random variables X_n, we say that $X_n = UO_p(1)$ over $\mathcal{F}$ if

$$\lim_{M\to\infty} \lim_{n\to\infty} \sup_{F\in\mathcal{F}} P_F\{|X_n| > M\} = 0.$$

We say that $X_n = Uo_p(1)$ over $\mathcal{F}$ if, for all $\delta > 0$,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} P_F\{|X_n| > \delta\} = 0.$$

Finally, we say that X_n is uniformly asymptotically normal (hereafter u.a.n.) over $\mathcal{F}$ if

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{x\in\mathbb{R}} |P_F\{X_n \leq x\} - \Phi(x)| = 0,$$

where Φ is the standard normal distribution function.

3 Uniform asymptotic normality and generalized estimating equations

From equation (3) we see that GS-scales are the solutions to an estimating equation defined by a U-statistics with kernel $\chi[(x_i - x_j)/s] - b$. A consequence of this fact is that the defining equation of location M-estimates (1) can also be approximately cast in terms of U-statistics. Therefore, it is useful for our purposes to have general results on the uniform asymptotic normality of estimates that solve a generalized estimating equation based on U-statistics. The goal of this section is to establish such results. The conclusions are also general in a way that they are not exclusively applicable to contamination neighborhoods. They cover arbitrary families of distributions $\mathcal{F}$ satisfying certain regularity assumptions.

Let $\widehat{\theta}_n$ be the solution of a generalized estimating equation $h_n(t) = 0$, where

$$h_n(t) \doteq \binom{n}{2}^{-1} \sum_{i<j} \eta(t; x_i, x_j)$$

is a U-statistic with a kernel η of degree 2. Define

$$H(t, F) \doteq \mathrm{E}_{F\times F}\eta(t; X, Y) \doteq \mathrm{E}_F\bar{\eta}(t; X, F),$$

where $\bar{\eta}(t; x, F) \doteq \mathrm{E}_F\eta(t; x, Y)$ and let $\theta = \theta(F)$ be the solution of the equation $H(t, F) = 0$. In Berrendero (2003) it is shown that, under some regularity conditions on η and the class of distributions $\mathcal{F}$, $\widehat{\theta}_n$ is uniformly consistent, that is, $\widehat{\theta}_n = \theta + Uo_p(1)$ over $\mathcal{F}$. Here, we will study the uniform asymptotic normality of $\widehat{\theta}_n$ given that the consistency and other regularity conditions hold. Basically, the kernel η has to be bounded and smooth and we must also ensure that the asymptotic variance is uniformly bounded away from zero. More precisely, we assume the following conditions:

A1. *The function* $\eta(\cdot, x, y)$ *is decreasing, bounded and twice continuously differentiable for all* $x, y \in \mathbb{R}$*.*

A2. *There exists* $0 < K < \infty$ *such that* $\sup_{F\in\mathcal{F}} |\eta'(\theta(F); x, y)| \leq K$*.*

A3. *There exists* $0 < M < \infty$ *and* $\beta > 0$ *such that* $|\eta''(t; x, y)| \leq M$ *when* $|t - \theta(F)| < \beta$ *for some* $F \in \mathcal{F}$*.*

A4. $\inf_{F\in\mathcal{F}} c(F) > 0$*, where* $c(F) \doteq -E_{F\times F}\eta'(\theta; X, Y) > 0$*.*

A5. $\inf_{F\in\mathcal{F}} \tilde{\sigma}(F) > 0$*, where* $\tilde{\sigma}^2(F) = Var_F[\bar{\eta}(\theta; X, F)]$*.*

A6. $\widehat{\theta}_n = \theta + Uo_p(1)$ *over* $\mathcal{F}$*.*

The assumptions above include different classes of estimates and neighborhoods of distributions. For instance, GS-scales are covered whenever the score function χ satisfies some usual regularity conditions (see **A7** below). The assumptions on $\mathcal{F}$ must be checked for each particular instance. The important case of contamination neighborhoods is addressed in Sections 4 and 5. On the other hand, if the family $\mathcal{F}$ is too large, the uniform asymptotic normality may fail to hold, at least for non-smooth estimates. For instance, when $\mathcal{F} = \{F : F \text{ is strictly increasing}\}$, then sample quantiles are not even uniformly consistent (see Zielinski (1998), example 1.)

We have the following general result:

Theorem 1. *Under* **A1** - **A6***,*

$$\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{\sqrt{n}h_n(\theta)}{c(F)} + Uo_p(1). \tag{4}$$

Furthermore, if $v(F) = [2\tilde{\sigma}(F)/c(F)]^2$*, then*

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} \sup_{x\in\mathbb{R}} \left| P_F\left\{ \frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sqrt{v(F)}} \leq x \right\} - \Phi(x) \right| = 0. \tag{5}$$

4 Uniform asymptotic normality for GS-scales

In this section we analyze the uniform asymptotic behavior of $\sqrt{n}(\hat{\sigma}_n - \sigma)$ over contamination neighborhoods $\mathcal{F}(\epsilon, F_0)$ [see (2)], where $\hat{\sigma}_n$ is a GS-scale [see (3)] and σ is implicitly defined as

$$\mathrm{E}_{F\times F}\chi\left(\frac{X-Y}{\sigma}\right) = b. \tag{6}$$

In this case, the kernel of the U-statistics is $\eta(s;x,y) = \chi[(x-y)/s] - b$, where

$$b = \mathrm{E}_{F_0\times F_0}\chi(X-Y) \tag{7}$$

for the estimate to be Fisher-consistent. The breakdown point of these estimates is given by $\epsilon^* = \min\{\sqrt{1-b}, 1-\sqrt{1-b}\}$. Then, for $\epsilon < \epsilon^*$, the explosion and implosion maxbias curves satisfy $S^+(\epsilon) = \sup_{F\in\mathcal{F}}\sigma(F) < \infty$ and $S^-(\epsilon) = \inf_{F\in\mathcal{F}}\sigma(F) > 0$. Condition **A7** on χ is sufficient to apply Theorem 1 above:

A 7. *$\chi(x)$ is even, bounded, non-decreasing in $(0,\infty)$, twice continuously differentiable with $\chi(0) = 0$ and $\chi(x) = 1$ for $|x| \geq k$, where $k > 0$ is a tuning constant.*

We have the following result:

Theorem 2. *Let $\mathcal{F} = \mathcal{F}(\epsilon, F_0)$, a contamination neighborhood centered at a distribution F_0 with an even, unimodal and strictly positive density. Let $\tilde{\epsilon}$ be such that $E_{F_0}\chi[X/S^-(\tilde{\epsilon})] = (b-\tilde{\epsilon})/(1-\tilde{\epsilon})$. Let $\hat{\sigma}_n$ be a GS-scale as defined in (3). Then, under* **A7**, *$\sqrt{n}(\hat{\sigma}_n - \sigma)/\sqrt{v(F)}$ is u.a.n. over $\mathcal{F}$, for all $\epsilon < \epsilon_0 \doteq \min\{\epsilon^*, \max\{\tilde{\epsilon}, 1-b\}\}$, where $v(F)$ is defined in Theorem 1.*

If we particularize the expression of the asymptotic variance to the case of GS-scales, we have that $v(F) = 4\tilde{\sigma}^2(F)/c^2(F)$, where $\tilde{\sigma}^2(F) = \mathrm{Var}_F[\bar{\chi}(X,F)]$, with

$$\bar{\chi}(x,F) = \mathrm{E}_F\chi\left(\frac{x-Y}{\sigma}\right) - b,$$

and

$$c(F) = \frac{1}{\sigma}\mathrm{E}_{F\times F}\left[\chi'\left(\frac{X-Y}{\sigma}\right)\left(\frac{X-Y}{\sigma}\right)\right]. \tag{8}$$

k	ϵ^*	$1-b$	$\tilde{\epsilon}$	ϵ_0
0.99	0.50	0.25	0.15	0.25
1.25	0.44	0.31	0.15	0.31
1.50	0.40	0.36	0.16	0.36
1.58	0.38	0.38	0.16	0.38

Table 1: Breakdown point ϵ^* and size ϵ_0 of the largest neighborhood over which a dispersion GS-estimate with Tukey's score function is uniformly asymptotically normal.

Tukey's score function, defined as $\chi_k(x) = \min\{3(y/k)^2 - 3(y/k)^4 + (y/k)^6, 1\}$, fulfills all the conditions of Theorem 2 above. Notice that the choice of k determines a value for b [through equation (7)] which in turns determines the breakdown point ϵ^* of the estimate. When F_0 is the standard normal distribution, taking $k = 0.99$ leads to the maximum breakdown point $\epsilon^* = 0.5$. In this case, $b = 0.75$, $\tilde{\epsilon} = 0.15$ and, therefore, $\epsilon_0 = 0.25$. As a consequence, the corresponding dispersion GS-estimate is u.a.n. over $\mathcal{F}(\epsilon, F_0)$ for $\epsilon < 0.25$. However, by decreasing the breakdown point, one can enlarge the size of the neighborhood over which the uniform asymptotic normality holds, as shown in Table 1. The optimal value for the tuning constant seems to be $k = 1.582$. In this case $\epsilon^* = 1 - b = \epsilon_0 = 0.38$.

5 Uniform asymptotic normality for location M-estimates

In this section we state the main result of this paper, that ensures the uniform asymptotic normality over contamination neighborhoods of $\sqrt{n}(\hat{\mu}_n - \mu)$, where $\hat{\mu}_n$ is a location M-estimate [see (1)] with an auxiliar GS-scale $\hat{\sigma}_n$. Let μ be given by the equation

$$\mathrm{E}_F \psi \left(\frac{X - \mu}{\sigma} \right) = 0,$$

where σ was defined in (6). The regularity conditions on the score function ψ are:

A8. *The function ψ is bounded, odd, twice continuously differentiable with $\psi(0) = 0$ and $\psi(x) = \operatorname{sgn}(x)$, for $|x| \geq r$, where $r > 0$ is a tuning constant.*

To give a precise expression for the asymptotic variance, we introduce the following notation: $a = a(F) = \sigma^{-1}\mathrm{E}_F\psi'[(X-\mu)/\sigma]$,

$$d = d(F) = \frac{1}{\sigma}\mathrm{E}_F\left[\psi'\left(\frac{X-\mu}{\sigma}\right)\left(\frac{X-\mu}{\sigma}\right)\right] \tag{9}$$

and

$$\overline{\psi}(x,F,\chi) = \frac{1}{2}\psi\left(\frac{x-\mu}{\sigma}\right) + \frac{d}{c}\overline{\chi}(x,F),$$

where $c = c(F)$ was defined in (8). Finally, let $\tilde{\sigma}^2(F) = \mathrm{E}_F[\overline{\psi}^2(X,F,\chi)]$, and

$$v(F) = \frac{4\tilde{\sigma}^2(F)}{a^2(F)} = 4\sigma^2\frac{\mathrm{E}_F[\overline{\psi}^2(X,F,\chi)]}{[\mathrm{E}_F\psi'\big((X-\mu)/\sigma\big)]^2}. \tag{10}$$

We have the following result:

Theorem 3. *Let $\mathcal{F} = \mathcal{F}(\epsilon, F_0)$ be a contamination neighborhood centered at a distribution F_0 with an even, unimodal and strictly positive density. Let $\hat{\mu}_n$ be a location M-estimate as defined in (1), where $\hat{\sigma}_n$ is a GS-scale based on a score function χ as defined in (3). Let ϵ_0 be as defined in Theorem 2. Then, under* **A7** *and* **A8**, *it holds that $\sqrt{n}(\hat{\mu}_n-\mu)/\sqrt{v(F)}$ is u.a.n. over $\mathcal{F}$, for all $\epsilon < \epsilon_0$, where the asymptotic variance, $v(F)$, is given in equation (10).*

Notice that the maximum size of the neighborhood over which uniform asymptotic normality holds, ϵ_0, does not depend on the choice of the location score function ψ.
Notice also that when the contaminated distribution F is symmetric, then $\mu(F) = 0$ and $d(F) = 0$, because ψ is an odd function. Therefore, $\overline{\psi}(x,F,\chi) = 0.5\psi(x/\sigma)$. As a consequence, the choice of $\hat{\sigma}_n$ does not affect the asymptotic distribution of $\hat{\mu}_n$ any longer. Indeed, the expression for the asymptotic variance that we obtain in this case applying the formula (10) for $v(F)$ in Theorem 3 is well-known:

$$v(F) = \sigma^2\frac{\mathrm{E}_F[\psi^2(X/\sigma)]}{\big[\mathrm{E}_F\psi'(X/\sigma)\big]^2}.$$

Fraiman *et al.* (2001) provide an example of a smooth score funtion ψ that meets all the conditions required in Theorem 3. It is defined as follows

$$\psi_r(x) = \text{sgn}(x) \begin{cases} |x/r| & |x| \leq 0.8r \\ p(|x/r|) & 0.8r < |x| \leq r \\ p(1) & |x| > r \end{cases}, \quad (11)$$

where $r > 0$ is a tuning constant and $p(x) = 38.4 - 175x + 300x^2 - 225x^3 + 62.5x^4$.

6 Conclusions

In this paper, we propose to use location M-estimates calculated with a smooth function ψ_r, as that defined in (11), and using a previous GS-scale calculated with a Tukey's score function χ_k. These choices, when appropriately tuned, allow us to obtain location estimates with the following desirable properties: (a) translation and scale equivariance, (b) high breakdown point, (c) high efficiency under the nominal model, (d) uniform consistency and asymptotic normality over a contamination neighborhood whose size is known.

Using a GS-scale has two advantages over the choice of an S-scale suggested recently by Salibian-Barrera and Zamar (2004): firstly, it produces a simpler expression for the asymptotic variance, because GS-scales do not depend on any auxiliar location estimate; secondly, the size of the neighborhood over which the uniform asymptotic normality holds is larger.

For instance, for a Tukey's score function χ_k with $k = 0.99$ and a location score function ψ_r with $r = 1.525$ we obtain a scale-equivariant location M-estimate with 50% breakdown point, 95% efficiency under the standard normal model, and asymptotically normal, uniformly over a contamination neighborhood of the standard normal model which may include up to a 25% of contamination. On the other hand, choosing $k = 1.58$ and $r = 1.525$ yields a scale-equivariant location M-estimate with 38% breakdown point, 95% efficiency under the gaussian model, and asymptotically normal, uniformly over a contamination neighborhood which may include up to a 38% of contamination.

Appendix: proofs

The facts gathered in Lemma 1 are mentioned in Salibian-Barrera and Zamar (2004). We also include them here for further reference. Lemma 2 is also useful, and the proof is straightforward, so that it is omitted.

Lemma 1. *Let X_n and Y_n be two sequences of random variables and let $\mathcal{F}$ be a set of distribution functions. Then,*

(a) If $X_n = UO_p(1)$ and $Y_n = Uo_p(1)$ over $\mathcal{F}$, then $X_n \cdot Y_n = Uo_p(1)$ over $\mathcal{F}$.

(b) If X_n is u.a.n. and $Y_n = Uo_p(1)$ over $\mathcal{F}$, then $X_n + Y_n$ is u.a.n. over $\mathcal{F}$.

(c) If $X_n = UO_p(1)$ over $\mathcal{F}$ and there exists $c > 0$ such that $Y_n - c = Uo_p(1)$ over $\mathcal{F}$, then $X_n/Y_n = X_n/c + Uo_p(1)$ over $\mathcal{F}$.

Lemma 2. *Let X_n be a sequence of random variables such that X_n is u.a.n. over a set of distributions $\mathcal{F}$. Let $Y_n = c(F)X_n + Uo_p(1)$, where $c_0 \doteq \inf_{F\in\mathcal{F}} c(F) > 0$. Then, $Y_n/c(F)$ is also u.a.n. over $\mathcal{F}$.*

The following lemmas are also useful in the proofs of the main results:

Lemma 3. *Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution $F \in \mathcal{F}$. Let $f_n(\boldsymbol{\theta}; x_1, \ldots, x_n)$ be any real function that depends on the sample and a parameter $\boldsymbol{\theta} = \boldsymbol{\theta}(F) \in \mathbb{R}^d$ such that, for every $F \in \mathcal{F}$,*

$$\sup_{\substack{x_1,\ldots,x_n \\ x_i' \in \mathbb{R}^d}} |f_n(\boldsymbol{\theta}; x_1, \ldots, x_i, \ldots, x_n) - f_n(\boldsymbol{\theta}; x_1, \ldots, x_i', \ldots, x_n)| \leq K_n = O(n^{-1}). \quad (12)$$

Let $Y_n(\boldsymbol{\theta}) \doteq f_n(\boldsymbol{\theta}, X_1, \ldots, X_n) - E_F[f_n(\boldsymbol{\theta}, X_1, \ldots, X_n)]$. Then, $Y_n(\boldsymbol{\theta}) = Uo_p(1)$ over $\mathcal{F}$, and $\sqrt{n}Y_n(\boldsymbol{\theta}) = UO_p(1)$ over $\mathcal{F}$.

Proof of Lemma 3: Under the assumptions of the lemma, we can apply McDiarmid's bounded differences inequality (see, for instance, Devroye and Lugosi (2001), p. 8) to conclude that, for all $\delta > 0$ and n,

$$\sup_{F\in\mathcal{F}} \mathrm{P}_F\{|f_n(\boldsymbol{\theta}; X_1,\ldots,X_n) - \mathrm{E}_F f_n(\boldsymbol{\theta}; X_1,\ldots,X_n)| > \delta\} \le 2\exp\Big(-\frac{2\delta^2}{nK_n^2}\Big) \longrightarrow 0$$

as $n\to\infty$, since $nK_n^2 = o(1)$. Moreover, there exists $L>0$ such that for all n,

$$\begin{aligned} & \mathrm{P}_F\{\sqrt{n}|f_n(\boldsymbol{\theta}; X_1,\ldots,X_n) - \mathrm{E}_F f_n(\boldsymbol{\theta}; X_1,\ldots,X_n)| > M\} \\ \le\ & 2\exp\Big(-\frac{2M^2}{n^2K_n^2}\Big) \le 2\exp\Big(-\frac{2M^2}{L}\Big), \end{aligned}$$

since $n^2K_n^2 = (nK_n)^2$ is bounded by the assumption on K_n. Therefore,

$$\lim_{M\to\infty}\lim_{n\to\infty}\sup_{F\in\mathcal{F}} \mathrm{P}_F\{\sqrt{n}|f_n(\boldsymbol{\theta}; X_1,\ldots,X_n) - \mathrm{E}_F f_n(\boldsymbol{\theta}; X_1,\ldots,X_n)| > M\} = 0,$$

which amounts to the second part of the result. □

Remark: When $f_n(\boldsymbol{\theta}; x_1,\ldots,x_n) = \binom{n}{2}^{-1}\sum_{i<j}\eta(\boldsymbol{\theta}; x_i, x_j)$ is a U-statistic with bounded kernel η, then (12) holds since, $|\eta(\boldsymbol{\theta}; x, y)| \le K$ for all $\boldsymbol{\theta}, x, y$, implies

$$\begin{aligned} & \sup_{\substack{x_1,\ldots,x_n \\ x_i'\in\mathbb{R}}} |f_n(\boldsymbol{\theta}; x_1,\ldots,x_i,\ldots,x_n) - f_n(\boldsymbol{\theta}; x_1,\ldots,x_i',\ldots,x_n)| \\ \le\ & \binom{n}{2}^{-1}(n-1)2K = O(n^{-1}). \end{aligned}$$

Lemma 4. *Let $X_1,\ldots,X_n$ be i.i.d. random variables with distribution $F\in\mathcal{F}$. Let*

$$f_n(\boldsymbol{\theta}) \doteq \binom{n}{2}^{-1}\sum_{i<j}\eta(\boldsymbol{\theta}; x_i, x_j),$$

a U-statistic with bounded kernel η, Assume that $\inf_{F\in\mathcal{F}}\tilde\sigma^2(F) > 0$, where $\tilde\sigma^2(F) = Var[\bar\eta(\boldsymbol{\theta}; X, F)]$. Then, $\sqrt{n}f_n(\boldsymbol{\theta})/[2\tilde\sigma(F)]$ is u.a.n. over $\mathcal{F}$.

Proof of Lemma 4: We apply a Berry-Essen Theorem for U-statistics [see, for instance, Lee (1990), p. 97] to conclude that there exist constants c_1, c_2 and c_3 depending neither on n, η nor F, such that

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}_F\left\{\frac{\sqrt{n}f_n(\boldsymbol{\theta})}{2\tilde{\sigma}(F)}\le x\right\}-\Phi(x)\right|\le\frac{1}{\sqrt{n}}\left[c_1\rho(F)+c_2\lambda_{5/3}(F)+c_3(\rho(F)\lambda_{5/3}(F))^{2/3}\right],$$

where $\rho(F)$ and $\lambda_j(F)$ are precisely defined in Lee (1990), p. 98. Under our assumptions, it is easy to check that $\sup_{F\in\mathcal{F}}\rho(F)<\infty$ and $\sup_{F\in\mathcal{F}}\lambda_{5/3}(F)<\infty$. As a consequence, $\sqrt{n}f_n(\boldsymbol{\theta})/[2\tilde{\sigma}(F)]$ is u.a.n. over $\mathcal{F}$. $\square$

Proof of Theorem 1: Since η is twice continuously differentiable, we have the following Taylor expansion,

$$0=h_n(\widehat{\theta}_n)=h_n(\theta)+h_n'(\theta)(\widehat{\theta}_n-\theta)+\frac{1}{2}h''(\tilde{\theta}_n)(\widehat{\theta}_n-\theta)^2,$$

where $\tilde{\theta}_n$ lies between θ and $\widehat{\theta}_n$. Then,

$$\sqrt{n}(\widehat{\theta}_n-\theta)=-\frac{\sqrt{n}h_n(\theta)}{h_n'(\theta)+R_n/2},\tag{13}$$

where $R_n=h''(\tilde{\theta}_n)(\widehat{\theta}_n-\theta)$. By Lemma 3 [applied to a U-statistic with kernel η', which is uniformly bounded by **A2**] it follows that $h_n'(\theta)=-c(F)+Uo_p(1)$. Now we prove that $R_n=Uo_p(1)$. Given an arbitrary $\delta>0$, we have that

$$\mathrm{P}_F\{|R_n|>\delta\}=\mathrm{P}_F\{|R_n|>\delta,\,|\widehat{\theta}_n-\theta|\ge\beta\}+\mathrm{P}_F\{|R_n|>\delta,\,|\widehat{\theta}_n-\theta|<\beta\},$$

where $\beta>0$ is small enough so that condition **A3** holds. Notice that

$$\sup_{F\in\mathcal{F}}\mathrm{P}_F\{|R_n|>\delta,\,|\widehat{\theta}_n-\theta|\ge\beta\}\le\sup_{F\in\mathcal{F}}\mathrm{P}_F\{|\widehat{\theta}_n-\theta|\ge\beta\}\to 0,\tag{14}$$

as $n\to\infty$, by **A6**. On the other hand, when $|\widehat{\theta}_n-\theta|<\beta$, we have $\sup_{F\in\mathcal{F}}|h''(\tilde{\theta}_n)|\le M$, by **A3**. Then,

$$\sup_{F\in\mathcal{F}}\mathrm{P}_F\{|R_n|>\delta,\,|\widehat{\theta}_n-\theta|<\beta\}\le\sup_{F\in\mathcal{F}}\mathrm{P}_F\{|\widehat{\theta}_n-\theta|>\delta/M,\,|\widehat{\theta}_n-\theta|<\beta\}=0,\tag{15}$$

if $\beta<\delta/M$. As a consequence of (14) and (15), $R_n=Uo_p(1)$ so that the denominator in (13) satisfies $h_n'(\theta)+R_n/2=-c(F)+Uo_p(1)$. Notice also that **A1** and **A5** imply the assumptions of Lemma 4 and, therefore, $\sqrt{n}h_n(\theta)/[2\tilde{\sigma}(F)]$ is u.a.n. over $\mathcal{F}$. This fact, together with Lemma 1(c) applied to (13), shows equation (4). Furthermore,

$$\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sqrt{v(F)}} = \frac{c(F)}{2\tilde{\sigma}(F)}\sqrt{n}(\widehat{\theta}_n - \theta) = \frac{\sqrt{n}h_n(\theta)}{2\tilde{\sigma}(F)} + Uo_p(1).$$

To end the proof, we just have to apply Lemma 1(b). □

Proof of Theorem 2: Obviously, **A7** implies **A1**. Assumptions **A2** and **A3** are not difficult to check for any $\epsilon < \epsilon^*$. Moreover, in Berrendero (2003) it is shown that **A7** also implies **A6**. Therefore, it remains to show **A4** and **A5**. Notice that **A4** reduces to show that

$$\inf_{F\in\mathcal{F}} \mathrm{E}_{F\times F}\left[\chi'\left(\frac{X-Y}{\sigma(F)}\right)\left(\frac{X-Y}{\sigma(F)}\right)\right] > 0.$$

To prove this fact, note that for $\epsilon < \epsilon^*$, $0 < S^-(\epsilon) \le \sigma(F) \le S^+(\epsilon) < \infty$, and, moreover, $\chi'(x/s)(x/s)$ is continuous for all $x \in \mathbb{R}$ and $s > 0$. Therefore,

$$\mathrm{E}_{F\times F}\left[\chi'\left(\frac{X-Y}{\sigma(F)}\right)\left(\frac{X-Y}{\sigma(F)}\right)\right] \ge (1-\epsilon)^2 \inf_{F\in\mathcal{F}} \mathrm{E}_{F_0\times F_0}\left[\chi'\left(\frac{X-Y}{\sigma(F)}\right)\left(\frac{X-Y}{\sigma(F)}\right)\right] > 0,$$

since $\chi'(x)x \ge 0$, $\sup_x \chi'(x)x > 0$, and F_0 has a strictly positive density. Finally, to prove **A5** note that, by Chebychev inequality we have that for all $\delta > 0$,

$$\begin{aligned}
\tilde{\sigma}^2(F) &\ge \delta^2(1-\epsilon)\mathrm{P}_{F_0}\{|\bar{\eta}(\sigma; X, F)| > \delta\} \\
&\ge \delta^2(1-\epsilon)\max\left[\mathrm{P}_{F_0}\{\bar{\eta}(\sigma; X, F) > \delta\}, \mathrm{P}_{F_0}\{\bar{\eta}(\sigma; X, F) < -\delta\}\right] \\
&\doteq \delta^2(1-\epsilon)\max\left[\mathrm{P}_{F_0}\{X \in A_F\}, \mathrm{P}_{F_0}\{X \in B_F\}\right],
\end{aligned}$$

where $A_F \doteq \{x \in \mathbb{R} : \bar{\eta}(\sigma; x, F) > \delta\}$ and $B_F \doteq \{x \in \mathbb{R} : \bar{\eta}(\sigma; x, F) < -\delta\}$. Notice that

$$\bar{\eta}(\sigma; x, F) = \mathrm{E}_F\chi\left(\frac{x-Y}{\sigma(F)}\right) - b \ge (1-\epsilon)\mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^+(\epsilon)}\right) - b,$$

so that

$$A_F \supset \left\{x \in \mathbb{R} : \mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^+(\epsilon)}\right) > \frac{b+\delta}{1-\epsilon}\right\}. \tag{16}$$

and the smaller set does not depend on F. By **A7**, it holds that

$$\sup_{x\in\mathbb{R}} \mathrm{E}_{F_0}\chi\left[\frac{x-Y}{S^+(\epsilon)}\right] = 1.$$

Furthermore, for $\epsilon < 1-b$ and $0 < \delta < 1-b-\epsilon$, we have that $(b+\delta)/(1-\epsilon) < 1$. Then, from (16) and the assumptions on F_0 it follows that $\inf_{F\in\mathcal{F}} \mathrm{P}_{F_0}\{X| \in A_F\} > 0$, for $\epsilon < 1-b$.

On the other hand,

$$\bar{\eta}(\sigma; x, F) = \mathrm{E}_F \chi\left(\frac{x-Y}{\sigma(F)}\right) - b \leq (1-\epsilon)\mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^-(\epsilon)}\right) + \epsilon - b.$$

It follows that

$$B_F \supset \left\{x \in \mathbb{R} : \mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^-(\epsilon)}\right) < \frac{b-\epsilon-\delta}{1-\epsilon}\right\},$$

where, again, the smaller set does not depend on F. The properties of F_0 and χ imply that

$$\inf_{x\in\mathbb{R}} \mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^-(\epsilon)}\right) = \mathrm{E}_{F_0}\chi\left(\frac{Y}{S^-(\epsilon)}\right).$$

Therefore, for small enough $\delta > 0$, we have $\inf_{F\in\mathcal{F}} \mathrm{P}_{F_0}\{X \in B_F\} > 0$ whenever

$$\mathrm{E}_{F_0}\chi\left(\frac{Y}{S^-(\epsilon)}\right) < \frac{b-\epsilon}{1-\epsilon},$$

but this inequality holds if $\epsilon < \tilde{\epsilon}$. □

Proof of Theorem 3: First, we introduce the following notation:

$$\begin{aligned}
g_n(t,s) &\doteq \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{x_i-t}{s}\right), \\
a_n = a_n(F) &\doteq -\left[\frac{\partial}{\partial t}g_n(t,s)\right]_{(t,s)=(\mu,\sigma)} = \frac{1}{\sigma}\sum_{i=1}^{n}\psi'\left(\frac{x_i-\mu}{\sigma}\right), \\
d_n = d_n(F) &\doteq \left[\frac{\partial}{\partial s}g_n(t,s)\right]_{(t,s)=(\mu,\sigma)} = -\frac{1}{\sigma}\sum_{i=1}^{n}\left[\psi'\left(\frac{x_i-\mu}{\sigma}\right)\left(\frac{x_i-\mu}{\sigma}\right)\right].
\end{aligned}$$

Finally, let $\psi_i \doteq \psi[(x_i-\mu)/\sigma]$ and $\chi_{ij} \doteq \chi[(x_i-x_j)/\sigma]$. Applying Theorem 4 in Berrendero (2003) and Theorem 1 in Salibian-Barrera and Zamar (2004) we see that $\hat{\sigma}_n = \sigma + Uo_p(1)$ and $\hat{\mu}_n = \mu + Uo_p(1)$ over $\mathcal{F}(\epsilon, F_0)$ for $\epsilon < \epsilon^*$. Moreover, since $\psi'(x)$ is bounded, we can apply Lemma 3 to deduce that

$$a_n = a + Uo_p(1), \tag{17}$$

where $a = a(F) = \sigma^{-1}\mathrm{E}_F\psi'[(X-\mu)/\sigma]$. It is not difficult to prove that $\inf_{F\in\mathcal{F}} a(F) > 0$. Since $x\psi'(x)$ is also bounded, Lemma 3 implies

$$d_n = d + Uo_p(1), \tag{18}$$

where $d = d(F)$ is defined in equation (9). Next, we use the following first order Taylor expansion, which is valid because ψ is twice continuously differentiable:

$$g_n(\hat{\mu}_n, \hat{\sigma}_n) = g_n(\mu,\sigma) - a_n(\hat{\mu}_n - \mu) + d_n(\hat{\sigma}_n - \sigma) + \frac{1}{2}\left[\frac{\partial^2}{\partial t^2} g_n(t,s)\right]_{(t,s)=(\tilde{\mu}_n,\tilde{\sigma}_n)} (\hat{\mu}_n - \mu)^2$$
$$+ \quad \frac{1}{2}\left[\frac{\partial^2}{\partial s^2} g_n(t,s)\right]_{(t,s)=(\tilde{\mu}_n,\tilde{\sigma}_n)} (\hat{\sigma}_n - \sigma)^2 + \frac{1}{2}\left[\frac{\partial^2}{\partial t \partial s} g_n(t,s)\right]_{(t,s)=(\tilde{\mu}_n,\tilde{\sigma}_n)} (\hat{\mu}_n - \mu)(\hat{\sigma}_n - \sigma),$$

where $(\tilde{\mu}_n, \tilde{\sigma}_n)$ lies in the segment linking (μ, σ) and $(\hat{\mu}_n, \hat{\sigma}_n)$. Notice that

$$\left[\frac{\partial^2}{\partial t^2} g_n(t,s)\right]_{(t,s)=(\tilde{\mu}_n,\tilde{\sigma}_n)} (\hat{\mu}_n - \mu) = \left(\frac{\hat{\mu}_n - \mu}{\tilde{\sigma}_n^2}\right) \frac{1}{n}\sum_{i=1}^{n} \psi''\left(\frac{x_i - \tilde{\mu}_n}{\tilde{\sigma}_n}\right) = \left(\frac{\hat{\mu}_n - \mu}{\tilde{\sigma}_n^2}\right) UO_p(1)$$

since, by **A8**, $\psi''(x)$ is bounded. By Lemma 1,

$$\left(\frac{\hat{\mu}_n - \mu}{\tilde{\sigma}_n^2}\right) UO_P(1) = \left(\frac{\hat{\mu}_n - \mu}{\sigma^2}\right) UO_p(1) + Uo_p(1) = Uo_p(1).$$

A similar argument works to prove that the rest of second order terms in the Taylor expansion are also $Uo_p(1)$. Since $g_n(\hat{\mu}_n, \hat{\sigma}_n) = 0$, applying (17) and (18) and Theorem 1 [to ensure that $\sqrt{n}(\hat{\sigma}_n - \sigma) = UO_p(1)$] we deduce

$$[a + Uo_p(1)]\sqrt{n}(\hat{\mu}_n - \mu) = \sqrt{n}[g_n(\mu,\sigma) + d(\hat{\sigma}_n - \sigma)] + Uo_p(1).$$

Applying Lemma 1(c),

$$\sqrt{n}(\hat{\mu}_n - \mu) = \frac{1}{a}\sqrt{n}[g_n(\mu,\sigma) + d(\hat{\sigma}_n - \sigma)] + Uo_p(1). \tag{19}$$

By equation (4) in Theorem 1,

$$\sqrt{n}[g_n(\mu,\sigma) + d(\hat{\sigma}_n - \sigma)] = \sqrt{n}\binom{n}{2}^{-1} \sum_{i<j} \eta(\boldsymbol{\theta}; x_i, x_j) + Uo_p(1),$$

where $\boldsymbol{\theta} = (\mu, \sigma, d, c)$ and

$$\eta(\boldsymbol{\theta}; x_i, x_j) = \frac{\psi_i + \psi_j}{2} + \frac{d}{c}(\chi_{ij} - b). \tag{20}$$

The next step is to show that $\sqrt{n}[g_n(\mu,\sigma) + d(\hat{\sigma}_n - \sigma)]/[2\tilde{\sigma}(F)]$ is u.a.n. over $\mathcal{F}$, where $\tilde{\sigma}^2(F) = \mathrm{Var}_F \bar{\eta}(\boldsymbol{\theta}, X, F)$. For that purpose we apply Lemma 4. What we need is to check that the kernel η is bounded and also that $\inf_{F\in\mathcal{F}} \tilde{\sigma}(F) > 0$. The functions ψ and χ are bounded, and we have already verified that $\inf_{F\in\mathcal{F}} c(F) > 0$ (see the proof of Theorem 2). Furthermore, it is obvious that $\sup_{F\in\mathcal{F}} |d(F)| < \infty$ since $|x\psi'(x)|$ is bounded and $\sigma(F) \geq S^-(\epsilon) > 0$, for all $\epsilon < \epsilon^*$. As a consequence $\eta(\boldsymbol{\theta}; x_i, x_j)$ is bounded. Regarding $\inf_{F\in\mathcal{F}} \tilde{\sigma}(F) > 0$, define $\mathcal{F}^+ = \{F \in \mathcal{F} : d(F) \geq 0\}$ and $\mathcal{F}^- = \{F \in \mathcal{F} : d(F) < 0\}$. Then,

$$\inf_{F\in\mathcal{F}} \tilde{\sigma}(F) = \min\{\inf_{F\in\mathcal{F}^+} \tilde{\sigma}(F), \inf_{F\in\mathcal{F}^-} \tilde{\sigma}(F)\}. \tag{21}$$

We will separately analyze the two terms inside the minimum of equation (21). As in the proof of Theorem 2, it is enough to show that for $\epsilon < \epsilon_0$ there exists $\delta > 0$ such that

$$\max\left[\inf_{F\in\mathcal{F}} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) > \delta\}, \inf_{F\in\mathcal{F}} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) < -\delta\}\right] > 0,$$

for $\mathcal{F} = \mathcal{F}^+$ and $\mathcal{F} = \mathcal{F}^-$. First, consider $F \in \mathcal{F}^+$. Denote $T(\epsilon) \doteq \sup_{F\in\mathcal{F}} |\mu(F)|$. Since $\epsilon < \epsilon^*$, we have $T(\epsilon) < \infty$. In this case, for $x > T(\epsilon)$, it follows that

$$\begin{aligned}
\inf_{F\in\mathcal{F}^+} \bar{\eta}(\boldsymbol{\theta}, x, F) &= \inf_{F\in\mathcal{F}^+} \frac{1}{2}\psi\left(\frac{x-\mu}{\sigma}\right) + \frac{d(F)}{c(F)}\left[\mathrm{E}_F \chi\left(\frac{x-Y}{\sigma}\right) - b\right] \\
&\geq \frac{1}{2}\psi\left(\frac{x-T(\epsilon)}{S^+(\epsilon)}\right) + \left[(1-\epsilon)\mathrm{E}_{F_0}\chi\left(\frac{x-Y}{S^+(\epsilon)}\right) - b\right] \inf_{F\in\mathcal{F}^+} \frac{d(F)}{c(F)} \\
&\geq \frac{1}{2}\psi\left(\frac{x-T(\epsilon)}{S^+(\epsilon)}\right) > 0,
\end{aligned}$$

if $\epsilon < 1-b$ and x is large enough. As a consequence, if $\epsilon < 1-b$, there exists $\delta > 0$ such that

$$\inf_{F\in\mathcal{F}^+} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) > \delta\} \geq \mathrm{P}_{F_0}\{\inf_{F\in\mathcal{F}^+} \bar{\eta}(\boldsymbol{\theta}; X, F) > \delta\} > 0.$$

Likewise, for $x < -T(\epsilon)$, we have that

$$\begin{aligned}
\sup_{F\in\mathcal{F}^+} \bar{\eta}(\boldsymbol{\theta}, x, F) &\leq \frac{1}{2}\psi\left(\frac{x+T(\epsilon)}{S^+(\epsilon)}\right) + \left[(1-\epsilon)\mathrm{E}_{F_0}\chi\left(\frac{Y}{S^-(\epsilon)}\right) + \epsilon - b\right] \sup_{F\in\mathcal{F}^+} \frac{d(F)}{c(F)} \\
&\leq \frac{1}{2}\psi\left(\frac{x+T(\epsilon)}{S^+(\epsilon)}\right) < 0,
\end{aligned}$$

if $\epsilon < \tilde{\epsilon}$ and $x < -T(\epsilon)$. As a consequence, if $\epsilon < \tilde{\epsilon}$, there exists $\delta > 0$ such that

$$\inf_{F \in \mathcal{F}^+} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) < -\delta\} \geq \mathrm{P}_{F_0}\{\sup_{F \in \mathcal{F}^+} \bar{\eta}(\boldsymbol{\theta}; X, F) < -\delta\} > 0.$$

A symmetric proof allows us to show that for $\epsilon < \epsilon_0$, there exists $\delta > 0$ such that

$$\max\left[\inf_{F \in \mathcal{F}^-} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) > \delta\}, \inf_{F \in \mathcal{F}^-} \mathrm{P}_{F_0}\{\bar{\eta}(\boldsymbol{\theta}; X, F) < -\delta\}\right] > 0.$$

Finally, let $v(F) = 4\tilde{\sigma}^2(F)/a^2(F)$. From (19) it follows that

$$\frac{\sqrt{n}(\hat{\mu}_n - \mu)}{\sqrt{v(F)}} = \frac{\sqrt{n}[g_n(\mu, \sigma) + d(\hat{\sigma}_n - \sigma)]}{2\tilde{\sigma}(F)} + Uo_p(1),$$

where, by Lemma 1(b), the right-hand side is u.a.n. over $\mathcal{F}$ for $\epsilon < \epsilon_0$. The result follows from this fact. □

References

Bednarski, T. and Müller, C. (2001). Optimal bounded influence regression and scale M-estimators in the context of experimental design. *Statistics*, **35**, 349–369.

Bednarski, T. and Zontek, S. (1996). Robust estimation of parameters in a mixed unbalanced model. *Ann. Statist.*, **24**, 1493–1510.

Berrendero, J.R. (2003). Uniform strong consistency of robust estimators. *Statist. Probab. Lett.*, **64**, 159–168.

Clarke, B.R. (1986). Nonsmooth analysis and Fréchet differentiability of M-functionals. *Probability theory and related fields*, **73**, 197-209.

Croux, C., Rousseeuw, P.J. and Hössjer, O. (1994). Generalized S-estimators. *J. Amer. Statist. Assoc.*, **89**, 1271–1281.

Davies, P.L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *Ann. Statist.*, **26**, 1103–1125.

Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation.* Springer, New York.

Fraiman, R., Yohai, V.J. and Zamar, R.H. (2001). Optimal robust M-estimates of location. *Ann. Statist.*, **29**, 194–223.

Hampel, F. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887–1896.

Hössjer, O., Croux, C. and Rousseeuw, P.J. (1994). Asymptotics of generalized S-estimators. *J. Multivariate Anal.*, **51**, 148–177.

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

Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

Lee, A.J. (1990). *U-Statistics, Theory and Practice*, Marcel Dekker, New York.

Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series* (J. Franke, W. Hardle and D. Martin, eds.) *Lecture notes in Statistics,* **26**, 256–272. Springer-Verlag, Berlin.

Salibian-Barrera, M. and Zamar, R.H. (2004). Uniform asymptotics for robust location estimates when the scale is unknown. *Ann. Statist.*, **32**.

Tukey, J.W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, I. Olkin, Ed., Standford University Press, Stanford, California.

Zielinski, R. (1998). Uniform strong consistency of sample quantiles. *Statist. Probab. Lett.*, **37**, 115–119.