.

UNIFORM STRONG CONSISTENCY OF ROBUST ESTIMATORS

José R. Berrendero [1]

Departamento de Matemáticas
Universidad Autónoma de Madrid
28049 Madrid, Spain

Abstract

In the robustness framework, the distribution underlying the data is not totally specified and, therefore, it is convenient to use estimators whose properties hold uniformly over the whole set of possible distributions. In this paper we give two general results on uniform strong consistency and apply them to study the uniform consistency of some classes of robust estimators over contamination neighborhoods. Some instances covered by our results are Huber's M-estimators, quantiles, or generalized S-estimators.

KEY WORDS: Uniform strong consistency, robustness, M-estimators, GS-estimators.
MSC: 62F12, 62F35.

[1]This research has been partially supported by Spanish Grant BFM2001-0169.

1 Introduction

A key feature of the statistical robustness theory is the uncertainty about the probabilistic model F underlying the data. The distribution F is not completely specified but it is rather assumed to belong to a family $\mathcal{F}$ of plausible distributions. From this point of view, it would be convenient to study the properties of the estimators that hold, in some sense, uniformly over the family $\mathcal{F}$ of possible distributions.

For instance, the strong consistency of an estimator $\hat{\theta}_n$ guarantees that the error of estimation $|\hat{\theta}_n - \theta|$ will be small with arbitrarily high probability as long as the sample size n is large enough. However, given a lower bound for $\mathrm{P}\{\sup_{m\geq n} |\hat{\theta}_m - \theta| < \delta\}$, the sample size required to attain it may be considerably different for different distributions $F \in \mathcal{F}$ unless the consistency holds uniformly on $\mathcal{F}$. This is the kind of uniform asymptotic property that we address in this paper. That is, our goal is to show, for a large class of estimators,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta| > \delta\} = 0, \quad \text{for every } \delta > 0.$$

There exists some previous work on uniform asymptotic properties: In his classic book, Huber (1981), p. 51, already mentioned the convenience of establishing conditions under which the estimators are asymptotically normal, uniformly over a neighborhood of distributions. Further comments on this question can be found in Davies (1993). Shorack (2000), p. 252, credits Chung with a result that establishes uniform integrability conditions on $\mathcal{F}$ that imply the uniform strong consistency of the sample mean. Zielinski (1998) characterized the strong uniform consistency of sample quantiles: if $\theta_p = \theta_p(F)$ is the pth quantile of F and $X_{(k_n)}$ is an order statistic from the sample such that $k_n/n \to p$, then

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{\sup_{m\geq n} |X_{(k_n)} - \theta_p| > \delta\} = 0, \text{ for every } \delta > 0$$

if and only if

$$\inf_{F\in\mathcal{F}} \min\{p - F(\theta_p - \delta), F(\theta_p + \delta) - p\} > 0, \text{ for every } \delta > 0. \tag{1}$$

Fraiman, Yohai and Zamar (2001) gave sufficient conditions for the uniform asymptotic normality of location M-estimators over contamination neighborhoods and also addressed the usefulness of this result to obtain robust confidence intervals. Salibian-Barrera and Zamar (2001) studied the uniform asymptotic properties of location M-estimators (when the scale is unknown and estimated with an S-estimator) over contamination neighborhoods. In particular, they gave sufficient conditions for the uniform strong consistency in this case. More uniform asymptotic results can be found in Davies (1998). These are locally uniform results meaning that for each distribution there exists a neighborhood of distributions where the convergence holds uniformly.

In this paper, we give two general results on uniform strong consistency (Sections 2 and 3) and apply them to study the uniform consistency of some classes of robust estimators over contamination neighborhoods (Section 4). Further comments and open questions are included in Section 5. A final appendix contains the proofs.

2 A general result on uniform strong consistency

Assume that we have a sample $X_1, \ldots, X_n$ of independent, identically distributed d-dimensional random vectors drawn from a distribution F on $\mathbb{R}^d$, which may be any distribution of certain family $\mathcal{F}$. We will consider the general class of estimators that solve an estimating equation. That is, provided that we have a meaningful monotone score function $h : \mathbb{R} \times \mathbb{R}^{nd} \to \mathbb{R}$, the estimator $\hat{\theta}_n$ satisfies $h(\hat{\theta}_n; X_1, \ldots, X_n) = 0$. Actually, $\hat{\theta}_n$ is precisely defined as

$$\hat{\theta}_n \doteq \inf\{t : \ h(t; X_1, \ldots, X_n) < 0\} \tag{2}$$

to cope with discontinuous or non strictly monotone score functions. Different specifications of the score function lead to some well-known classes of estimators (see Section 4 below). Let $H(t, F) \doteq \mathrm{E}_F[h(t; X_1, \ldots, X_n)]$ and define

$$\theta = \theta(F) \doteq \inf\{t : \ H(t, F) < 0\}. \tag{3}$$

Under some conditions on h and $\mathcal{F}$, we will show that the difference between $\hat{\theta}_n$ and θ vanishes a.s. uniformly over $\mathcal{F}$ as $n \to \infty$. We are implicitly assuming that $H(t, F)$

does not depend on the sample size n, but this condition holds in all the applications of Section 4. We also need the following assumptions:

A1. Monotonicity*: The score function $h(t; x_1, \dots, x_n)$ is decreasing in t, for all $x_1, \dots, x_n \in \mathbb{R}^d$.*

A2. Bounded differences*: For each $t \in \mathbb{R}$ and $i = 1, \dots, n$ there exists $c_{in} \in \mathbb{R}$ such that*

$$\sup_{\substack{x_1, \dots, x_n \\ x_i' \in \mathbb{R}^d}} |h(t; x_1, \dots, x_i, \dots, x_n) - h(t; x_1, \dots, x_i', \dots, x_n)| \leq c_{in},$$

and $\sum_{n=1}^{\infty} \exp(-\gamma / \sum_{i=1}^{n} c_{in}^2) < \infty$, for all $\gamma > 0$.

A3. Uniform lower bound*: The set $I_\theta(\mathcal{F}) \doteq \{\theta(F) : F \in \mathcal{F}\}$ is bounded. Moreover, for every $\delta > 0$,*

$$\inf_{F \in \mathcal{F}} \min\{H[\theta(F) - \delta, F], -H[\theta(F) + \delta, F]\} > 0. \tag{4}$$

The following theorem is the main result of this paper:

Theorem 1. *Let $\hat{\theta}_n$ and θ be defined as in (2) and (3) respectively. Under assumptions* **A1**, **A2** *and* **A3**, *it holds*

$$\lim_{n \to \infty} \sup_{F \in \mathcal{F}} P_F\{\sup_{m \geq n} |\hat{\theta}_m - \theta| > \delta\} = 0, \quad \textit{for every } \delta > 0. \tag{5}$$

Remark: From the proof of Theorem 1 it actually follows that, for all $\delta > 0$,

$$\sum_{n=1}^{\infty} \sup_{F \in \mathcal{F}} \mathrm{P}_F\{|\hat{\theta}_n - \theta| > \delta\} < \infty.$$

This condition is slightly stronger than uniform strong consistency and it is usually called (uniform) complete convergence.

Assumption **A2** requires that the function h is robust to changes in the value of each observation. Since this robustness is usually inherited by the estimator, our result is mainly aimed at robust estimators. Moreover, notice that as $n \to \infty$ the upper bounds

c_{in} in **A2** must decrease to zero at a rate such that certain series converges. It is not difficult to check that $c_{in} = O(n^{-1})$ suffices.

The aspect of condition (4) is similar to that of Zielinski's condition (1) and they are indeed equivalent when applied to quantiles. To clarify the meaning of (4), notice that if $H(t, F)$ is differentiable at $t = \theta$, then

$$H(\theta - \delta, F) = -\delta H'(\theta, F) + o(\delta)$$

and

$$H(\theta + \delta, F) = \delta H'(\theta, F) + o(\delta)$$

Therefore, a sufficient condition for (4) is $\inf_{F\in\mathcal{F}} H'(\theta, F) < 0$, that is, the minimum slope of the functions $H(\cdot, F)$ at θ is strictly negative.

Note that the only assumption depending on the family of distributions $\mathcal{F}$ is **A3**. If we apply **A3** to $\mathcal{F} = \{F\}$, a family of distributions consisting of a single distribution F, then condition (4) reduces to

$$\min\{H(\theta - \delta, F), -H(\theta + \delta, F)\} > 0, \text{ for every } \delta > 0. \tag{6}$$

Therefore, **A1**, **A2** and (6) provide a simple set of conditions that enables a unified approach to prove the (non uniform) strong consistency of a large class of estimators.

The mathematical tool that gives to Theorem 1 its generality is a powerful result due to McDiarmid (1989). This is a large deviation-type inequality that allows to handle complicated functions of independent random variables. Further applications of this result to the field of non parametric density estimation can be found in Devroye and Lugosi (2000).

3 Nuisance parameters

Although Theorem 1 is relatively simple and satisfactory for several purposes, it cannot be applied straightforwardly when a nuisance parameter must be estimated together with the parameter of main interest. For instance, if we want a location M-estimator to be scale-equivariant (see Section 4), we must use a previous auxiliary scale estimator. In this section, we address the adjustments needed to handle these problems. Roughly speaking, adding a continuity condition to the function h, and assuming that the auxiliary estimator is uniformly strongly consistent, suffices to obtain the uniform strong consistency of the main estimator.

Let $h(t, s, x_1, \ldots, x_n)$ be the function that defines the estimating equation. To handle the auxiliary estimator of the nuisance parameter we add the new argument s. Let $\hat{\sigma}_n$ be the auxiliary estimator and assume that $\hat{\sigma}_n$ is completely convergent to some value $\sigma = \sigma(F)$, uniformly over $\mathcal{F}$. This means that

$$\sum_{n=1}^{\infty} \sup_{F \in \mathcal{F}} \mathrm{P}_F\{|\hat{\sigma}_n - \sigma| > \delta\} < \infty, \quad \text{for all } \delta > 0. \tag{7}$$

We can use Theorem 1 to check (7) as observed in the remark after Theorem 1. The estimator $\hat{\theta}_n$ is now defined as

$$\hat{\theta}_n \doteq \inf\{t : \; h(t, \hat{\sigma}_n; X_1, \ldots, X_n) < 0\}. \tag{8}$$

and its tentative limit is given by

$$\theta = \theta(F) \doteq \inf\{t : \; H(t, \sigma, F) < 0\}, \tag{9}$$

where $H(t, s, F) \doteq \mathrm{E}_F[h(t, s; X_1, \ldots, X_n)]$. Let $I_\theta(\mathcal{F}) \doteq \{\theta(F) : \; F \in \mathcal{F}\}$ and $I_\sigma(\mathcal{F}) \doteq \{\sigma(F) : \; F \in \mathcal{F}\}$. In this new context, Assumptions **A1**, **A2** and **A3** are replaced respectively by the following ones:

A4. *The score function $h(t, s; x_1, \ldots, x_n)$ is decreasing in t, for all $s \in I_\sigma(\mathcal{F})$ and for all $x_1, \ldots, x_n \in \mathbb{R}^d$.*

A5. *For each $t \in \mathbb{R}$, $s \in \mathbb{R}$, and $i = 1, \ldots, n$ there exists $c_{in} \in \mathbb{R}$ such that*

$$\sup_{\substack{x_1, \ldots, x_n \\ x_i' \in \mathbb{R}^d}} |h(t, s; x_1, \ldots, x_i, \ldots, x_n) - h(t, s; x_1, \ldots, x_i', \ldots, x_n)| \leq c_{in},$$

and $\sum_{n=1}^{\infty} \exp(-\gamma / \sum_{i=1}^{n} c_{in}^2) < \infty$, for all $\gamma > 0$.

A 6. *The sets* $I_\theta(\mathcal{F})$ *and* $I_\sigma(\mathcal{F})$ *are bounded. Moreover,* $\inf_{F\in\mathcal{F}} \sigma(F) > 0$ *and, for every* $\delta > 0$,

$$\inf_{F\in\mathcal{F}} \min\{H[\theta(F)-\delta, \sigma(F), F], -H[\theta(F)+\delta, \sigma(F), F]\} > 0. \tag{10}$$

We will also need to impose an additional uniform continuity condition on the estimating function h.

A 7. *For every* $[\underline{t}, \overline{t}] \subset \mathbb{R}$, $[\underline{s}, \overline{s}] \subset (0, \infty)$ *and* $\alpha > 0$, *there exists* $\eta > 0$ *such that if* $s_1, s_2 \in (\overline{s}, \underline{s})$ *and* $|s_1 - s_2| < \eta$ *then* $|h(t, s_1; x_1, \ldots, x_n) - h(t, s_2; x_1, \ldots, x_n)| < \alpha$, *for all* $t \in [\underline{t}, \overline{t}]$, *and all* $x_1, \ldots, x_n \in \mathbb{R}^d$.

We have the following result:

Theorem 2. *Let* $\hat{\theta}_n$ *and* θ *be defined as in (8) and (9) respectively, where* $\hat{\sigma}_n$ *and* σ *satisfy (7). Under assumptions* **A4**, **A5**, **A6** *and* **A7**, *it holds*

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}} P_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta| > \delta\} = 0, \quad \textit{for every } \delta > 0.$$

4 Contamination neighborhoods

First introduced by Tukey (1960), contamination neighborhoods have been extensively used in robustness theory to formalize the presence in the sample of gross errors, outliers and other deviations from a prototypical data set coming from a specified model F_0. Given a distribution function F_0, we define the ϵ-contamination neighborhood $\mathcal{F}_\epsilon$ as

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

Therefore, if a sample has been drawn from any distribution $F \in \mathcal{F}_\epsilon$, most of the data come from a “nominal” specified distribution F_0 but there is still a fraction ϵ of outliers coming from an arbitrary distribution H.

Given a statistical functional $\theta(F)$ defined on $\mathcal{F}_\epsilon$ and taking values in $\mathbb{R}$, its contamination breakdown point ϵ^* is defined as $\epsilon^* \doteq \inf\{\epsilon > 0 : \, I_\theta(\mathcal{F}) \text{ is bounded}\}$. With this definition at hand, the following result gives a simple way to check **A3** and **A6** for contamination neighborhoods.

Lemma 1. *Let $\mathcal{F}_\epsilon$ be a contamination neighborhood centered at F_0 and assume that $0 < \epsilon < \epsilon^*$, where ϵ^* is the minimum of the contamination breakdown points of the functionals $\theta(F)$ and $\sigma(F)$. Then,*
(a) When there is no nuisance parameter, if $H(t, F_0)$ is continuous and strictly decreasing in t, then assumption **A3** *holds.*
(b) When there is a nuisance parameter, if $\inf_{F\in\mathcal{F}_\epsilon} \sigma(F) > 0$ and $H(t, s, F_0)$ is continuous in (t, s), and strictly decreasing in t for all $s \in I_\sigma(\mathcal{F}_\epsilon)$, then assumption **A6** *holds.*

In the remainder of this section we apply Theorems 1 and 2 together with Lemma 1 to prove the uniform strong consistency over contamination neighborhoods of two classes of estimators.

4.1 Location M-estimators with general estimated scale

Given a score function ψ, Huber (1964) defined a location M-estimator, $\hat{\theta}_n$, as

$$\hat{\theta}_n = \inf\left\{t : \frac{1}{n}\sum_{i=1}^{n} \psi\left(\frac{X_i - t}{\hat{\sigma}_n}\right) < 0\right\}, \tag{11}$$

where $\hat{\sigma}_n$ is a preliminary scale estimator required to obtain scale-equivariant location estimators. We assume the following conditions on ψ:

A8. *The function ψ is increasing, bounded and continuous. Moreover, there exists a finite constant c such that $\psi(x) = 1$, if $x \geq c$, and $\psi(x) = -1$, if $x \leq -c$.*

The application of Theorem 2 leads to the next result:

Theorem 3. *Let $\hat{\theta}_n$ be a location M-estimator based on a score function ψ satisfying* **A8**. *Let $\hat{\sigma}_n$ be a previous auxiliary dispersion estimator satisfying (7). Let ϵ^* be the minimum of the contamination breakdown points of the functionals $\theta(F)$ and $\sigma(F)$ and suppose that $\inf_{F\in\mathcal{F}_\epsilon} \sigma(F) > 0$. Finally, let $\mathcal{F}_\epsilon$ be an ϵ-contamination neighborhood centered at a continuous and strictly increasing distribution F_0 with $\epsilon < \epsilon^*$. Then,*

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}_\epsilon} \mathrm{P}_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta| > \delta\} = 0, \quad \textit{for all } \delta > 0.$$

Notice that the conditions imposed on ψ are fulfilled by many score functions that have been proposed in the literature. In particular, Huber's functions (defined as $\psi_c(x) = \min\{\max\{-1, x/c\}, 1\}$) meet all the assumptions.

When an auxiliary dispersion estimate is not needed (e.g. for sample quantiles) a similar result can be obtained applying Theorem 1. In this case, the continuity of ψ is not necessary. The sufficiency of condition (1) for the strong consistency of sample quantiles -established by Zielinski (1998)- arises as a particular case of this result.

4.2 Dispersion generalized S-estimators

The class of dispersion generalized S-estimators (GS-estimators) was defined by Croux, Rousseeuw and Hössjer (1994) in the context of linear regression. They are defined as

$$\hat{\theta}_n = \inf\left\{t > 0 : \binom{n}{2}^{-1} \sum_{i<j} \chi\left(\frac{X_i - X_j}{t}\right) < b\right\}, \tag{12}$$

for a score function χ and a constant $b \in (0, 1)$. An interesting feature of these estimators is that they do not depend on a specific location estimator since are based exclusively on pairwise differences of data. Hössjer, Croux and Rousseeuw (1994) studied some asymptotic properties of GS-estimators. Their breakdown point is given by $\epsilon^* = \min\{\sqrt{1-b}, 1-\sqrt{1-b}\}$, so that, for $\epsilon < \epsilon^*$, $0 < \inf_{F\in\mathcal{F}_\epsilon} \sigma(F) \le \sup_{F\in\mathcal{F}_\epsilon} \sigma(F) < \infty$.

GS-estimators are another instance of estimators included in the framework of Section 2. Take $h(s, X_1, \ldots, X_n) = \binom{n}{2}^{-1} \sum_{i<j} \chi[(X_i - X_j)/s)] - b$, with expectation $H(t, F) = \mathrm{E}_F\chi[(X - Y)/t] - b$ (here, X and Y are two independent copies of a random variable with distribution F) and $\theta = \inf\{t > 0 : H(t, F) < 0\}$.

We impose the following conditions on χ:

A 9. *$\chi(x)$ is even, bounded and increasing in $(0, \infty)$, with at most a finite number of discontinuities, $\chi(0) = 0$ and $\chi(\infty) \doteq \sup_x \chi(x) = 1$.*

Theorem 4. *Let $\hat{\theta}_n$ be a dispersion GS-estimator based on a constant $b \in (0, 1)$ and a score function χ satisfying* **A9**. *Let $\mathcal{F}_\epsilon$ be an ϵ-contamination neighborhood centered at*

F_0, a strictly increasing, continuous and symmetric [$F_0(-x) = 1 - F_0(x)$] distribution, with $\epsilon < \min\{\sqrt{1-b}, 1-\sqrt{1-b}\}$. Then,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}_\epsilon} P_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta| > \delta\} = 0, \quad \textit{for all } \delta > 0.$$

As a consequence of Theorem 4 (see also the remark after Theorem 2), GS-estimators fulfill condition (7) and, therefore, are a suitable choice as auxiliary scale in (11) to obtain scale equivariant and uniform strong consistent location M-estimators.

5 Further remarks

We have shown that several robust estimators are strongly consistent, uniformly over contamination neighborhoods, provided that the fraction of contamination is less than their breakdown point. When the neighborhood $\mathcal{F}_\epsilon$ is centered at a "nominal" distribution F_0, the interest is very often focused on the estimation of $\theta_0 = \theta(F_0)$, the value of the parameter under the nominal distribution. When the true underlying distribution is $F \in \mathcal{F}_\epsilon$, we have $\hat{\theta}_n \to \theta(F) \neq \theta_0$, so that there is an asymptotic bias $|\theta(F) - \theta_0|$. The maximum asymptotic bias or *maxbias*, $B(\epsilon) \doteq \sup_{F\in\mathcal{F}_\epsilon} |\theta(F) - \theta_0|$, is usually taken as a measure of the robustness of $\hat{\theta}_n$. Although the maxbias is an asymptotic measure of robustness, the results of this paper show its finite sample relevance since, by the triangle inequality,

$$\sup_{m\geq n} |\hat{\theta}_m - \theta_0| \leq \sup_{m\geq n} |\hat{\theta}_m - \theta(F)| + B(\epsilon),$$

and, from (5), for every $\delta > 0$,

$$\lim_{n\to\infty} \sup_{F\in\mathcal{F}_\epsilon} \mathrm{P}_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta_0| > B(\epsilon) + \delta\} = 0.$$

That is, under uniform strong consistency, the maxbias can be understood as an asymptotic upper bound of the finite sample estimation error $\sup_{m\geq n} |\hat{\theta}_m - \theta_0|$, which holds uniformly over the set of possible distributions.

It would be desirable to extend the methods of this article to vector valued estimators that solve systems of estimating equations. This would allow to apply the results in a multivariate context.

Usually, consistency is a condition needed to obtain more enlightening asymptotic properties of the estimators such as their asymptotic distributions. Therefore, the results of this paper could be considered as a starting point to achieve general results on uniform asymptotic distributions of robust statistics.

Appendix: proofs

Proof of Theorem 1: Fix an arbitrary $\delta > 0$. By **A1**, the score function $h(t; X_1, \dots, X_n)$ is decreasing as a function of t. Therefore, for all $F \in \mathcal{F}$,

$$\mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \le \mathrm{P}_F\{h(\theta - \delta; X_1, \dots, X_n) < 0\}. \tag{13}$$

Let $\alpha = \alpha(\delta) \doteq \inf_{F \in \mathcal{F}} H(\theta - \delta, F)$. By **A3**, $\alpha > 0$. Notice that α does not depend on F and that $H(\theta - \delta, F) \ge \alpha > 0$, for all $F \in \mathcal{F}$. Therefore, for all n and $F \in \mathcal{F}$, if $h(\theta - \delta; X_1, \dots, X_n) \le 0$, then

$$|h(\theta - \delta; X_1, \dots, X_n) - H(\theta - \delta, F)| \ge \alpha.$$

As a consequence, for all n and $F \in \mathcal{F}$,

$$\mathrm{P}_F\{h(\theta - \delta; X_1, \dots, X_n) \le 0\} \le \mathrm{P}_F\{|h(\theta - \delta; X_1, \dots, X_n) - H(\theta - \delta, F)| \ge \alpha\}. \tag{14}$$

From (13) and (14), we have that for all n and $F \in \mathcal{F}$,

$$\mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \le \mathrm{P}_F\{|h(\theta - \delta; X_1, \dots, X_n) - H(\theta - \delta, F)| \ge \alpha\}.$$

Since $H(\theta - \delta, F)$ is the expectation under F of $h(\theta - \delta; X_1, \dots, X_n)$, by **A2** we can apply McDiarmid's bounded difference inequality (see, for instance, Devroye and Lugosi (2001), p. 8) to conclude that, for each n, $\mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \le \exp\left(-2\alpha^2 / \sum_{i=1}^n c_{in}^2\right)$. Notice that the right hand side of this bound does not depend on F so that it is actually a uniform upper bound over $\mathcal{F}$:

$$\sup_{F \in \mathcal{F}} \mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \le \exp\left(-\frac{2\alpha^2}{\sum_{i=1}^n c_{in}^2}\right). \tag{15}$$

Using a totally analogous argument, it can be shown the companion inequality

$$\sup_{F \in \mathcal{F}} \mathrm{P}_F\{\hat{\theta}_n > \theta + \delta\} \le \exp\left(-\frac{2\beta^2}{\sum_{i=1}^n c_{in}^2}\right), \tag{16}$$

where $\beta = \beta(\delta) \doteq -\inf_{F\in\mathcal{F}} H(\theta+\delta, F) > 0$ by **A3**. Equations (15) and (16) and the convergence of the series in **A2** imply $\sum_{n=1}^{\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{|\hat{\theta}_n - \theta| > \delta\} < \infty$. Finally, (5) holds since

$$\begin{aligned}\sup_{F\in\mathcal{F}} \mathrm{P}_F\{\sup_{m\geq n} |\hat{\theta}_m - \theta| > \delta\} &\leq \sup_{F\in\mathcal{F}} \sum_{m=n}^{\infty} \mathrm{P}_F\{|\hat{\theta}_m - \theta| > \delta\} \\ &\leq \sum_{m=n}^{\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{|\hat{\theta}_m - \theta| > \delta\} \longrightarrow 0,\end{aligned}$$

as $n \to \infty$ □

Proof of Theorem 2: As in the proof of Theorem 1 we can show that, for any $\delta > 0$,

$$\mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \leq \mathrm{P}_F\{|h(\theta-\delta, \hat{\sigma}_n; X_1, \ldots, X_n) - H(\theta-\delta, \sigma, F)| \geq \alpha\},$$

where $\alpha = \inf_{F\in\mathcal{F}} H(\theta-\delta, \sigma, F) > 0$. As a consequence, by triangle inequality,

$$\begin{aligned}&\mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \leq \mathrm{P}_F\{|h(\theta-\delta, \hat{\sigma}_n; X_1, \ldots, X_n) - h(\theta-\delta, \sigma; X_1, \ldots, X_n)| \geq \alpha/2\} \\ + \;& \mathrm{P}_F\{|h(\theta-\delta, \sigma; X_1, \ldots, X_n) - H(\theta-\delta, \sigma, F)| \geq \alpha/2\} \doteq p_1(F) + p_2(F).\end{aligned}$$

By **A6**, there exist $\underline{t}$, $\overline{t}$, $\underline{s} > 0$ and $\overline{s}$ (all of them independent of F) with $\theta \in [\underline{t}, \overline{t}]$ and $\sigma \in (\underline{s}, \overline{s})$. From **A7**, there exists $\eta > 0$ (which does not depend on F) such that $p_1(F) \leq \mathrm{P}_F\{|\hat{\sigma}_n - \sigma| > \eta\}$. Furthermore, by McDiarmid's inequality and **A5**, $p_2(F) \leq \exp\{-\alpha^2/(2\sum_{i=1}^n c_{in}^2)\}$. Then,

$$\sum_{n=1}^{\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{\hat{\theta}_n < \theta - \delta\} \leq \sum_{n=1}^{\infty} \sup_{F\in\mathcal{F}} \mathrm{P}_F\{|\hat{\sigma}_n - \sigma| > \eta\} + \sum_{n=1}^{\infty} \exp\left(-\frac{\alpha^2}{2\sum_{i=1}^n c_{in}^2}\right) < \infty.$$

A similar inequality can be obtained for $\mathrm{P}_F\{\hat{\theta}_n < \theta + \delta\}$ so that the result follows. □

Proof of Lemma 1: We will only prove part (b) since part (a) is analogous. Fix $F = (1-\epsilon)F_0 + \epsilon\tilde{F} \in \mathcal{F}_\epsilon$ and let $\theta = \theta(F)$ and $\sigma = \sigma(F)$. If $\epsilon < \epsilon^*$, then there exists a compact set $K \subset \mathbb{R}^2$ such that $\{(\theta, \sigma) : F \in \mathcal{F}_\epsilon\} \subset K$. Define $H(\theta-, \sigma, F) \doteq \lim_{\delta\downarrow 0} H(\theta-\delta, \sigma, F)$, the left limit of $H(\cdot, \sigma, F)$ at θ. By monotonicity, $H(\theta-\delta, \sigma, \tilde{F}) - H(\theta-, \sigma, \tilde{F}) \geq 0$ for any distribution $\tilde{F}$. Therefore, for all $\delta > 0$,

$$\begin{aligned}&H(\theta-\delta, \sigma, F) - H(\theta-, \sigma, F) = (1-\epsilon)[H(\theta-\delta, \sigma, F_0) - H(\theta, \sigma, F_0)] \\ + \;& \epsilon[H(\theta-\delta, \sigma, \tilde{F}) - H(\theta-, \sigma, \tilde{F})] \geq (1-\epsilon)[H(\theta-\delta, \sigma, F_0) - H(\theta, \sigma, F_0)] \\ \geq \;& (1-\epsilon) \inf_{(\theta,\sigma)\in K} [H(\theta-\delta, \sigma, F_0) - H(\theta, \sigma, F_0)] \doteq (1-\epsilon)C(F_0, \delta).\end{aligned}$$

By the assumptions, $C(F_0,\delta)$ (which does not depend on F) is the infimum of a strictly positive continuous function over a compact set and, therefore, it is strictly positive. Moreover, by definition of θ, $H(\theta-\delta,\sigma,F)\geq 0$ for all $\delta>0$ and therefore $H(\theta-,\sigma,F)\geq 0$. As a consequence

$$\inf_{F\in\mathcal{F}_\epsilon} H(\theta-\delta,\sigma,F)\geq \inf_{F\in\mathcal{F}_\epsilon} H(\theta-,\sigma,F)+(1-\epsilon)C(F_0,\delta)\geq(1-\epsilon)C(F_0,\delta)>0.$$

An analogous argument works for $-\inf_{F\in\mathcal{F}_\epsilon} H(\theta+\delta,\sigma,F)$. $\square$

Proof of Theorem 3: We have to check **A4**, **A5** and **A6**. Since $\psi(x)$ is increasing and $\sigma(F)>0$ for all $F\in\mathcal{F}_\epsilon$, **A4** trivially holds. Regarding **A5**, since ψ is bounded, we can assume w.l.o.g. that $\psi(x)\leq M/2$ for all $x\in\mathbb{R}$. Then,

$$\begin{aligned}&\sup_{\substack{x_1,\ldots,x_n\\ x_i'\in\mathbb{R}}} |h(t,s;x_1,\ldots,x_i,\ldots,x_n)-h(t,s;x_1,\ldots,x_i',\ldots,x_n)|\\ =\;&\frac{1}{n}\left|\psi\left(\frac{x_i-t}{s}\right)-\psi\left(\frac{x_i'-t}{s}\right)\right|\leq M/n.\end{aligned}$$

Thus, $c_{in}=M/n$ for $i=1,\ldots,n$ and, for $\gamma>0$,

$$\sum_{n=1}^{\infty}\exp\left(\frac{\gamma}{\sum_{i=1}^n c_{in}^2}\right)=\sum_{n=1}^{\infty}\exp\left(-\frac{\gamma n}{M^2}\right)<\infty.$$

Therefore, **A5** also holds. For checking **A6**, notice that since ψ is bounded and continuous, $H(t,s,F_0)$ is also continuous. Moreover, if $t_1<t_2$, then $\psi[(x-t_2)/s]-\psi[(x-t_1)/s]\leq 0$ for all $s\in I_\sigma(\mathcal{F})$ and $x\in\mathbb{R}$ and also, since F_0 is strictly increasing and $\psi(-\infty)<\psi(\infty)$,

$$\mathrm{P}_{F_0}\{\psi[(X-t_2)/s]-\psi[(X-t_1)/s]<0\}>0,\quad\text{for all } s\in I_\sigma(\mathcal{F}_\epsilon).$$

Hence $H(t_2,s,F_0)-H(t_1,s,F_0)<0$, for all $s\in I_\sigma(\mathcal{F}_\epsilon)$. Applying Lemma 1(b) we have that **A6** also hold. Finally, for proving **A7** define $\overline{x}\doteq\overline{t}+\overline{s}c$ and $\underline{x}\doteq\underline{t}-\underline{s}c$. Since we are assuming that $|\psi(x)|=1$ for $|x|\geq c$, then $x<\underline{x}$ or $x>\overline{x}$ implies that $|\psi[(x-t)/s_1]-\psi[(x-t)/s_2]|=0$ for all $s_1,s_2\in(\underline{s},\overline{s})$ and $t\in[\underline{t},\overline{t}]$. Then, we can concentrate on $x\in[\underline{x},\overline{x}]$. In this interval, ψ is uniformly continuous and therefore we just have to bound $|(x-t)/s_1-(x-t)/s_2|$. Now, for $s_1,s_2\in(\underline{s},\overline{s})$, $t\in[\underline{t},\overline{t}]$, and $x\in[\underline{x},\overline{x}]$, it holds

$$\left|\frac{x-t}{s_1}-\frac{x-t}{s_2}\right|=\frac{|x-t|}{s_1s_2}|s_1-s_2|\leq\frac{|\overline{x}|+|\overline{t}|}{\underline{s}^2}|s_1-s_2|\doteq M|s_1-s_2|,$$

and the last quantity is arbitrarily small if $|s_1 - s_2|$ is sufficiently small. □

Proof of Theorem 4: Arguments quite similar to those in the proof of Theorem 3 show that **A1** and **A2** also hold in this case. Regarding **A3**, let F_0^* be the distribution of $X - Y$, where X and Y are independent random variables with distribution F_0 and notice that, under the assumptions on F_0, the distribution F_0^* is continuous, strictly increasing and $F_0^*(-x) = 1 - F_0^*(x)$ hold for all $x \in \mathbb{R}$. Then, since χ is even,

$$H(t, F_0) = 2\int_0^\infty \chi(x/t) dF_0^*(x) - b.$$

Since χ is bounded and has at most a finite number of discontinuities, $H(\cdot, F_0)$ is continuous by Dominated Convergence Theorem. Let $0 < t_1 < t_2$, then $\chi(x/t_2) - \chi(x/t_1) \leq 0$ for all $x > 0$ and also, since F_0^* is strictly increasing, $\mathrm{P}_{F_0^*}\{x > 0 : \chi(x/t_2) - \chi(x/t_1) < 0\} > 0$, and hence

$$H(t_2, F_0) - H(t_1, F_0) = 2\int_0^\infty \big(\chi(x/t_2) - \chi(x/t_1)\big) dF_0^*(x) < 0,$$

that is, $H(\cdot, F_0)$ is strictly decreasing. Applying Lemma 1(a) we get that **A3** also hold. □

References

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

Davies, P.L. (1993). Aspects of robust linear regression. *Ann. Statist.*, **21**, 1843–1899.

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

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

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

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

McDiarmid, C. (1989). On the method of bounded differences, in: *Surveys in Combinatorics 1989,* pp. 148–188, Cambridge University Press, Cambridge.

Salibian-Barrera, M. and Zamar, R.H. (2001). Uniform asymptotics for robust location estimates when the scale is unknown. Unpublished manuscript available at `http://mathstat.carleton.ca/~matias/pubs.html`.

Shorack, G.R. (2000). *Probability for Statisticians.* Springer, New York.

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

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