Ann. Inst. Statist. Math.
Vol. 56, No. 4, 771–790 (2004)


# ANOTHER APPROACH TO ASYMPTOTICS AND BOOTSTRAP OF RANDOMLY TRIMMED MEANS

ZHIQIANG CHEN[1]* AND EVARIST GINÉ[2]**

[1] *Department of Mathematics, William Paterson University, Wayne, NJ 07470, U.S.A.,*
e-mail: ChenZ@wpunj.edu
[2] *Department of Mathematics, U-3009, University of Connecticut, Storrs, CT 06269, U.S.A.,*
e-mail: gine@math.uconn.edu

(Received February 17, 2003; revised November 11, 2003)

**Abstract.** A unified, empirical processes based approach to the central limit theorem and to the bootstrap for randomly trimmed and Winsorized means is developed, with emphasis on Hampel's means.

*Key words and phrases*: Randomly trimmed means, median absolute deviation.

## 1. Introduction

Since the sample mean is very sensitive to outliers in the data, alternative estimators (and location parameters) have been in use since time immemorial. Randomly trimmed means, which include as extremes the mean and the median, constitute a very general class of location parameters whose sample counterparts that can be made as sensitive or insensitive to outliers as one wishes, and also as asymptotically efficient with respect to the sample mean as one wishes when considered as estimators of the center of symmetry of a smooth symmetric distribution. A randomly trimmed mean is constructed as follows: one fixes levels $a_n < b_n$ depending on the sample, and takes the average of those data points that fall between $a_n$ and $b_n$. Typically these levels are obtained by evaluation at the empirical distribution function $F_n$ of functions $a(F)$, $b(F)$ defined on probability laws; then, if $F$ is the distribution function from which the data are drawn, and one takes $a_n = a(F_n)$ and $b_n = b(F_n)$, the trimmed mean estimates the expectation of $X$ (with distribution $F$) conditioned to $X \in [a(F), b(F)]$. If $a(F) = F^{-1}(\alpha/2)$ and $b(F) = F^{-1}(1-\alpha/2)$, one has the classical trimming, with a breakdown point of $\alpha$. If $a(F) = \mu(F) - cs(F)$ and $b(F) = \mu(F) + cs(F)$, where $\mu$ denotes median, $s$ denotes median absolute deviation (MAD), and $c$ is a constant of choice, then one has Hampel's trimmed mean (Hampel (1971, 1985)), which has the optimal breakdown point of $1/2$; in this case one does not know in advance what proportion of the data are being discarded (which may have some advantages). These are just two examples, but there are many more. A companion to trimmed means are Winsorized means, which are the plug-in estimators of the population parameters $E((X \wedge b) \vee a)$.

---

*Research partially supported by the Center for Research, College of Science and Health, and by the ART program of the William Paterson University.

**Research partially supported by NSF Grant No. DMS–0070382.

It seems advantageous to have a unified approach at getting robustness properties, limit distributions and the validity of the bootstrap for this relatively general and very useful class of statistics. Shorack (1974) has a unifed approach to the central limit theorem in the case of symmetric distributions; Shorack and Wellner (1986) have asymptotic expressions for randomly trimmed and Winsorized means that fall short from being limit theorems because the levels $a$ and $b$ are, in a sense, too general. Kim (1992) proposes a type of trimmed means (not unrelated to Hampel's) for which he obtains the central limit theorem using the delta method. Hall and Padmanahban (1992) contains a detailed study of the bootstrap for the studentized classical trimmed mean. In this article we present a unified, empirical process based approach to the central limit theorem and the bootstrap central limit theorem for the general trimmed mean under very mild assumptions on the levels $a(F)$ and $b(F)$. The basic assumption is that $\sqrt{n}(a_n - a)$ can be asymptotically linearized, in the sense that it is asymptotically equivalent to $\sum_{i=1}^{n}(h_1(X_i) - Eh_1(X))/\sqrt{n}$ for some function $h_1$ square integrable for $F$, and likewise for $\sqrt{n}(b_n - b)$. This assumption is very natural because on one hand it allows for very simple proofs, and on the other, it is satisfied by the median, any quantiles and the MAD (and hence also by $\mu - cs$), as we show below. Robustness properties are also breifly reviewed.

Then, we apply the general results to Hampel's means and to means based on box plots (precise definitions are given in the next section). We emphasize Hampel's means because they have optimal breakdown point, and, as Davies (1998) concludes after commenting on their simplicity and favorably comparing the performance of Hampel's trimmed mean and standard deviation with that of other robust location/scale parameters, 'It (Hmean/Hsdv) can certainly be recommended for a first course in data analysis as an alternative to the usual mean/sdv.' (Here we do not deal with Hampel's standard deviation—the standard deviation of the data between $\mu_n - cs_n$ and $\mu_n + cs_n$.) Also, as estimators of the center of symmetry of a smooth symmetric distribution, the asymtotic efficiency of Hampel's trimmed and Winsorized sample means with respect to the sample mean tends to one as the tuning parameter $c$ tends to infinity, whereas the breakdown point remains $1/2$ regardless of the value of $c$.

Section 2 contains definitions and examples, Section 3 reviews robustness properties, in Section 4 we obtain asymptiotic normality and in Section 5 we show that the bootstrap is valid a.s.

## 2. Definitions

Let $X, X_1, \ldots, X_n, \ldots$ be indepedent identically distributed real random variables with common probability law $P$ and, for each $n \in \boldsymbol{N}$, let $P_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i}$ be the empirical measure corresponding to the first $n$ observations $X_1, \ldots, X_n$. $F$ and $F_n$ will denote the cumulative distribution functions associated respectively to $P$ and $P_n$. Let $a(Q) = a(F_Q)$ and $b(Q) = b(F_Q)$ be real valued functionals defined on a subset of the set of probability measures $Q$ on $\boldsymbol{R}$ containing $P$ and $P_n(\omega)$ for all $n \in \boldsymbol{N}$ and $\omega \in \Omega$, and such that $a \leq b$. Then, the *trimmed mean of $P$ based on $a$ and $b$* is defined as

$$\theta = \theta(P) := \frac{\int_{a(P)}^{b(P)} x dF(x)}{F(b(P)) - F(a(P)-)} = E(X \mid X \in [a(P), b(P)]),$$

and we will usually write $a$ and $b$ for $a(P)$ and $b(P)$. Likewise, we will write $a_n$ for $a(P_n)$ and $b_n$ for $b(P_n)$, and we impose without further mention that $a_n$ and $b_n$ are measurable functions, that is, random variables. The *empirical trimmed mean based on a and b* is analogously defined as

$$\theta_n = \theta(P_n) = \frac{\sum_{i=1}^n X_i I_{[a_n,b_n]}(X_i)}{\sum_{i=1}^n I_{[a_n,b_n]}(X_i)} = \frac{\int_{a_n}^{b_n} x dF_n(x)}{F_n(b_n) - F_n(a_n-)}.$$

Here and elsewhere, $\int_c^d = \int_{[c,d]}$. The *Winsorized population and empirical means based on a and b* are respectively defined as

$$\xi = \xi(P) = \int_a^b x dF(x) + aF(a-) + b(1 - F(b)) = E((X \wedge b) \vee a)$$

and

$$\xi_n = \xi_n(P) = \int_{a_n}^{b_n} x dF_n(x) + a_n F_n(a_n-) + b_n(1 - F_n(b_n))$$
$$= \frac{1}{n}\sum_{i=1}^n [X_i I_{[a_n,b_n]}(X_i) + b_n I_{(b_n,\infty)}(X_i) + a_n I_{(-\infty,a_n)}(X_i)].$$

Most trimmed and Winsorized means in the literature are obtained by appropriately choosing $a$ and $b$. Given a distribution function $F$ and a number $\alpha \in (0,1)$ define, as usual, $Q_\alpha(F) = F^{-1}(\alpha) := \inf\{x : F(x) \geq \alpha\}$, the $\alpha$-quantile of $F$ (quartiles: $\alpha = 1/4, 3/4$, median: $\alpha = 1/2$). We will denote by $\mu$ the population median $Q_{1/2}(F)$ and by $\mu_n$ the sample median $Q_{1/2}(F_n)$; likewise, we simplify the notation for quartiles as $Q_1 = Q_{1/4}(F)$, $Q_{n,1} = Q_{1/4}(F_n)$, $Q_3 = Q_{3/4}(F)$ and $Q_{n,3} = Q_{3/4}(F_n)$, and the interquartile ranges are denoted by $R = Q_3 - Q_1$, $R_n = Q_{n,3} - Q_{n,1}$. Finally, the *median absolute deviation* or MAD $s$ for $P$ is defined as the median of the random variable $|X - \mu|$ (where $X$ has law $P$), and the empirical or sample MAD, $s_n$, as the median of the (random) set of points $|X_1 - \mu_n|, \ldots, |X_n - \mu_n|$.

*Examples.*

1. *Hampel's means.* (*Hampel* (1971, 1985)). The trimmed and Winsorized Hampel means are obtained by taking $a = \mu - cs$ and $b = \mu + cs$ for some $c > 0$ (Hampel takes $c = 5.2$). So, the Hampel means are averages of the data points that are close to the (sample) median by at most a fixed multiple of the (sample) MAD. These means are remarkable because they inherit the excellent robustness properties of the median and the MAD. Their asymptotic normality in the symmetric case can be obtained from Shorack (1974), Example 9, given the asymptotic normality of the MAD, that was proved by Hall and Welsh (1985). We are not aware of any proofs in the literature of the asymptotic normality of these estimators in general (without assuming symmetry of $F$) or of the validity of the bootstrap for them.

2. *Box plot means.* The box plot trimmed and Winsorized means are obtained by taking $a = Q_1 - cR$ and $b = Q_3 + cR$ for some $c \geq 0$, or what is the same, $a = (Q_1 + Q_3)/2 - dR$, $b = (Q_1 + Q_3)/2 + dR$ for some $d \geq 1/2$. Asymptotic normality in the symmetric case follows from Shorack (1974). We call these box plot means for obvious reasons. We could take other quantiles as well in the definition of these trimmed means.

3. *Symmetrically trimmed means.* For these means, one takes $a = Q_\alpha$ and $b = Q_{1-\alpha}$. Classical, well studied (for asymptotic normality, see e.g., Huber (1969) and Stigler (1973), and for the bootstrap, Hall and Padmanabhan (1992)).

4. *Kim's metrically trimmed means.* This (Kim (1992)) is obtained by taking $a = \mu - F^{-1}_{|X-\mu|}(1-\alpha)$ and $b = \mu + F^{-1}_{|X-\mu|}(1-\alpha)$. For $\alpha = 1/2$ Kim's metrically trimmed mean coincides with Hampel's with the choice of $c = 1$. The asymptotic normality when '$F$ possesses a density $f$ which is positive and absolutely continuous on its support' was obtained by Kim (1992); in the symmetric case it does follow from Shorack (1974) given the asymptotic normality of $a_n$ and $b_n$, that was obtained later. The results in the next sections imply that Kim's theorem is true under the weaker assumption that the density exists in an open set containing $\mu$, $a$ and $b$ and is positive and continuous there; they also imply that the bootstrap works as well. However, these results will neither be stated nor proved because of their similarity with the corresponding results for Hampel's means.

## 3. Robustness

It is obvious that if $a$ and $b$, as functions of $P$, are equivariant under affine transformations, then so are $\theta_n$ and $\xi_n$, and this is a basic property to have for location parameters and estimators. It is also interesting to note that if $P$ is symmetric about a point $c$, then $\theta(P) = \xi(P) = c$ and $\theta_n$ and $\xi_n$ are unbiased estimators of the center of symmetry $c$. Although these are desirable and important properties, randomly trimmed means are most appealing because of their robustness properties. We look now a little more closely at robustness in terms of breakdown points and influence curves.

We recall (Hampel (1974)) that the finite sample breakdown point of an estimator is 'the smallest percentage of free contamination that can carry the value of the estimator over any bounds'.

PROPOSITION 3.1. *The finite sample breakdown points of $\theta_n$ and $\xi_n$, defined from $a_n$ and $b_n$ as in Section 2, are the smallest of the breakdown points of $a_n$ and $b_n$. In particular, Hampel's trimmed and Winsorized means have breakdown points of* 50%, *whereas the box plot trimmed and Winsorized means have breakdown points of* 25%.

PROOF. For any portion of contamination points that keep $a_n$ and $b_n$ unbroken (bounded), the estimators $\theta_n$ and $\xi_n$ also remain bounded by their very definition. Conversely, if a portion of contamination can take $a_n$ or $b_n$ beyond any bound, then it can also take $\theta_n$ and $\xi_n$ beyond any bound, proving the general result. The result for Hampel's and box plot means follows from this because, as is well known, the breakdown points of the median and the MAD are both equal to 1/2, and that of the quartiles is 1/4. □

The influence curve or function at $P$ of a parameter $\theta = \theta(P)$ is defined (Hampel (1974)) as

$$IC(x, \theta, P) = \lim_{\varepsilon \to 0} \frac{\theta(P(\varepsilon, x)) - \theta(P)}{\varepsilon},$$

where $P(\varepsilon, x) := (1-\varepsilon)P + \varepsilon\delta_x$, $\varepsilon \in [0, 1]$. The next proposition follows by direct simple computations that we omit.

PROPOSITION 3.2. *Assume the distribution function $F$ of $P$ has a continuous derivative $f$ on an open set containing $a = a(P)$ and $b = b(P)$. Let $a_x(\varepsilon) := a(P(\varepsilon, x))$ and $b_x(\varepsilon) := b(P(\varepsilon, x))$, and assume $a_x$ and $b_x$ are differentiable (with respect to $\varepsilon$) at* *zero, with derivatives* $a'_x$ *and* $b'_x$. *If* $\theta = \theta(P)$ *and* $\xi = \xi(P)$ *are defined from* $a$ *and* $b$ *as in Section* 2, *we have*:

$$IC(x, \theta, P) = \frac{bb'_x f(b) - aa'_x f(a) + xI_{[a,b]}(x)}{F(b) - F(a)} - \frac{b'_x f(b) - a'_x f(a) + I_{[a,b]}(x)}{F(b) - F(a)} \cdot \theta,$$

*and, for* $x \neq a, b$,

$$IC(x, \xi, P) = a'_x F(a) + b'_x(1 - F(b)) + xI_{[a,b]}(x) + aI_{(-\infty,a)}(x) + bI_{(b,\infty)}(x) - \xi.$$

*In particular, these parameters have bounded influence functions.*

The previous proposition gives the influence functions of the Hampel and the box plot means just by replacing $a$ and $b$ by their values in terms of the median, the MAD and the quartiles, whose influence functions are (Hampel (1974), Huber (1981) or direct computation):

$$IC(x, Q_\alpha, P) = \begin{cases} 0 & \text{if} \quad x = Q_\alpha \\ \dfrac{\alpha - I_{(-\infty, Q_\alpha)}(x)}{f(Q_\alpha)} & \text{otherwise} \end{cases}$$

and

$$IC(x, s, P) = \frac{1/2 - I_{(\mu - s, \mu + s]}(x) - [f(\mu + s) - f(\mu - s)]IC(x, \mu, P)}{f(\mu + s) + f(\mu - s)}.$$

## 4. Limit distributions

Given $X, X_i$, $i \in \boldsymbol{N}$, we let $P$ be their common law, with c.d.f. $F$ and density $f$, and, for each $n \in \boldsymbol{N}$, we let $P_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$ be the empirical measure and $\nu_n := \sqrt{n}(P_n - P)$ the empirical process. In this notation, the classical empirical process is $\nu_n(-\infty, x] = \sqrt{n}(F_n(x) - F(x))$. Given a measure $\mu$ (in paticular, $P$, $P_n$ or $\nu_n$) and an integrable function $f$, we will often write $\mu(f)$ for $\int f d\mu$.

### 4.1 *General trimmed and Winsorized means*

Let $-\infty < a < b < \infty$ and let $a_n$, $b_n$ be random variables such that $-\infty < a_n \leq b_n < \infty$ a.s.. The following assumptions will be in force in this subsection:

(D.1) The c.d.f. $F$ has a derivative $f$ on an open set containing $a$ and $b$ and $f$ is continuous there, hence, $f$ is uniformly continuous on a compact set $K$ whose interior contains $a$ and $b$.

(D.2) $f(a) + f(b) \neq 0$.

(L) There exist measurable, $P$-square integrable functions $h_1$ and $h_2$ such that

$$\sqrt{n}(a_n - a) = \nu_n(h_1) + o_P(1) \quad \text{and} \quad \sqrt{n}(b_n - b) = \nu_n(h_2) + o_P(1). \tag{4.1}$$

In particular, by the central limit theorem, $a_n$ and $b_n$ are weakly consistent (that is, $a_n \to a$ and $b_n \to b$ in probability) and asymptotically normal (when centered at $a$, resp. $b$, and multiplied by $\sqrt{n}$). Note that, although fairly general, these assumptions are slightly stronger than the assumptions in Theorems 1 and 2, Shorack and Wellner (1986), pp. 678 and 682. But, on one hand, they will allow us to get stronger results than in their theorems and, on the other, they are satisfied by all the examples in Section 2.

LEMMA 4.1. *Assume* (D.1) *and* (L). *Define, for* $x \in \boldsymbol{R}$,

$$g_1(x) = I_{[a,b]}(x) + f(b)h_2(x) - f(a)h_1(x)$$

*and*

$$g_2(x) = xI_{[a,b]}(x) + bf(b)h_2(x) - af(a)h_1(x).$$

*Then*,

$$\sqrt{n}\left[\int_{a_n}^{b_n} dF_n(x) - \int_a^b dF(x)\right] = \nu_n(g_1) + o_P(1) \tag{4.2}$$

*and*

$$\sqrt{n}\left[\int_{a_n}^{b_n} xdF_n(x) - \int_a^b xdF(x)\right] = \nu_n(g_2) + o_P(1). \tag{4.3}$$

PROOF. We have

$$\sqrt{n}\left[\int_{a_n}^{b_n} dF_n - \int_a^b dF\right] = \sqrt{n}\int_{a_n}^{b_n} d(F_n - F) + \sqrt{n}\int_{a_n}^{a} dF + \sqrt{n}\int_{b}^{b_n} dF.$$

Since $f$ is uniformly bounded on $K$ we have that $\|f\|_K := \sup_{x\in K} f(x)$ is finite, and that if $[x,y] \subseteq K$ then $P[x,y] \le \|f\|_K(y-x)$, hence, $P[x,y] - (P[x,y])^2 \le \|f\|_K(y-x)$. Also, there is $\delta_0 > 0$ such that $[a-\delta_0, a+\delta_0] \cup [b-\delta_0, b+\delta_0] \subseteq K$. Therefore, by the asymptotic equicontinuity of the classical empirical process (e.g., Theorem 3.7.2 and Corollary 6.3.17, pp. 118 and 215, in Dudley (1999)), we have

$$\lim_{\delta\to 0}\limsup_n \Pr\left\{\sup_{|x-a|\le\delta, |y-b|\le\delta} |\nu_n[x,y] - \nu_n[a,b]| \ge \varepsilon\right\} = 0 \tag{4.4}$$

for all $\varepsilon > 0$. Since

$$\Pr\left\{\left|\sqrt{n}\int_{a_n}^{b_n} d(F_n - F) - \nu_n[a,b]\right| \ge \varepsilon\right\} \le \Pr\left\{\sup_{|x-a|\le\delta, |y-b|\le\delta} |\nu_n[x,y] - \nu_n[a,b]| \ge \varepsilon\right\}$$
$$+ \Pr\{|b_n - b| > \delta\} + \Pr\{|a_n - a| > \delta\},$$

the equicontinuity condition and condition (L) give, upon taking limits first as $n \to \infty$ and then as $\delta \to 0$,

$$\sqrt{n}\int_{a_n}^{b_n} d(F_n - F) = \nu_n[a,b] + o_P(1).$$

Given $0 < \delta \le \delta_0$ and $M_n$, $n \in \boldsymbol{N}$, such that $M_n \to \infty$ and $M_n/\sqrt{n} \to 0$, set $\tau_n := \sup\{|f(b) - f(c)| : |b-c| \le (\delta + M_n)/\sqrt{n}\}$, which tends to zero by uniform continuity of $f$ on $K$. Then, on the event where $|\nu_n(h_2) - \sqrt{n}(b_n - b)| \le \delta$ and $|\nu_n(h_2)| \le M_n$, we have

$$\left|\sqrt{n}\int_b^{b_n} dF - f(b)\nu_n(h_2)\right| \le \sqrt{n}\int_b^{b_n} |f(x) - f(b)|dx + |\sqrt{n}(b_n - b) - \nu_n(h_2)|f(b)$$
$$\le \sqrt{n}|b_n - b|\tau_n + |\sqrt{n}(b_n - b) - \nu_n(h_2)|f(b),$$

and this bound tends to zero in probability. So, for any $\varepsilon > 0$,

$$\begin{aligned}\Pr\left\{\left|\sqrt{n}\int_b^{b_n} dF - f(b)\nu_n(h_2)\right| > \varepsilon\right\} \\ &\leq \Pr\{\sqrt{n}|b_n - b|\tau_n + |\sqrt{n}(b_n - b) - \nu_n(h_2)|f(b) > \varepsilon\} \\ &\quad + \Pr\{|\nu_n(h_2) - \sqrt{n}(b_n - b)| > \delta\} + \Pr\{|\nu_n(h_2)| > M_n\} \to 0,\end{aligned}$$

proving

$$\sqrt{n}\int_b^{b_n} dF = f(b)\nu_n(h_2) + o_P(1).$$

A similar proof gives

$$\sqrt{n}\int_{a_n}^{a} dF = -f(a)\nu_n(h_1) + o_P(1).$$

Collecting terms we obtain (4.2). To prove (4.3) we write

$$\sqrt{n}\left[\int_{a_n}^{b_n} x dF_n - \int_a^b x dF\right] = \sqrt{n}\int_{a_n}^{b_n} x d(F_n - F) + \sqrt{n}\int_{a_n}^{a} x dF + \sqrt{n}\int_b^{b_n} x dF$$

and then proceed as in the proof of (4.2) (we note that the asymptotic equicontinuity condition also holds for the empirical process indexed by the class of functions $xI_{[c,d]}(x)$, $|c-a| \leq \delta_0$, $|d-b| \leq \delta_0$, also by Theorem 3.7.2 and Corollary 6.3.17 in Dudley (1999)). □

THEOREM 4.1. *Assume* (D.1), (D.2) *and* (L), *and set*

$$g(x) := \frac{1}{\int_a^b dF(t)} g_2(x) - \frac{\int_a^b t dF(t)}{(\int_a^b dF(t))^2} g_1(x), \qquad x \in \boldsymbol{R},$$

*with $g_1$ and $g_2$ as defined in Lemma* 3.1. *Let $\theta_n$ be the trimmed mean based on $a_n$ and $b_n$ and let $\theta$ be its population counterpart, as defined in Section* 2. *Then,*

$$\sqrt{n}(\theta_n - \theta) \to \sqrt{\mathrm{Var}_F(g)}Z$$

*in distribution, where $Z$ is standard normal.*

PROOF. By (D.2), $\int_a^b dF > 0$ and, with probability tending to 1 by (4.2), $\int_{a_n}^{b_n} dF_n > 0$. We then have

$$\begin{aligned}\sqrt{n}(\theta_n - \theta) &= \sqrt{n}\left[\frac{\int_{a_n}^{b_n} x dF_n}{\int_{a_n}^{b_n} dF_n} - \frac{\int_a^b x dF}{\int_a^b dF}\right] \\ &= \left(\frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n}\right)\frac{\sqrt{n}(\int_a^b dF\int_{a_n}^{b_n} x dF_n - \int_{a_n}^{b_n} dF_n\int_a^b x dF)}{(\int_a^b dF)^2} \\ &:= (I) \times (II).\end{aligned}$$

By (4.2),

$$\left|\frac{\int_{a_n}^{b_n} dF_n}{\int_a^b dF} - 1\right| = \frac{1}{\int_a^b dF}\left|\int_{a_n}^{b_n} dF_n - \int_a^b dF\right| \to 0 \quad \text{in pr.}$$

Hence also

$$\left| \frac{\int_a^b dF}{\int_{a_n}^{b_n} dF_n} - 1 \right| \to 0 \qquad \text{in pr.,}$$

that is,

$$(I) = 1 + o_P(1).$$

And, also by Lemma 4.1,

$$(II) = \nu_n(g) + o_P(1).$$

Now the theorem follows because, by the central limit theorem, $\nu_n(g) \to \sqrt{\mathrm{Var}_F(g)} Z$ in distribution. □

Next we establish the asymptotic normality of the Winsorized means $\xi_n$.

LEMMA 4.2. *Under* (D.1) *and* (L), *we have both,*

$$\sqrt{n}\left(\frac{b_n \sum_{i=1}^n I_{(b_n,\infty)}(X_i)}{n} - b(1-F(b))\right) = \nu_n(bI_{(b,\infty)} + [1 - F(b) - bf(b)]h_2) + o_P(1)$$

*and*

$$\sqrt{n}\left(\frac{a_n \sum_{i=1}^n I_{(-\infty,a_n)}(X_i)}{n} - aF(a)\right) = \nu_n(aI_{(-\infty,a)} + [F(a) + af(a)]h_1) + o_P(1).$$

PROOF. If we write

$$\sqrt{n}\left(\frac{b_n \sum_{i=1}^n I_{(b_n,\infty)}(X_i)}{n} - b(1-F(b))\right) = \sqrt{n}(b_n - b)\int_{b_n}^{\infty} dF_n + b\nu_n(b_n,\infty) + \sqrt{n} b \int_{b_n}^{b} dF$$

and

$$\sqrt{n}\left(\frac{a_n \sum_{i=1}^n I_{(-\infty,a_n)}(X_i)}{n} - aF(a)\right) = \sqrt{n}(a_n - a)\int_{-\infty}^{a_n} dF_n + a\nu_n(-\infty,a_n) + \sqrt{n} a \int_{a}^{a_n} dF,$$

and then apply the Glivenko-Cantelli theorem together with (L) to the first term at the right hand side of each of these two identities, asymptotic equicontinuity of the empirical process to the second, and (D.1) and (L) to the third, the result follows. □

THEOREM 4.2. *Assume* (D.1), (D.2) *and* (L), *and set*

$$w(x) := xI_{[a,b]}(x) + bI_{(b,\infty)}(x) + aI_{(-\infty,a)}(x) + (1 - F(b))h_2(x) + F(a)h_1(x), \qquad x \in \boldsymbol{R}.$$

*Let $\xi_n$ be the Winsorized mean based on $a_n$ and $b_n$ and let $\xi$ be its population counterpart, as defined in Section 2. Then,*

$$\sqrt{n}(\xi_n - \xi) \to \sqrt{\mathrm{Var}_F(w)} Z$$

*in distribution, where $Z$ is standard normal.*

PROOF. Since

$$\sqrt{n}(\xi_n - \xi) = \sqrt{n}\left[\int_{a_n}^{b_n} x dF_n - \int_a^b x dF\right] + \sqrt{n}\left[a_n \int_{-\infty}^{a_n} dF_n - aF(a)\right] + \sqrt{n}\left[b_n \int_{b_n}^{\infty} dF_n - b(1 - F(b))\right],$$

the result follows from Lemmas 4.1 and 4.2 together with the central limit theorem. □

With somewhat different assumptions, namely, that $a$ and $b$ are Hadamard differentiable tangentially to the distribution functions with non-vanishing densities at $a(P)$ and $b(P)$, and that $a_x(\varepsilon)$, $b_x(\varepsilon)$ are differentiable at zero, Theorems 4.1 and 4.2 may be deduced by the delta method together with Proposition 2.2 (see, e.g., van der Vaart and Wellner (1996)). This was the approach taken by Kim (1992) for metrically trimmed means, assuming much more regularity than is needed with our approach, which, moreover, is more direct and elementary. This comment applies as well to the remaining results in this section and the next.

4.2 *Hampel's trimmed and Winsorized means*

To apply the results from the previous section to Hampel's trimmed mean, it suffices to check that $a_n = \mu_n - cs_n$ and $b_n = \mu_n + cs_n$, where $\mu_n$ and $s_n$ are respectively the sample median and the sample MAD and $c > 0$ is a fixed constant, satisfy condition (L) for appropriate functions $h_1$ and $h_2$. This will be a consequence of the following lemma, which may have some independent interest.

LEMMA 4.3. *Assuming the c.d.f. $F$ has a continuous derivative $f$ on an open set containing $\mu$, $\mu + s$ and $\mu - s$, and that $f(\mu) > 0$ and $f(\mu + s) + f(\mu - s) > 0$, we have:*

$$\mu_n \to \mu \quad a.s., \qquad s_n \to s \quad a.s., \tag{4.5}$$

$$\sqrt{n}(\mu_n - \mu) = -\frac{1}{f(\mu)}\nu_n(-\infty, \mu] + o_P(1) \tag{4.6}$$

*and*

$$\sqrt{n}(s_n - s) = -\frac{1}{f(\mu + s) + f(\mu - s)}\nu_n[\mu - s, \mu + s] + \frac{1}{f(\mu)}\frac{f(\mu + s) - f(\mu - s)}{f(\mu + s) + f(\mu - s)}\nu_n(-\infty, \mu] + o_P(1). \tag{4.7}$$

PROOF. It is well known that $\mu_n \to \mu$ a.s. and that the sequence $\{\sqrt{n}(\mu_n - \mu)\}$ is stochastically bounded, in fact, asymptotically normal with mean zero and variance $(2f(\mu))^{-2}$ (e.g., Pollard (1984), pp. 7 and 53). Next we prove (4.6) by adapting an argument from Pollard (1984), p. 98 (which we basically repeat because it is also used in the proof of (4.7)). The existence and continuity of $f$ near $\mu$ and the convergence of $\mu_n$ to $\mu$ (in pr. suffices) then give

$$F(\mu_n) = \frac{1}{2} + \int_{\mu}^{\mu_n} dF(x) = \frac{1}{2} + (\mu_n - \mu)(f(\mu) + o_P(1))$$

and

$$F_n(\mu_n) = \frac{1}{2} + O(1/n) \qquad \text{a.s.}$$

(note that the sample points in a neighborhood of $\mu$, hence eventually of $\mu_n$, are all different with probability 1). Then, by subtracting, the stochastic boundedness of the sequence $\{\sqrt{n}(\mu_n - \mu)\}$ give

$$\sqrt{n}(F_n - F)(\mu_n) = -\sqrt{n}(\mu_n - \mu)(f(\mu) + o_P(1)) + o_P(1) = -\sqrt{n}(\mu_n - \mu)f(\mu) + o_P(1).$$

Continuity of $f$, and the asymptotic equicontinuity condition for the processes $\sqrt{n}(F_n - F)(y) = \nu_n(-\infty, y]$, $-\infty < y < \infty$, (e.g., results cited above from Dudley (1999)), give, just as in (4.4),

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{\sup_{\substack{x,y:0\le x-y<\delta,\\ |x-a|<\delta,|y-a|<\delta}}|\nu_n[x,y]| > \varepsilon\right\} = 0 \qquad \text{for all} \quad \varepsilon > 0, \tag{4.8}$$

for $a = \mu, \mu + s, \mu - s$. Then, (4.8) together with the fact that $\mu_n - \mu \to 0$ in probability (actually, a.s.) finally give

$$\sqrt{n}(F_n - F)(\mu) = -\sqrt{n}(\mu_n - \mu)f(\mu) + o_P(1),$$

which is (4.6). (To see that $\sqrt{n}(F_n - F)(\mu_n) = \sqrt{n}(F_n - F)(\mu) + o_P(1)$ just consider that for all $\varepsilon > 0$ and $\delta > 0$,

$$\Pr\{|\nu_n(-\infty,\mu_n) - \nu_n(-\infty,\mu)| > \varepsilon\} \le \Pr\left\{\sup_{|x-\mu|<\delta}|\nu_n(-\infty,x) - \nu_n(-\infty,\mu)| > \varepsilon\right\} + \Pr\{|\mu_n - \mu| > \delta\}$$

and apply (4.8) and that $\mu_n \to \mu$.)

Next we prove the second part of (4.5) and (4.7). If $\tilde{s}_n$ is the median of the set $|X_1 - \mu|, \ldots, |X_n - \mu|$, by the definition of $s_n$ we have

$$|s_n - s| \le |s_n - \tilde{s}_n| + |\tilde{s}_n - s| \le |\mu_n - \mu| + |\tilde{s}_n - s|,$$

and, $s$ being the median of $Y = |X - \mu|$ and $\tilde{s}_n$ its empirical counterpart, it follows that $\tilde{s}_n \to s$ a.s. and that $\sqrt{n}(\tilde{s}_n - s)$ is asymptotically centered normal with variance $(2g(s))^{-2}$, where

$$g(x) := f(\mu + x) + f(\mu - x), \qquad x \ge 0, \tag{4.9}$$

is the density of $|X - \mu|$, for $x \ge 0$ in a neighborhood of $s$. So,

$$s_n - s \to 0 \quad \text{a.s.} \quad \text{and} \quad \sqrt{n}(s_n - s) = O_P(1). \tag{4.10}$$

In particular, the limits (4.5) hold. Then, using (4.10) and letting $\bar{F}$ be the c.d.f. of $|X - \mu|$ and $\tilde{F}_n$ be the empirical c.d.f. for the data $|X_i - \mu_n|, \ldots, |X_n - \mu_n|$, arguing as in the first part of the proof of (4.6) (but now comparing $\tilde{F}_n(s_n)$ with $\bar{F}_n(s_n)$), we get

$$\sqrt{n}(s_n - s) = -\frac{\sqrt{n}(\tilde{F}_n - \bar{F})(s_n)}{g(s)} + o_P(1). \tag{4.11}$$

Next, using (4.8) and (4.10),

$$\begin{aligned}\sqrt{n}(\tilde{F}_n - \bar{F})(s_n) &= \sqrt{n}(P_n[\mu_n - s_n, \mu_n + s_n] - P[\mu - s_n, \mu + s_n]) \\ &= \nu_n[\mu_n - s_n, \mu_n + s_n] + \sqrt{n}(P[\mu_n - s_n, \mu_n + s_n] - P[\mu - s_n, \mu + s_n]) \\ &= \nu_n[\mu - s, \mu + s] + (f(\mu + s) - f(\mu - s))\sqrt{n}(\mu_n - \mu) + o_P(1).\end{aligned}$$

Now, (4.7) follows from this, (4.6), (4.11) and $g(s) \neq 0$. □

*Remark.* The asymptotic normality of $\sqrt{n}(s_n - s)$ was obtained by Hall and Welsh (1985), and Falk (1997) gave the asymptotic joint distribution of $\sqrt{n}(\mu_n - \mu, s_n - s)$, from where he deduced, in particular, that in the symmetric case $\mu_n$ and $s_n$ are asymptotically independent. Of course Lemma 4.1 provides another, shorter, proof of Falk's result: by the bivariate central limit theorem, this lemma readily implies that the limiting joint distribution of $\sqrt{n}(\mu_n - \mu, s_n - s)$ is centered normal with covariance

$$\mathrm{Cov}_F \left\{ -\frac{I_{(-\infty,\mu]}}{f(\mu)}, -\frac{I_{[\mu-s,\mu+s]}}{f(\mu + s) + f(\mu - s)} + \frac{(f(\mu + s) - f(\mu - s))I_{(-\infty,\mu]}}{f(\mu)(f(\mu + s) + f(\mu - s))} \right\},$$

which is Falk's result. Then, one immediately sees that the off-diagonal terms of this covariance matrix, in the case when $F$ is the distribution of a probability law symmetric about $\mu$, are zero because $E(I_{(-\infty,\mu]}I_{[\mu-s,\mu+s]}) = 1/4 = E(I_{(-\infty,\mu]})E(I_{[\mu-s,\mu+s]})$ and $f(\mu - s) = f(\mu + s)$.

Lemma 4.3 shows that condition (L) is satisfied for $a = \mu - cs$, $a_n = \mu_n - cs_n$, $b = \mu + cs$ and $b_n = \mu_n + cs_n$, with

$$\begin{aligned}h_1 = &-\frac{1}{f(\mu)} I_{(-\infty,\mu]} \\ &+ \frac{c}{f(\mu + s) + f(\mu - s)} \left[ I_{[\mu-s,\mu+s]} - \frac{f(\mu + s) - f(\mu - s)}{f(\mu)} I_{(-\infty,\mu]} \right]\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}h_2 = &-\frac{1}{f(\mu)} I_{(-\infty,\mu]} \\ &- \frac{c}{f(\mu + s) + f(\mu - s)} \left[ I_{[\mu-s,\mu+s]} - \frac{f(\mu + s) - f(\mu - s)}{f(\mu)} I_{(-\infty,\mu]} \right].\end{aligned} \tag{4.13}$$

So, Theorem 4.1, and likewise Theorem 4.2, give the following:

THEOREM 4.3. (Asymptotic normality of Hampel's trimmed and winsorized means.) *Let $F$ be a distribution function with median $\mu$ and MAD $s$, and let $c$ be a positive constant. Assume $F$ has a continuous density on an open set containing $\mu$, $\mu + s$, $\mu - s$, $\mu + cs$ and $\mu - cs$, and that $f(\mu) > 0$, $f(\mu + s) + f(\mu - s) > 0$ and $f(\mu + cs) + f(\mu - cs) > 0$.* a) *Let $\theta_n$, $n \in \boldsymbol{N}$, be Hampel's trimmed mean corresponding to the constant $c$ and let $\theta$ be its population counterpart. With $h_1$ and $h_2$ defined by (4.12) and (4.13), set*

$$\begin{aligned}g(x) := &\frac{xI_{[\mu-cs,\mu+cs]}(x) + bf(b)h_2(x) - af(a)h_1(x)}{\int_{\mu-cs}^{\mu+cs} dF} \\ &- \frac{(\int_{\mu-cs}^{\mu+cs} t dF(t))(I_{[\mu-cs,\mu+cs]}(x) + f(b)h_2(x) - f(a)h_1(x))}{(\int_{\mu-cs}^{\mu+cs} dF)^2}.\end{aligned}$$

*Then,*

$$\sqrt{n}(\theta_n - \theta) \to \sqrt{\mathrm{Var}_F(g)}Z,$$

*in distribution, where $Z$ is standard normal.* b) *Let $\xi_n$, $n \in \boldsymbol{N}$, be Hampel's Winsorized mean corresponding to c and let $\xi$ be its population counterpart. With $h_1$ and $h_2$ defined by* (4.12) *and* (4.13), *set*

$$\begin{aligned} w(x) := {} & xI_{[\mu-cs,\mu+cs]}(x) + (\mu+cs)I_{(\mu+cs,\infty)}(x) + (\mu-cs)I_{(-\infty,\mu-cs)}(x) \\ & + (1-F(\mu+cs))h_2(x) + F(\mu-cs)h_1(x). \end{aligned}$$

*Then,*

$$\sqrt{n}(\xi_n - \xi) \to \sqrt{\mathrm{Var}_F(w)}Z,$$

*in distribution, where $Z$ is standard normal.*

The expressions for $g$ and $w$ are quite complicated. They simplify if $F$ is the c.d.f. of a symmetric distribution (meaning $F(x) + F(-x) = 1$ for all $x$). In this case

$$g(x) = \frac{1}{2\int_0^{cs} dF}\left(xI_{[-cs,cs]}(x) - \frac{2csf(cs)}{f(0)}I_{(-\infty,0]}(x)\right),$$

and

$$\mathrm{Var}_F(g) = \frac{1}{(2\int_0^{cs} dF)^2}\left(\int_{-cs}^{cs} x^2 dF(x) + \left(\frac{csf(cs)}{f(0)}\right)^2 + \frac{4csf(cs)}{f(0)}\int_0^{cs} xdF\right).$$

And (also in the symmetric case),

$$w(x) = xI_{[-cs,cs]}(x) + csI_{(cs,\infty)}(x) - csI_{(-\infty,-cs)} - \frac{2F(-cs)}{f(0)}I_{(-\infty,0)}$$

and

$$\begin{aligned} \mathrm{Var}_F(w) = {} & \int_{-cs}^{cs} x^2 dF + 2(cs)^2 F(-cs) + \left(\frac{F(-cs)}{f(0)}\right)^2 \\ & - \frac{4F(-cs)}{f(0)}\int_{-cs}^{0} xdF + \frac{4csF^2(-cs)}{f(0)}. \end{aligned}$$

Suppose $F$ is symmetric and $f(x)$ is continuous and bounded, and decreasing on $[0, \infty)$. Then, if we let $g_c$ and $w_c$ be the functions $g$ and $w$ in the above two theorems corresponding to the tuning parameter $c$, we see that both, $\mathrm{Var}\, g_c$ and $\mathrm{Var}\, w_c$ tend to 1 as $c \to \infty$ and tend to $1/(4f^2(0))$ as $c \to 0$. In particular, it is possible to choose, within the family of Hampel's trimmed (Winsorized) means, estimators of the center of symmetry of $F$ with breakdown point $1/2$ and with asymptotic efficiency with respect to the mean as good as one wishes. Hampel's means are always more asymptotically efficient than the median but if $c$ is chosen too small they become almost as 'asymptotically inefficient'.

### 4.3 *Box plot trimmed and Winsorized means*

For any $0 < \alpha < 1$ let $Q = F^{-1}(\alpha)$ and assume $F$ has a non-vanishing continuous density $f$ in a neighborhood of $Q$. Then, as with the median,

$$\sqrt{n}(Q_n - Q) = -\frac{1}{f(Q)}\nu_n(-\infty, Q] + o_P(1).$$

Therefore, assuming that $f$ exists and is strictly positive and continuous on an open set containing $Q_1$ and $Q_3$, we obtain

$$\sqrt{n}(R_n - R) = \nu_n\left(\frac{1}{f(Q_1)}I_{(-\infty,Q_1]} - \frac{1}{f(Q_3)}I_{(-\infty,Q_3]}\right) + o_P(1).$$

For the box plot trimmed mean we use $a_n = Q_{n,1} - cR_n$ and $b_n = Q_{n,3} + cR_n$ (see Section 2), so that, in this case,

$$\sqrt{n}(a_n - a) = \sqrt{n}[Q_{n,1} - cR_n - (Q_1 - cR)] = \nu_n(h_1) + o_P(1)$$

where

$$h_1 := -\frac{1}{f(Q_1)}I_{(-\infty,Q_1]} + c\left[\frac{1}{f(Q_3)}I_{(-\infty,Q_3]} - \frac{1}{f(Q_1)}I_{(-\infty,Q_1]}\right], \tag{4.14}$$

and

$$\sqrt{n}(b_n - b) = \sqrt{n}\left[Q_{n,3} + cR_n - (Q_3 + cR)\right] = \nu_n(h_2) + o_P(1) \tag{4.15}$$

where

$$h_2 := -\frac{1}{f(Q_3)}I_{(-\infty,Q_3]} - c\left[\frac{1}{f(Q_3)}I_{(-\infty,Q_3]} - \frac{1}{f(Q_1)}I_{(-\infty,Q_1]}\right].$$

So, the results of Subsection 4.1 imply the asymptotic normality of the box plot trimmed mean. Computations simplify in the symmetric case since then we have $\int_a^b x dF = 0$, $Q_1 = -Q_3$, $b = (2c+1)Q_3$ and $a = -b$. Here is the result.

THEOREM 4.4. *Assume that $F$ has a continuous density on an open set containing $Q_1$, $Q_3$, $Q_1 - cR$ and $Q_3 + cR$ for some constant $c \geq 0$, and that $f(Q_1) \neq 0$, $f(Q_3) \neq 0$ and $f(Q_1 - cR) + f(Q_3 + cR) \neq 0$. Let $\theta_n$ be the box plot trimmed mean corresponding to $c$, and let $\theta$ be its population counterpart. Then, $\sqrt{n}(\theta_n - \theta) \to \sigma Z$ converges in distribution to a normal random variable with variance equal to the variance of the random variable $g(X)$, $g$ defined as in Theorem 4.1 with $h_1$ and $h_2$ given respectively by* (4.14) *and* (4.15). *If $F$ is the c.d.f. of a symmetric distribution, then the variance of the normal limit is*

$$\sigma^2 = \frac{1}{[2F(b)-1]^2}\left(\int_{-b}^{b} x^2 dF(x) + \left(\frac{bf(b)}{f(Q_3)}\right)^2 + \frac{8bf(b)\int_{Q_3}^{b} x dF(x)}{f(Q_3)}\right),$$

*with $b = (2c+1)Q_3$.*

Likewise, the box plot Winsorized mean is also asymptotically normal, and in the symmetric case (with the same assumptions as in Theorem 4.4) the variance of the normal limit of $\sqrt{n}(\xi_n - \xi)$ is

$$\tau^2 = \int_{-b}^{b} x^2 dF(x) + 2b^2 F(-b) + \frac{F^2(-b)}{2f^2(Q_3)} + \frac{4F(-b)\int_{Q_3}^{b} x dF(x)}{f(Q_3)} + \frac{4bF^2(-b)}{f(Q_3)},$$

with $b = (2c+1)Q_3$.

## 5. Bootstrap

The variances of the normal limits of $\sqrt{n}(\theta_n - \theta)$ and $\sqrt{n}(\xi_n - \xi)$ are rather complicated and, even worse, depend on $f$. So, the results of Section 4 cannot be used to obtain asymptotic confidence intervals for $\theta$ or for $\xi$. The bootstrap provides a way out of this difficulty. One can always bootstrap *without* replacement and with a resampling sample size $m_n \to \infty$ with $m_n/n \to 0$ (Politis and Romano (1994)). However, as we prove in this section, depending on the properties of the estimators $a_n$ and $b_n$, Efron's bootstrap with resampling size $m_n = n$ may also be valid. We show in addition that this is the case for Hampel's and the box plot means. We present a justification of this, in all details, for Hampel's trimmed mean, and state without proofs the corresponding results for Hampel's Winsorized mean and the box plot means.

We recall that the $n$-th bootstrap sample $X^b_{n,1}, \ldots, X^b_{n,n}$ is obtained by sampling with replacement $n$ times from the original sample $X_1, \ldots, X_n$. We denote $F^b_n$, $P^b_n$ and $\nu^b_n$ respectively the empirical c.d.f., the empirical measure and the empirical process based on the $n$-th bootstrap sample: $P^b_n(A) = n^{-1}\sum_{i=1}^n \delta_{X^b_{n,i}}(A)$, $F^b_n(x) = P^b_n(-\infty, x]$ and $\nu_n = \sqrt{n}(P^b_n - P_n)$. The bootstrap median, $\mu^b_n$, is the median of the bootstrap sample, and the bootstrap MAD, $s^b_n$, is the median of the set of points $|X^b_{n,1} - \mu^b_n|, \ldots, |X^b_{n,n} - \mu^b_n|$. Finally, we denote by $\Pr_b = \Pr_b(\omega)$ the conditional probability given the sample $X_1, \ldots, X_n$ (its dependence on $\omega$ will not be displayed unless absolutely necessary). Also, $\mathcal{L}^b$ will denote conditional law given the sample. The symbol $o_{P_b}(1)$ will mean the following: $U_n(X^b_{n,1}, X^n_{n,n}, X_1, \ldots, X_n)$ is $o_{P_b}(1)$ a.s. if a.s. $\Pr_b\{|U_n| > \varepsilon\} \to 0$ for every $\varepsilon > 0$.

We briefly list the extra key ingredients needed in the proofs that follow. Since the class of functions $\mathcal{F} = \{I_{[a,b]}(x), I_{(-\infty,a]}(x), xI_{[a,b]}(x) : -\infty < c_1 \le a < b \le c_2 < \infty\}$ is uniform $P$-Donsker (or, equivalently, uniformly pregaussian) (see Giné (1997), Example 2.5.2), we have (Giné (1997), Theorem 2.6.1, p. 139)

$$\lim_{n\to\infty} \sup_{\substack{g,h\in\mathcal{F} \\ E_P(g-h)^2<1/(\log\log n)^2}} |\nu_n(g) - \nu_n(h)| = 0 \quad \text{a.s.} \tag{5.1}$$

Also, by the Giné-Zinn bootstrap central limit theorem (e.g., Giné (1997), Theorem 2.3.2),

$$\nu^b_n \to_{\mathcal{L}^b} G_P \quad \text{in} \quad \ell_\infty(\mathcal{F}) \quad \text{a.s.} \tag{5.2}$$

where $\to_{\mathcal{L}^b}$ denotes convergence in law conditionally on the sample and $G_P$ is a centered Gaussian process indexed by $\mathcal{F}$ with covariance $P(gh) - P(g)P(h)$. Moreover, by the asymptotic equicontinuity associated to a limit theorem in $\ell_\infty(\mathcal{F})$, we also have

$$\lim_{\delta\to 0}\limsup_n \Pr_b\left\{\sup_{\substack{h,g\in\mathcal{F} \\ E_P(g-h)^2\le\delta}} |\nu^b_n(g) - \nu^b_n(h)| > \varepsilon\right\} = 0 \tag{5.3}$$
$$\text{for all} \quad \varepsilon > 0, \quad \text{a.s.}$$

Also (Bickel and Freedman (1981), see also Giné (1997)) the median and the quantiles bootstrap a.s., that is,

$$\mathcal{L}^b(\sqrt{n}(\mu^b_n - \mu_n)) \to N(0, 1/(4f^2(\mu))),$$

$$\mathcal{L}^b(\sqrt{n}(Q_{n,1}^b - Q_{n,1})) \to N(0, 3/(16f^2(Q_1)))$$
$$\mathcal{L}^b(\sqrt{n}(Q_{n,3}^b - Q_{n,3})) \to N(0, 3/(16f^2(Q_3))) \quad \text{a.s.} \tag{5.4}$$

weakly a.s., assuming $f$ is continuous and non-vanishing at $\mu$ for the first limit, at $Q_1$ for the second and at $Q_3$ for the third.

Even though we could do with less, we will assume

(D.3) $P$ has a density $f$ on $\boldsymbol{R}$, the set $B_f = \{f > 0\}$ is open and $f$ is continuous on $B_f$.

Then we have the following general theorem for randomly trimmed and Winsorized means, asserting that if the $a_n$ and $b_n$ bootstrap, so do the corresponding means:

THEOREM 5.1. *Assume* (D.2), (D.3), (L), $a, b \in B_f$, $a_n \to a$ *a.s. and* $b_n \to b$ *a.s. Assume also that* $a_n^b$ *and* $b_n^b$, *defined respectively as* $a_n^b = a(F_n^b)$ *and* $b_n^b = b(F_n^b)$, *satisfy*

$$(Lb) \qquad \sqrt{n}(a_n^b - a_n) = \nu_n^b(h_1) + o_{P_b}(1), \qquad \sqrt{n}(b_n^b - b_n) = \nu_n^b(h_2) + o_{P_b}(1) \qquad a.s.$$

*Then, if* $\theta_n^b$ *is the bootstrap trimmed mean,*

$$\theta_n^b = \frac{\sum_{i=1}^n X_{n,i}^b I_{[a_n^b, b_n^b]}(X_{n,i}^b)}{\sum_{i=1}^n I_{[a_n^b, b_n^b]}(X_{n,i}^b)}, \qquad n \in \boldsymbol{N},$$

*and* $\theta_n$, $\theta$, $g$ *and* $Z$ *are as in Theorem* 4.1, *we have*

$$\sqrt{n}(\theta_n^b - \theta_n) \to_{\mathcal{L}^b} \sqrt{\mathrm{Var}_F(g)}Z \qquad a.s.$$

*Moreover, if* $\xi_n$, $\xi$ *and* $w$ *are as in Theorem* 4.2, *and* $w_n^b$ *is the bootstrap counterpart of* $w_n$ *(defined in analogy with* $\theta_n^b$*), we also have*

$$\sqrt{n}(\xi_n^b - \xi_n) \to_{\mathcal{L}^b} \sqrt{\mathrm{Var}_F(w)}Z \qquad a.s.$$

PROOF. (Sketch) We only prove the result for trimmed means, as the proof for Winsorized means is not different, aside from formalities. The proof follows the steps of the proofs of Lemma 4.1 and Theorem 4.1. The main point consists in proving the representations

$$\sqrt{n}\int_{a_n^b}^{b_n^b} d(F_n^b - F_n) = \nu_n^b[a, b] + o_{P_b}(1) \qquad \text{a.s.},$$
$$\sqrt{n}\int_{b_n}^{b_n^b} dF_n = f(b)\nu_n^b(h_2) + o_{P_b}(1) \quad \text{a.s.} \qquad \text{and}$$
$$\sqrt{n}\int_{a_n^b}^{a_n} dF_n = -f(a)\nu_n^b(h_1) + o_{P_b}(1) \quad \text{a.s.},$$

as well as similar representations for $\sqrt{n}\int_{a_n^b}^{b_n^b} xd(F_n^b - F_n)$, $\sqrt{n}\int_{b_n}^{b_n^b} xdF_n$ and $\sqrt{n}\int_{a_n^b}^{a_n} xdF_n$, as they give bootstrap versions of (4.2) and (4.3), which in turn can be used to complete the proof of the theorem just as in the proof of Theorem 4.1, but now invoking the bootstrap CLT instead of the regular CLT.

Let $\Omega_1$ be the subset of $\Omega$ of probability one where (Lb) holds, $a_n \to a$, $b_n \to b$, and (5.3) hold. Then, for any $\omega \in \Omega_1$ fixed, $a_n^b$ is close in $\Pr_b(\omega)$ probability to $a_n$, which is in turn close to $a$, and likewise for $b_n^b$, and therefore we can use the bootstrap asymptotic equicontinuity (5.3) to get the three representations above (in the same way as we used the asymptotic equicontinuity of $\nu_n$ in the proof of Lemma 4.1). □

Next we will apply Theorem 5.1 to show that Hampel's means bootstrap a.s. By Theorem 5.1, this reduces to obtaining representations for $\sqrt{n}(\mu_n^b - \mu_n)$ and $\sqrt{n}(s_n^b - s_n)$ in terms of the bootstrap empirical process $\nu_n^b$ (i.e., the bootstrap analogue of Lemma 4.3). Here is the representation for the bootstrap median.

LEMMA 5.1. *If* (D.3) *holds and* $\mu \in B_f$, *then*,

$$\sqrt{n}(\mu_n^b - \mu_n) = -\frac{1}{f(\mu)}\nu_n^b(-\infty, \mu] + o_{P_b}(1) \qquad a.s.$$

PROOF. The sample points $X_i$ are all different a.s. by the assumption on $f$. Also by the assumption on $f$, if $a$ and $b$ are near $\mu$, then $e_P^2(a,b) := (P(a,b])^2 \geq c|b-a|^2$ for some $c > 0$, hence, by (5.1),

$$\sup_{\substack{|a-b|<1/\log n \\ |a-\mu|<1/\log n, |b-\mu|<1/\log n}} \left| \int_a^b d\nu_n \right| \to 0 \quad \text{a.s.,} \tag{5.5}$$

and, by (5.3), we have that, almost surely,

$$\lim_{\delta\to 0} \limsup_n \Pr_b \left\{ \sup_{a,b: P[a,b]<\delta} |\nu_n^b(a,b]| > \varepsilon \right\} \to 0 \tag{5.6}$$

for all $\varepsilon > 0$. Let $\Omega_1 \subseteq \Omega$ be a set of probability one where i) all the $X_i$ are different, ii) $\mu_n \to \mu$, iii) $\sqrt{n}(\mu_n^b - \mu_n)$ converges in law conditionally on the sample, iv) the limit (5.5) holds, and v) the limit (5.6) holds. If $m_{n,i}$ denotes the number of terms from the bootstrap sample $X_{n,1}^b, \ldots, X_{n,n}^b$ equal to $X_i$, then $m_{n,i}$ is Binomial$(n, 1/n)$, and therefore, by a well known inequality for binomial probabilities, for all $k \leq n$,

$$\Pr_b \left\{ \max_{1\leq i\leq n} m_{n,i} > k \right\} \leq n\Pr_b\{m_{n,1} > k\} \leq n(e/k)^k,$$

which tends to zero for instance for $k = [\log n]$. That is, on $\Omega_1$, the conditional probability that $F_n^b$ has a jump of size larger than $n^{-1}\log n$ tends to zero, which implies

$$F_n^b(\mu_n^b) = \frac{1}{2} + O\left(\frac{\log n}{n}\right) \qquad \text{with } \Pr_b\text{-probability tending to 1.}$$

(This statement holds a.s., more concretely, it holds for each $\omega \in \Omega_1$; this will also be true, without further mention, for the identities in this proof that follow.) Also,

$$F_n(\mu_n^b) = \frac{1}{2} + O\left(\frac{1}{n}\right) + (F_n(\mu_n^b) - F_n(\mu_n))$$

so that, subtracting this identity from the previous one and multiplying by $\sqrt{n}$,

$$\nu_n^b(-\infty,\mu_n^b] = -\int_{\mu_n}^{\mu_n^b} d\nu_n - \sqrt{n}\int_{\mu_n}^{\mu_n^b} dF + o_{P_b}(1)$$

(note that $\alpha_n \to 0$ trivially implies $\alpha_n = o_{P_b}(1)$). Now, since $\mu_n^b - \mu_n = O_{P_b}(1/\sqrt{n})$, (5.5) gives

$$\int_{\mu_n}^{\mu_n^b} d\nu_n = o_{P_b}(1).$$

Since, moreover, by consistency and bootstrap consistency of the median,

$$\sqrt{n}\int_{\mu_n}^{\mu_n^b} dF = \sqrt{n}(\mu_n^b - \mu_n)(f(\mu) + o_{P_b}(1) + o(1)),$$

we obtain

$$\nu_n^b(-\infty,\mu_n^b] = -\sqrt{n}\int_{\mu_n}^{\mu_n^b} dF + o_{P_b}(1) = -f(\mu)\sqrt{n}(\mu_n^b - \mu_n) + o_{P_b}(1).$$

On the other hand, by the bootstrap of the median, the fact that $\mu_n \to \mu$ and the conditional asymptotic equicontinuity (5.6) of the bootstrap empirical process, we have that

$$\nu_n^b(-\infty,\mu_n^b] = \nu_n^b(-\infty,\mu] + o_{P_b}(1).$$

The last two identities prove the lemma. □

To get a similar representation for the bootstrapped MAD, we need an intermediate result. Recall that $s_n$ is the median of the set of points $|X_1 - \mu_n|, \dots, |X_n - \mu_n|$; we define now $\tilde{s}_n^b$ as the median of the set of points $|X_{n,1}^b - \mu_n|, \dots, |X_{n,n}^b - \mu_n|$, whereas the bootstrap MAD, $s_n^b$, is the median of $|X_{n,1}^b - \mu_n^b|, \dots, |X_{n,n}^b - \mu_n^b|$.

LEMMA 5.2. *If* (D.3) *holds and* $\mu, \mu+s, \mu-s \in B_f$, *then, the sequence* $\{\sqrt{n}(\tilde{s}_n^b - s_n)\}_{n=1}^{\infty}$ *converges in conditional distribution given the sample, almost surely.*

Note that if $\mu_n$ were replaced by $\mu$ in the definitions of $s_n$ and $\tilde{s}_n^b$, this lemma would follow from the bootstrap of the median for $|X - \mu|$. We cannot, however, do this directly because $\mu_n - \mu$ is not $o(1/\sqrt{n})$. On the other hand, we can prove Lemma 5.2 by a straightforward adaptation of the proof of the bootstrap of the median in Giné (1997) pp. 141–142, keeping in mind that 1) $s_n \to s$ a.s., 2) almost surely, there are no three equal terms among $|X_1 - \mu_n|, \dots, |X_n - \mu_n|$ for any $n$ (this is so because, given that all the $X_i$ are different, if three terms are equal here then, for some $i, j, k$, $(X_i + X_j)/2 = (X_j + X_k)/2 = m_n$, which implies $X_i = X_j$) and 3) one must apply (5.1), (5.2) and (5.3) to the empirical process indexed by the class of sets $\{x : |x - b| \le a\}$, $b \in \boldsymbol{R}$, $a \ge 0$, instead of to the classical empirical process. We skip the details.

LEMMA 5.3. *If* (D.3) *holds and* $\mu, \mu+s, \mu-s \in B_f$, *then,*

$$\begin{aligned}\sqrt{n}(s_n^b - s_n) &= -\frac{1}{f(\mu+s)+f(\mu-s)}\nu_n^b[\mu-s,\mu+s] \\ &\quad + \frac{1}{f(\mu)}\frac{f(\mu+s)-f(\mu-s)}{f(\mu+s)+f(\mu-s)}\nu_n^b(-\infty,\mu] + o_{P_b}(1) \qquad a.s.\end{aligned}$$

PROOF. Since

$$|s_n^b - s_n| \le |s_n^b - \tilde{s}_n^b| + |\tilde{s}_n^b - s_n| \le |\mu_n^b - \mu_n| + |\tilde{s}_n^b - s_n|,$$

it follows from the bootstrap of the median and from Lemma 5.1 that

$$(5.7) \qquad s_n^b - s_n = o_{P_b}(1) \quad \text{and} \quad \sqrt{n}(s_n^b - s_n) = O_{P_b}(1) \quad \text{a.s.}$$

Let now $H_n$ be the c.d.f. of $|X_1 - \mu_n|, \ldots, |X_n - \mu_n|$ and let $H_n^b$ be the c.d.f. of $|X_{n,1}^b - \mu_n^b|, \ldots, |X_{n,n}^b - \mu_n^b|$. Since all the $X_i$ are different a.s., by an argument just before the statement of the lemma, $H_n$ has jumps of size at most $2/n$ a.s., and then, by the argument below (5.6) in the proof of Lemma 5.1, $\tilde{H}_n^b$ has jumps of size most of the order of $n^{-1}\log n$ with bootstrap probability tending to 1 a.s., so that we have

$$\tilde{H}_n^b(s_n^b) = \frac{1}{2} + O\left(\frac{\log n}{n}\right) \quad \text{and} \quad H_n(s_n^b) = \frac{1}{2} + O\left(\frac{1}{n}\right) + (H_n(s_n^b) - H_n(s_n))$$

with $\Pr_b$-probability tending to 1 a.s. Also, (5.3) and (5.7) imply

$$\sqrt{n}\left[\int_{\mu_n - s_n^b}^{\mu_n + s_n^b} d(F_n - F) - \int_{\mu_n - s_n}^{\mu_n + s_n} d(F_n - F)\right] = o_{P_b}(1) \quad \text{a.s.}$$

and therefore we get

$$(5.8) \qquad \sqrt{n}(\tilde{H}_n^b(s_n^b) - H_n(s_n^b)) = \sqrt{n}\left[\int_{\mu_n - s_n^b}^{\mu_n + s_n^b} dF - \int_{\mu_n - s_n}^{\mu_n + s_n} dF\right] + o_{P_b}(1) \quad \text{a.s.}$$

By the hypothesis on $F$ and since $\mu_n \to \mu$ and $s_n \to s$ a.s. and $s_n^b - s_n \to 0$ in $\Pr_b$ a.s., this gives

$$\sqrt{n}(\tilde{H}_n^b(s_n^b) - H_n(s_n^b)) = \sqrt{n}(s_n^b - s_n)(f(\mu + s) + f(\mu - s)) + o_{P_b}(1) \quad \text{a.s.}$$

On the other hand, for the same reasons plus the conditional asymptotic equicontinuity of the bootstrap empirical process ((5.3) or (5.6)), the left hand side of (5.8) becomes

$$\begin{aligned}
\sqrt{n}(\tilde{H}_n^b - H_n)(s_n^b) &= \sqrt{n}(P_n^b[\mu_n^b - s_n^b, \mu_n^b + s_n^b] - P_n[\mu_n - s_n^b, \mu_n + s_n^b]) \\
&= \nu_n^b[\mu_n^b - s_n^b, \mu_n^b + s_n^b] \\
&\quad + \sqrt{n}(P_n[\mu_n^b - s_n^b, \mu_n^b + s_n^b] - P_n[\mu_n - s_n^b, \mu_n + s_n^b]) \\
&= \nu_n^b[\mu - s, \mu + s] + \sqrt{n}\left[\int_{\mu_n^b - s_n^b}^{\mu_n^b + s_n^b} d(F_n - F) - \int_{\mu_n - s_n^b}^{\mu_n + s_n^b} d(F_n - F)\right] \\
&\quad + \sqrt{n}\left[\int_{\mu_n^b - s_n^b}^{\mu_n^b + s_n^b} dF - \int_{\mu_n - s_n^b}^{\mu_n + s_n^b} dF\right] \\
&= \nu_n^b[\mu - s, \mu + s] + (f(\mu + s) - f(\mu - s))\sqrt{n}(\mu_n^b - \mu_n) \\
&\quad + o_{P_b}(1) \quad \text{a.s.}
\end{aligned}$$

The lemma follows from the last three displays and Lemma 5.1. □

In particular, this lemma shows that the MAD bootstraps a.s.

As a consequence of Theorem 5.1, Lemma 5.1 and Lemma 5.3 give the bootstrap central limit theorem for the Hampel's sample trimmed and Winsorized means. Here are the precise statements:

THEOREM 5.2. (Bootstrap of Hampel's trimmed and Winsorized means.) *Let $F$ be a c.d.f. satisfying condition* (D.3) *and such that $\mu$, $\mu+s$, $\mu-s$, $\mu+cs$ and $\mu-cs$ are in $B_f$ for some constant $c>0$. Let $\theta_n$ be Hampel's trimmed mean corresponding to the constant $c$ and let $\theta_n^b$ be its bootstrapped version,*

$$\theta_n^b := \frac{\sum_{i=1}^n X_{n,i}^b I_{[\mu_n^b - cs_n^b, \mu_n^b + cs_n^b]}(X_{n,i}^b)}{\sum_{i=1}^n I_{[\mu_n^b - cs_n^b, \mu_n^b + cs_n^b]}(X_{n,i}^b)},$$

*for each $n \in \boldsymbol{N}$. Then,*

$$\sqrt{n}(\theta_n^b - \theta_n) \to_{\mathcal{L}^b} \sqrt{\mathrm{Var}_F(g)} Z \qquad a.s.$$

*where $g$ and $Z$ are as in Theorem* 4.3. *Also, if $\xi_n$ is Hampel's Winsorized mean for some $c>0$ and $\xi_n^b$ is its bootstrap counterpart, then, almost surely, $\sqrt{n}(\xi_n^b - \xi_n)$ converges in conditional law given the sample to $\sqrt{\mathrm{Var}_F(w)}Z$, where $w$ and $Z$ are as in Theorem* 4.3.

Clearly, similar arguments prove that the bootstrap works as well for the box plot trimmed and Winsorized means. We skip the proofs since they are so similar to the proof of Theorem 5.2.

THEOREM 5.3. *Let $F$ be a c.d.f. satisfying condition* (D.3) *and such that the density $f$ is positive at $Q_1$, $Q_3$, $Q_1 - cR$ and $Q_3 + cR$ for some $c \geq 0$. Let $\theta_n$ and $\xi_n$ be the box plot trimmed and Winsorized means corresponding to the constant $c$ respectively, and let $\theta_n^b$ and $\xi_n^b$ be their bootstrap counterparts. Then, almost surely, the conditional distributions given the sample of $\sqrt{n}(\theta_n^b - \theta_n)$ and $\sqrt{n}(\xi_n^b - \xi_n)$ respectively converge to the limiting distributions of $\sqrt{n}(\theta_n - \theta)$ and $\sqrt{n}(\xi_n - \xi)$.*

## REFERENCES


Bickel, P. J. and Freedman, F. (1981). Some asymptotic theory for the bootstrap, *Annals of Statistics*, **9**, 1196–1216.

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

Dudley, R. M. (1999). *Uniform Central Limit Theorems*, Cambridge Press, Cambridge, U.K.

Falk, M. (1997). Asymptotic independence of median and MAD, *Statistics and Probability Letters*, **34**, 341–345.

Giné, E. (1997). *Lectures on Some Aspects of the BootStrap*, Lecture Notes in Mathematics, **1665**, 37–151.

Hall, P. and Padmanabhan, A. R. (1992). On the bootstrap and the trimmed mean, *Journal of Multivariate Analysis*, **41**, 132–153.

Hall, P. and Welsh, A. H. (1985). Limit theorems for the median deviation, *Annals of the Institute of Statistical Mathematics*, **37**, 27–36.

Hampel, F. R. (1971). A general qualitative definition of robustness, *Annals of Mathematical Statistics*, **42**, 1887–1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, **69**, 383–392.

Hampel, F. R. (1985). The breakdown points of the mean combined with some rejection rules, *Technometrics*, **27**, 95–106.

Huber, P. (1969). *Théorie de l'Inference Statistique Robuste*, Les Presses de l'Université de Montréal, Montréal.

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

Kim, S.-J. (1992). The metrically trimmed mean as a measure of location, *Annals of Statistics*, **20**, 1534–1547.

Politis, D. N. and Romano, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions, *Annals of Statistics*, **22**, 2031–2050.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Shorack, G. R. (1974). Random means, *Annals of Statistics*, **2**, 661–675.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications in Statistics*, Wiley, New York.

Stigler, S. M. (1973). The asymptotic distribution of the trimmed mean, *Annals of Statistics*, **1**, 472–477.

van der Vaart, A. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer, New York.