Exponential Probability Inequality and Convergence Results for the Median Absolute Deviation and Its Modifications

Robert Serfling[1] and Satyaki Mazumder[2]

University of Texas at Dallas

April 2009
(Revision for *Statistics & Probability Letters*)

[1]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA. Email: serfling@utdallas.edu. Website: www.utdallas.edu/~serfling. Support by NSF Grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

[2]Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083-0688, USA.

Abstract

The median absolute deviation from the median (MAD) is an important robust univariate spread measure. It also plays important roles with multivariate data through statistics based on the univariate projections of the data, in which case a modified sample MAD introduced by Tyler (1994) and Gather and Hilker (1997) is used to gain increased robustness. Here we establish for the modified sample MAD the same almost sure convergence to the population MAD shown by Hall and Welsh (1985) and Welsh (1986) for the usual sample MAD, and at the same time we eliminate the regularity assumptions imposed in the previous results. Our method is to establish for the sample MAD and modified versions an exponential probability inequality which yields the desired almost sure convergence and also carries independent interest. Further, the asymptotic joint normality of the sample median and the sample MAD established by Falk (1997) is extended to the modified sample MAD. Besides eliminating some regularity conditions, these results provide theoretical validation for use of the more general form of sample MAD.

AMS 2000 Subject Classification: Primary 62G05 Secondary 62G20

Key words and phrases: Median; Median absolute deviation; Exponential probability inequality; Almost sure convergence; Asymptotic normality.

1 Introduction

The median and MAD (median absolute deviation about the median) are nonparametric measures of location and scale with highly robust sample versions widely used in practice. Previous results on almost sure convergence of the sample MAD to the population MAD impose regularity conditions on the parent distribution. Here we remove these conditions via application of an exponential probability inequality, of general interest, that we prove for the sample MAD. Also, these results are obtained for a more general form of sample MAD that arises in certain contexts. In addition, it is shown that this modified sample MAD satisfies the same asymptotic joint normality with the sample median as holds for the usual sample MAD. Besides eliminating some regularity conditions, these results provide theoretical validation for use of the more general form of sample MAD.

Let us now make this precise. The *median* of a univariate distribution F, or $\mathrm{Med}(F)$, is defined by $\nu = F^{-1}(1/2) = \inf\{x : F(x) \geq 1/2\}$ and satisfies

$$F(\nu-) \leq 1/2 \leq F(\nu). \tag{1}$$

Let X have distribution F. The distribution G of $|X - \nu|$, i.e.,

$$G(y) = P(|X - \nu| \leq y) = F(\nu + y) - F(\nu - y-),\ y \in \mathbb{R}, \tag{2}$$

has median $\zeta = G^{-1}(1/2)$ satisfying

$$G(\zeta-) \leq 1/2 \leq G(\zeta). \tag{3}$$

The quantity ζ defines a scale parameter of F, the *median absolute deviation about the median (MAD)*, i.e., $\mathrm{Med}(G) = \mathrm{MAD}(F)$ (which is *not* the *mean* absolute deviation about the *mean*, sometimes also abbreviated by "MAD").

Sample versions Med_n and MAD_n for a random sample $\mathbb{X}_n = \{X_1, \ldots, X_n\}$ from F are defined as follows. With $X_{1:n} \leq \ldots \leq X_{n:n}$ the ordered sample values,

$$\mathrm{Med}_n = \frac{1}{2}\left(X_{\lfloor \frac{n+1}{2} \rfloor : n} + X_{\lfloor \frac{n+2}{2} \rfloor : n}\right).$$

Also, with $W^*_{1:n} \leq \ldots \leq W^*_{n:n}$ the ordered values of $W^*_i = |X_i - \mathrm{Med}_n|$, $1 \leq i \leq n$,

$$\mathrm{MAD}_n = \frac{1}{2}\left(W^*_{\lfloor \frac{n+1}{2} \rfloor : n} + W^*_{\lfloor \frac{n+2}{2} \rfloor : n}\right).$$

One attraction of Med_n and MAD_n is their 50% breakdown points. The interquartile range (IQR) also measures spread and agrees with the MAD for symmetric F, but the sample IQR has only 25% breakdown point. Among a wide range of applications using Med_n and MAD_n, we mention that metrically trimmed means based on observations within intervals of form $\mathrm{Med}_n \pm c\,\mathrm{MAD}_n$ yield high-breakdown analogues of the usual quantile-based trimmed means (see Hampel, 1985, Olive, 2001, and Chen and Giné, 2004).

In another direction of application, the $(\mathrm{Med}_n, \mathrm{MAD}_n)$ combination long has been used in univariate outlyingness functions of scaled deviation type (Mosteller and Tukey, 1977) and these arise also in connection with multivariate data sets $\mathbb{X}_n$ in dimension $d > 1$ through the associated univariate projections (Donoho and Gasko, 1992, and Liu, 1992). Then the breakdown points of various multivariate location and scatter statistics that are based on the projected $(\mathrm{Med}_n, \mathrm{MAD}_n)$ combinations can be improved by using instead of MAD_n a *modified sample MAD*. This is defined, for any choice of $k = 1, \ldots, n-1$, as

$$\mathrm{MAD}_n^{(k)} = \frac{1}{2}\left(W^*_{\lfloor \frac{n+k}{2} \rfloor : n} + W^*_{\lfloor \frac{n+k+1}{2} \rfloor : n}\right),$$

thus including MAD_n for $k = 1$.

The advantages of using $\mathrm{MAD}_n^{(k)}$ with $k > 1$ arise in a variety of settings involving data $\mathbb{X}_n$ in $\mathbb{R}^d$. For example, for $\mathbb{X}_n$ in "general position" (no more than d points of $\mathbb{X}_n$ in any $(d-1)$-dimensional subspace) with $n \geq d+1$ and with either $k = d$ or $k = d-1$, the uniform breakdown point of $(\mathrm{Med}_n, \mathrm{MAD}_n^{(k)})$ over all univariate projections attains an optimal value (Tyler, 1994, Gather and Hilker, 1997). Further, for data as sparse as $n \leq 2d$, the usual MAD_n is not even defined and the modification $\mathrm{MAD}_n^{(k)}$ for some $k > 1$ becomes essential, not merely an option for improving breakdown points. Also, again for $\mathbb{X}_n$ in general position, and with $n \geq 2(d-1)^2 + d$ and $k = d-1$, the projection median based on $(\mathrm{Med}_n, \mathrm{MAD}_n^{(k)})$ attains the optimal breakdown point possible for any translation equivariant location estimator (Zuo, 2003).

As $n \to \infty$ with k fixed, does $\mathrm{MAD}_n^{(k)}$ satisfy the same almost sure convergence to the population MAD, and does it have the same joint asymptotic normality with Med_n, as does MAD_n? These questions have not been explored previously. We establish in the present paper that the answers are affirmative, perhaps as expected and certainly as hoped, thus validating the use of $\mathrm{MAD}_n^{(k)}$. Also, by eliminating regularity conditions on F assumed in previous almost sure convergence results for MAD_n, we broaden the scope of practical application of MAD_n and extend to its modifications. (Of course, in the sparse data case, large n implies large d, and asymptotics with $n \to \infty$ but k fixed do not permit $k = d$ or $k = d-1$. Extension of the results of the present paper to allow $k = k_n \to \infty$ as $n \to \infty$ is deferred to a future investigation.)

Almost sure convergence of MAD_n to MAD was obtained by Hall and Welsh (1985), Welsh (1986), and Falk (1997a) under assumptions including continuity and strict monotonicity of F at ν and $\nu \pm \zeta$. Here we establish this convergence for $\mathrm{MAD}_n^{(k)}$ and without such restrictions, assuming only uniqueness of ν and ζ. This result is obtained by developing for $\mathrm{MAD}_n^{(k)}$ an exponential probability inequality similar to that which is well-known for Med_n. Besides yielding the desired almost sure convergence by a Borel-Cantelli argument, the inquality has interest and applications beyond the present paper.

Joint asymptotic normality of Med_n and MAD_n was derived by Falk (1997b) under mild regularity conditions on F and using elementary arguments. We show that exactly the same result holds with $\mathrm{MAD}_n^{(k)}$ in place of MAD_n. One can also derive joint asymptotic normality of Med_n and MAD_n using in-probability type Bahadur representations for Med_n

(Ghosh, 1971) and MAD_n (Hall and Welsh, 1985, Welsh, 1986, and Chen and Giné, 2004). However, these entail additional regularity assumptions on F not imposed in the simple and direct approach used here. (Development of a suitable Bahadur representation for MAD_n without such additional conditions, along with extension to $\text{MAD}_n^{(k)}$ as well as to an almost sure version, involves considerably more elaborate arguments and is deferred to a separate paper.) Also, weak convergence of a stochastic process based on the projected univariate (Med_n, MAD_n) combinations over all directions has been treated in Pan, Fung, and Fang (2000). However, that result entails considerable additional regularity on F and has not been extended to the modified MAD. It remains open to pursue these improvements and extensions.

For n and k both *odd*, Med_n and $\text{MAD}_n^{(k)}$ are given, conveniently, by single order statistics. Indeed, Falk (1997b) just treats the case n odd (with $k = 1$). For other cases of n and k, the desired results can be derived by simple arguments from identical results proved for certain choices of single order statistics. To this effect, for any fixed integers $\ell \geq 1$ and $m \geq 1$, let us put

$$\widehat{\nu}_n = X_{\lfloor \frac{n+\ell}{2} \rfloor : n} \tag{4}$$

and

$$\widehat{\zeta}_n = W_{\lfloor \frac{n+m}{2} \rfloor : n}, \tag{5}$$

with $W_{1:n} \leq \ldots \leq W_{n:n}$ the ordered values of $W_i = |X_i - \widehat{\nu}_n|$, $1 \leq i \leq n$. In particular, for $\ell = 1 = m$, the statistics $\widehat{\nu}_n$ and $\widehat{\zeta}_n$ agree with the *sample analogue estimators* of $\nu(F)$ and $\zeta(F)$, respectively,

$$\nu(\widehat{F}_n) = \widehat{F}_n^{-1}(1/2) = X_{\lfloor \frac{n+1}{2} \rfloor : n}$$

and

$$\zeta(\widehat{F}_n) = \text{MAD}(\widehat{F}_n) = W_{\lfloor \frac{n+1}{2} \rfloor : n},$$

where $\widehat{F}_n$ denotes the usual sample distribution function. Also, when n is odd, we have $\text{Med}_n = \nu(\widehat{F}_n)$ and $\text{MAD}_n = \zeta(\widehat{F}_n)$. When *both* n and k are odd, we have $\text{MAD}_n^{(k)} = \widehat{\zeta}_n$ for $m = k$.

Therefore, for $\widehat{\nu}_n$ and $\widehat{\zeta}_n$ given by (4) and (5), we establish the following convergence results:

$$\widehat{\zeta}_n \textit{ converges almost surely to } \zeta, \ n \to \infty, \tag{6}$$

$$\widehat{\nu}_n \textit{ and } \widehat{\zeta}_n \textit{ are asymptotically jointly normal}, \ n \to \infty. \tag{7}$$

In Section 2, we develop an exponential probability inequality for $\widehat{\zeta}_n$ and obtain (6). Section 3 establishes (7).

2 Exponential Probability Inequality for $\text{MAD}_n^{(k)}$, and Almost Sure Convergence

The following exponential probability inequality for $\widehat{\zeta}_n$ defined by (4) and (5) is of general interest.

Theorem 1 *Suppose that* $\nu = F^{-1}(1/2) = Med(F)$ *is the unique solution of (1) and that* $\zeta = G^{-1}(1/2) = MAD(F)$ *is the unique solution of (3). Define* $\widehat{\zeta}_n$ *by (4) and (5), for fixed positive integers* ℓ *and* m*. Then, for every* $\varepsilon > 0$,

$$P(|\widehat{\zeta}_n - \zeta| > \varepsilon) \leq 6e^{-2n\Delta^2_{\varepsilon,n}}, \tag{8}$$

where $\Delta_{\varepsilon,n}$ $(= \Delta_{\varepsilon,n}(\ell, m))$ $= \min\{a_0, b_0, c_0, d_0\}$*, with*

$$\begin{aligned}
a_0 &= (F(\nu + \varepsilon/2) - (\lfloor(n+\ell)/2\rfloor - 1)/n)^+ , \\
b_0 &= \lfloor(n+\ell)/2\rfloor/n - F(\nu - \varepsilon/2), \\
c_0 &= (F(\nu + \zeta + \varepsilon/2) - F(\nu - \zeta - \varepsilon/2) - \lfloor(n+m)/2\rfloor/n)^+ , \\
d_0 &= \lfloor(n+m)/2\rfloor/n - [F(\nu + \zeta - \varepsilon/2) - F(\nu - \zeta + \varepsilon/2)].
\end{aligned}$$

The proof will utilize the following similar probability inequality for $\widehat{\nu}_n$ defined by (4), which is a special case of Theorem 2.2 of Serfling (1992).

Lemma 2 *Suppose that* $\nu = F^{-1}(1/2) = Med(F)$ *is the unique solution of (1). Define* $\widehat{\nu}_n$ *by (4), for fixed positive integer* ℓ*. Then, for every* $\varepsilon > 0$,

$$P(|\widehat{\nu}_n - \nu| > \varepsilon/2) \leq 2e^{-2n\delta^2_{\varepsilon,n}}, \tag{9}$$

where $\delta_{\varepsilon,n}$ $(= \delta_{\varepsilon,n}(\ell))$ $= \min\{a_0, b_0\}$ *with* a_0 *and* b_0 *as above.*

Corollary 3 Strong convergence results. *Under the conditions of Theorem 1,* $\widehat{\nu}_n$ *and* $\widehat{\zeta}_n$ *converge almost surely to* ν *and* ζ*, respectively, as* $n \to \infty$.

Proof of Corollary 3. (i) The definition and assumption of uniqueness of ν imply $F(\nu - \varepsilon/2) < 1/2 < F(\nu + \varepsilon/2)$ and hence

$$\delta_{\varepsilon,n} \to \delta_\varepsilon = \min\{F(\nu + \varepsilon/2) - 1/2, 1/2 - F(\nu - \varepsilon/2)\} > 0, \ n \to \infty.$$

Then $\delta_{\varepsilon,n} > \delta_\varepsilon/2 > 0$ for all sufficiently large n, and, under no further conditions on F, a standard Borel-Cantelli argument yields almost sure convergence of $\widehat{\nu}_n$ to ν. In fact, this establishes "complete convergence" of $\widehat{\nu}_n$ to ν in the sense of Hsu and Robbins (1947).

(ii) Likewise, since also $\zeta = G^{-1}(1/2) = \mathrm{MAD}(F)$ is the unique solution of (3), it follows that $\Delta_{\varepsilon,n}$ has a positive limit Δ_ε as $n \to \infty$, and thus $\widehat{\zeta}_n$ converges completely to ζ. □

Proof of Theorem 1. Let $\varepsilon > 0$. Put $\alpha_n = \lfloor\frac{n+m}{2}\rfloor/n$. We have

$$P(|\widehat{\zeta}_n - \zeta| > \varepsilon) = P(\widehat{\zeta}_n < \zeta - \varepsilon) + P(\widehat{\zeta}_n > \zeta + \varepsilon).$$

We now introduce the sample analogue estimator for the distribution G induced via (2) relative to $\widehat{\nu}_n$ as estimator of ν,

$$\widehat{G}_n(y) = \widehat{F}_n(\widehat{\nu}_n + y) - \widehat{F}_n(\widehat{\nu}_n - y-), \ y \in \mathbb{R}. \tag{10}$$

Then

$$\begin{aligned} P(\widehat{\zeta}_n > \zeta + \varepsilon) &= P(\widehat{G}_n^{-1}(\alpha_n) > \zeta + \varepsilon) = P(\alpha_n > \widehat{G}_n(\zeta + \varepsilon)) \\ &\le P(A_n) \le P(B_n) + P(C_n), \end{aligned}$$

with

$$\begin{aligned} A_n &= \{\alpha_n > \widehat{F}_n(\widehat{\nu}_n + \zeta + \varepsilon) - \widehat{F}_n(\widehat{\nu}_n - \zeta - \varepsilon)\}, \\ B_n &= \{\alpha_n > \widehat{F}_n(\nu + \zeta + \varepsilon/2) - \widehat{F}_n(\nu - \zeta - \varepsilon/2)\}, \\ C_n &= \{|\widehat{\nu}_n - \nu| > \varepsilon/2\}. \end{aligned}$$

Here we have used $\widehat{F}_n(x-) \le \widehat{F}_n(x)$, any x. By Lemma 2 we have

$$P(C_n) \le 2e^{-2n\Delta^2_{(1,\varepsilon,n)}},$$

with $\Delta_{(1,\varepsilon,n)} = \delta_{\varepsilon,n}$. To establish a similar bound for $P(B_n)$, we write

$$\begin{aligned} P(B_n) &\le P\Big(\alpha_n > \widehat{F}_n(\nu + \zeta + \varepsilon/2) - \widehat{F}_n(\nu - \zeta - \varepsilon/2)\Big) \\ &\le P\left(n\alpha_n > \sum_{i=1}^n \mathbf{1}(\nu - \zeta - \varepsilon/2 < X_i \le \nu + \zeta + \varepsilon/2)\right) \\ &\le P\left(n\alpha_n > \sum_{i=1}^n Y_i\right) \\ &\le P\left(n(\alpha_n - p_1) > \sum_{i=1}^n (Y_i - E(Y_i))\right) \\ &\le e^{-2n\Delta^2_{(2,\varepsilon,n)}}, \end{aligned}$$

with $Y_n = \mathbf{1}(\nu - \zeta - \varepsilon/2 < X_i \le \nu + \zeta + \varepsilon/2)$, $p_1 = E(Y_n) = F(\nu + \zeta + \varepsilon/2) - F(\nu - \zeta - \varepsilon/2)$, and $\Delta_{(2,\varepsilon,n)} = (p_1 - \alpha_n)^+$, and in the last step using the Hoeffding inequality (Hoeffding, 1963, or Serfling, 1980, Lemma 2.3.2). Note that, by (3), $p_1 > 1/2$. Thus we obtain

$$P(\widehat{\zeta}_n > \zeta + \varepsilon) \le 2e^{-2n\Delta^2_{(1,\varepsilon,n)}} + e^{-2n\Delta^2_{(2,\varepsilon,n)}}.$$

Now, using $\widehat{F}_n(x) \ge \widehat{F}_n(x-) \ge \widehat{F}_n(x - \beta)$, any $\beta > 0$ and any x, we have

$$\begin{aligned} P(\widehat{\zeta}_n < \zeta - \varepsilon) &\le P(\widehat{G}_n^{-1}(\alpha_n) \le \zeta - \varepsilon) = P(\alpha_n \le \widehat{G}_n(\zeta - \varepsilon)) \\ &\le P(\widetilde{B}_n) + P(\widetilde{C}_n), \end{aligned}$$

with

$$\begin{aligned} \widetilde{B}_n &= \{\alpha_n \le \widehat{F}_n(\nu+\zeta-\varepsilon/2) - \widehat{F}_n(\nu-\zeta+\varepsilon/2)\}, \\ \widetilde{C}_n &= \{|\nu_n - \nu| \ge \varepsilon/2\}. \end{aligned}$$

Then, with $p_2 = F(\nu+\zeta-\varepsilon/2) - F(\nu-\zeta+\varepsilon/2)$ and $\Delta_{(3,\varepsilon,n)} = \alpha_n - p_2$ (> 0 since $\alpha_n \ge 1/2 > p_2$, by (3)), similar steps as above yield

$$P(\widehat{\zeta}_n < \zeta - \varepsilon) \le 2e^{-2n\Delta^2_{(1,\varepsilon,n)}} + e^{-2n\Delta^2_{(3,\varepsilon,n)}}.$$

With $\Delta_{\varepsilon,n} = \min\{\Delta_{(1,\varepsilon,n)}, \Delta_{(2,\varepsilon,n)}, \Delta_{(3,\varepsilon,n)}\}$, the proof is complete. □

3 Joint Asymptotic Normality

For n odd and $\ell = 1 = m$, Falk (1997b) establishes joint asymptotic normality of $(\widehat{\nu}_n, \widehat{\zeta}_n)$ under simple regularity conditions on F. Here, by our variant of the same line of argument, we extend this result to $\ell \ge 1$ and $m \ge 1$.

Theorem 4 *With $\nu = F^{-1}(1/2) = Med(F)$ and $\zeta = G^{-1}(1/2) = MAD(F)$, suppose that F is continuous in neighborhoods of ν and $\nu \pm \zeta$ and differentiable at these points with $F'(\nu) > 0$ and $G'(\zeta) = F'(\nu-\zeta) + F'(\nu+\zeta) > 0$. Define $\widehat{\nu}_n$ and $\widehat{\zeta}_n$ by (4) and (5). Put $\alpha = F(\nu-\zeta) + F(\nu+\zeta)$, $\beta = F'(\nu-\zeta) - F'(\nu+\zeta)$, and $\gamma = \beta^2 + 4(1-\alpha)\beta F'(\nu)$. Then*

$$n^{1/2}(\widehat{\nu}_n - \nu, \widehat{\zeta}_n - \zeta) \xrightarrow{d} N((0,0), (\sigma_{ij})_{2\times 2}), \tag{11}$$

where

$$\begin{aligned} \sigma_{11} &= \frac{1}{4F'(\nu)^2}, \\ \sigma_{12} &= \sigma_{21} = \frac{1}{4F'(\nu)G'(\zeta)}\left(1 - 4F(\nu-\zeta) + \frac{\beta}{F'(\nu)}\right), \\ \sigma_{22} &= \frac{1}{4G'(\zeta)^2}\left(1 + \frac{\gamma}{F'(\nu)^2}\right). \end{aligned}$$

Corollary 5 *Under the conditions of Theorem 4, we have, for any k, $1 \le k \le n-1$, asymptotic independence of Med_n and $MAD_n^{(k)}$ if and only if*

$$1 - 4F(\nu-\zeta) + \frac{\beta}{F'(\nu)} = 0,$$

which, in particular, holds if X is symmetric about its (unique) median ν.

Proof of Theorem 4. We need to show that

$$C_n(s,t) := P(n^{1/2}(\widehat{\nu}_n - \nu) \le s, n^{1/2}(\widehat{\zeta}_n - \zeta) \le t) \to C(s,t),\ n \to \infty, \tag{12}$$

where $C(s,t) = P(aZ_1 \le s, bZ_1 + cZ_2 \le t)$, with Z_1 and Z_2 independent standard normal random variables and

$$\begin{aligned} a &= \frac{1}{2F'(\nu)}, \\ b &= \frac{1}{2G'(\zeta)}\left(1 - 4F(\nu-\zeta) + \frac{\beta}{F'(\nu)}\right), \\ c &= \frac{1}{G'(\zeta)}\left(2F(\nu-\zeta)(1-2F(\nu-\zeta))\right)^{1/2}. \end{aligned}$$

We follow with appropriate modifications and some small alterations the approach of Falk (1997b) for the case n odd and $\ell = 1 = m$. Using a standard representation $X_i = F^{-1}(U_i)$, $1 \le i \le n$, where $U_1, \ldots, U_n$ are independent uniform $(0,1)$ random variables with ordered values $U_{1:n} \le \ldots \le U_{n:n}$, we have $X_{i:n} = F^{-1}(U_{i:n})$, $1 \le i \le n$. Then, with $I_n = \lfloor \frac{n+\ell}{2} \rfloor$ and $J_n = \lfloor \frac{n+m}{2} \rfloor$, and for n sufficiently large that $\zeta + t/\sqrt{n} > 0$, we have

$$\begin{aligned} C_n(s,t) &= P\left(F^{-1}\left(U_{\lfloor \frac{n+\ell}{2} \rfloor : n}\right) \le \nu + \frac{s}{\sqrt{n}}, \widehat{\zeta}_n \le \zeta + \frac{t}{\sqrt{n}}\right) \\ &= P\left(U_{I_n:n} \le F\left(\nu + \frac{s}{\sqrt{n}}\right), \sum_{i=1}^n \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(U_i) - F^{-1}(U_{I_n:n})|) \ge J_n\right) \\ &= P\left(n^{-1/2}(U_{I_n:n} - 1/2) \le s_n, \sum_{i=1}^n \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(U_i) - F^{-1}(U_{I_n:n})|) \ge J_n\right) \\ &= \int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{n^{-1/2}(U_{I_n:n}-1/2)}(u)\, du, \end{aligned}$$

with $s_n = n^{-1/2}(F(\nu + \frac{s}{\sqrt{n}}) - 1/2)$ and

$$\beta_n(u,t) = P\left(\sum_{i=1}^n \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(U_i) - F^{-1}(U_{I_n:n})|) \ge J_n \,\Big|\, n^{-1/2}(U_{I_n:n} - 1/2) = u\right).$$

Now, with $u_n := 1/2 + u/\sqrt{n}$, the conditional distribution of $(U_{1:n}, \ldots, U_{n:n})$, given $U_{I_n:n} = u_n$, is the distribution of $(\theta_{1:I_n-1}, \ldots, \theta_{I_n-1:I_n-1}, u_n, \eta_{1:n-I_n}, \ldots, \eta_{n-I_n:n-I_n})$, where the vectors $(\theta_{1:I_n-1}, \ldots, \theta_{I_n-1:I_n-1})$ and $(\eta_{1:n-I_n}, \ldots, \eta_{n-I_n:n-I_n})$ are independent and denote the order statistics from a sample of size $I_n - 1$ from uniform $(0, u_n)$ and a sample of size $n - I_n$ from uniform $(u_n, 1)$, respectively. Thus

$$\begin{aligned}&\beta_n(u,t)\\ &= P\left(\sum_{i=1}^{I_n-1} \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(\theta_i) - F^{-1}(u_n)|) + \sum_{i=1}^{n-I_n} \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(\eta_i) - F^{-1}(u_n)|) \geq J_n\right).\end{aligned}$$

Now the sum

$$S_1 := \sum_{i=1}^{I_n-1} \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(\theta_i) - F^{-1}(u_n)|)$$

is Binomial$(I_n - 1, p_n(u_n))$, where $p_n(u_n) = P(|F^{-1}(\theta_1) - F^{-1}(u_n)| \leq \zeta + \frac{t}{\sqrt{n}})$, and

$$S_2 := \sum_{i=1}^{n-I_n} \mathbf{1}_{[0,\zeta+\frac{t}{\sqrt{n}}]}(|F^{-1}(\eta_i) - F^{-1}(u_n)|)$$

is Binomial$(n - I_n, q_n(u_n))$, where $q_n(u_n) = P(|F^{-1}(\eta_1) - F^{-1}(u_n)| \leq \zeta + \frac{t}{\sqrt{n}})$, and these sums are independent. Using continuity of F at ν and $\nu \pm \zeta$, it is readily seen that

$$\begin{aligned} p_n(u_n) &\to 1 - 2F(\nu - \zeta) =: p, \\ q_n(u_n) &\to 2F(\nu + \zeta) - 1 =: q, \end{aligned}$$

and, by uniqueness of ζ, that $p + q = 1$. Further, using $I_n - 1 \sim n - I_n \sim J_n \sim n/2$ and all of the regularity conditions on F, a routine but somewhat lengthy derivation yields

$$\begin{aligned} &(n/2)^{-1/2}(J_n - (I_n - 1)p_n(u_n) - (n - I_n)q_n(u_n)) \\ &\sim (n/2)^{1/2}(1 - p_n(u_n) - q_n(u_n)) \\ &\to \sqrt{2}\left(1 - 4F(\nu - \zeta) + \frac{\beta}{F'(\nu))}\right)u - \sqrt{2}G'(\zeta)t =: D(u,t). \end{aligned}$$

Then, applying the Berry-Esséen and Slutsky's theorems, we have

$$\begin{aligned} &(n/2)^{-1/2}(S_1 + S_2 - (I_n - 1)p_n(u_n) - (n - I_n)q_n(u_n)) \\ &\qquad \xrightarrow{d} N(0, p(1-p) + q(1-q)) = N(0, 2p(1-p)) \end{aligned}$$

and hence

$$\beta_n(u,t) \to P(N(0, 2p(1-p)) \geq D(u,t)), \ n \to \infty.$$

Now

$$\begin{aligned} &\left|C_n(s,t) - \int_{-\infty}^{F'(\nu)s} P(N(0,2p(1-p)) \geq D(u,t))\, f_{N(0,1/4)}(u)\, du\right| \\ &\leq \left|\int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{n^{-1/2}(U_{I_n:n}-1/2)}(u)\, du - \int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{N(0,1/4)}(u)\, du\right| \\ &\quad + \left|\int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{N(0,1/4)}(u)\, du - \int_{-\infty}^{F'(\nu)s} P(N(0,2p(1-p)) \geq D(u,t))\, f_{N(0,1/4)}(u)\, du\right|. \end{aligned}$$

The first term on the right in the above inequality satisfies

$$\begin{aligned}&\left|\int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{n^{-1/2}(U_{I_n:n}-1/2)}(u)\, du - \int_{-\sqrt{n}/2}^{s_n} \beta_n(u,t)\, f_{N(0,1/4)}(u)\, du\right| \\ &\leq \int_{-\infty}^{\infty} \left| f_{n^{-1/2}(U_{I_n:n}-1/2)}(u) - f_{N(0,1/4)}(u)\right| du \to 0,\ n \to \infty,\end{aligned}$$

using convergence of the density $f_{n^{-1/2}(U_{I_n:n}-1/2)}(u)$ to $f_{N(0,1/4)}(u)$ and Scheffé's Theorem (e.g., Serfling, 1980, §32.3.4 and Theorem 1.5.1C), and the second term → 0 by routine arguments using the Dominated Convergence Theorem. Hence

$$C_n(s,t) \to \int_{-\infty}^{F'(\nu)s} P(N(0, 2p(1-p)) \geq D(u,t))\, f_{N(0,1/4)}(u)\, du,\ n \to \infty,$$

and the claimed result follows in straightforward fashion. □

Acknowledgment

The authors greatly appreciate the very constructive remarks by two reviewers. These have led to substantial improvements in the presentation. Also, support under National Science Foundation Grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is gratefully acknowledged.

References

[1] Chen, Z. and Giné, E. (2004). Another approach to asymptotics and bootstrap of randomly trimmed means. *Annals of the Institute of Statistical Mathematics* **56** 771–790.

[2] Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estinates based on halfspace depth and projected outlyingness. *Annals of Statistics* **20** 1803–1827.

[3] Falk, M. (1997a) On MAD and COMEDIANS. *Annals of Institute of Statistical Mathematics* **49** 615–644.

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

[5] Gather, U. and Hilker, T. (1997). A note on Tyler's modification of the MAD for the Stahel-Donoho estimator. *Annals of Statistics* **25** 2024-2026.

[6] Ghosh, J. K. (1971). A new proof of the Bahadur representation of sample quantiles and an application. *Annals of Mathematical Statistics* **42** 1957–1961.

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

[8] Hampel, F. R. (1985) The breakdown points of the mean combined with some rejection rules. *Technometrics* **27** 95–107.

[9] Hoeffding, W. (1963). Probability inequalities for sum of bounded random variables. *Journal of the American Statistical Association* **36** 369–408.

[10] Hsu, P. L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proceedings of the National Academy of Sciences U.S.A* **33** 25–31.

[11] Liu, R. Y. (1992). Data depth and multivariate rank tests. In *L_1-Statistics and Related Methods* (Y. Dodge, ed.), pp. 279–294, North-Holland, Amsterdam.

[12] Mosteller, C. F. and Tukey, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, Mass.

[13] Olive, D. J. (2001). High breakdown analogs of the trimmed mean. *Statistics and Probability Letters* **51** 87–92.

[14] Pan, J.-X., Fung, W.-K., and Fang, K.-T. (2000). Multiple outlier detection in multivariate data using projection pursuit techniques. *Journal of Statistical Planning and Inference* **83** 153–167.

[15] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

[16] Serfling, R. (1992). Nonparametric confidence intervals for generalized quantile parameters in multi-sample contexts. In *Nonparametric Statistics and Related Topics* (A. K. Md. E. Saleh, ed.), pp. 121–139, Elsevier Science Publishers B.V.

[17] Tyler, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Annals of Statistics* **22** 1024–1044.

[18] Welsh, A. H. (1986). Bahadur representations for robust scale estimators based on regression residuals. *Annals of Statistics* **14** 1246–1251.

[19] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.